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.

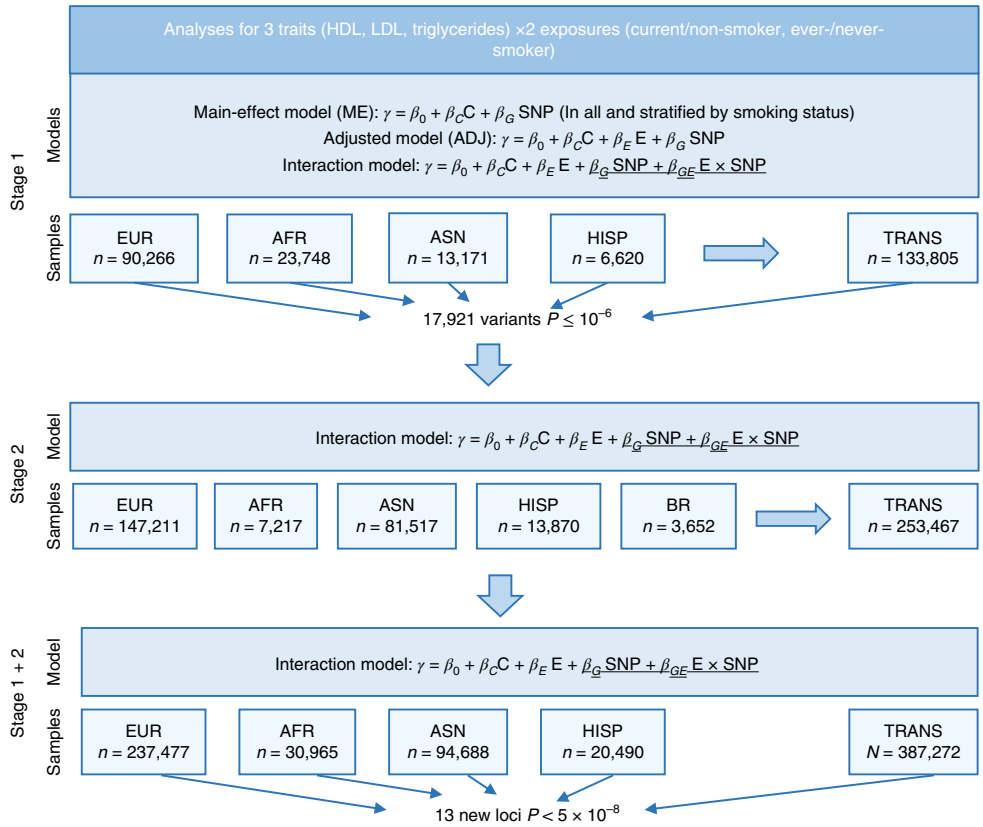


Fig. 1 | Study overview. Summary of data included in this study. Of the 17,921 associated variants from stage 1, 16,389 passed filtering criteria and were included in stage 2 analyses. Trans-ancestry combined stage 1 and 2 meta-analyses were performed on stage 1 trans-ancestry and stage 2 trans-ancestry 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 $P_{2df} < 5\times10^{-8}$) 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 $q < 2.4\times10^{-4}$) 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 ($P_{2df} = 1.4\times10^{-9}$), 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 ($P = 9.6\times10^{-12}$), HDL–ever smoking ($P = 5.9\times10^{-7}$), LDL–current smoking ($P = 8.4\times10^{-15}$), LDL–ever smoking ($P = 3.1\times10^{-5}$), triglycerides–current smoking ($P = 4.0\times10^{-3}$), and triglycerides–ever smoking ($P = 3.1\times10^{-4}$) 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 ($P < 5 \times 10^{-8}/8 = 6.25 \times 10^{-9}$).

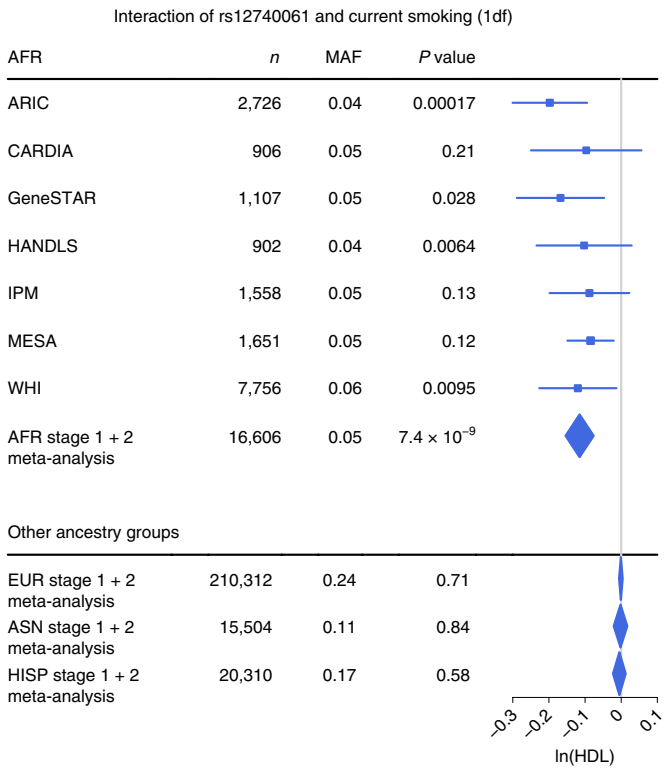


Fig. 2 | Interaction of rs12740061 (*LOC105378783*) and current smoking status (1df). Forest plots show β values (95% confidence intervals) and *P* values (1df) for the rs12740061 × current smoking interaction term in linear regression models of HDL adjusted for age, sex, study-specific covariates (if 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.

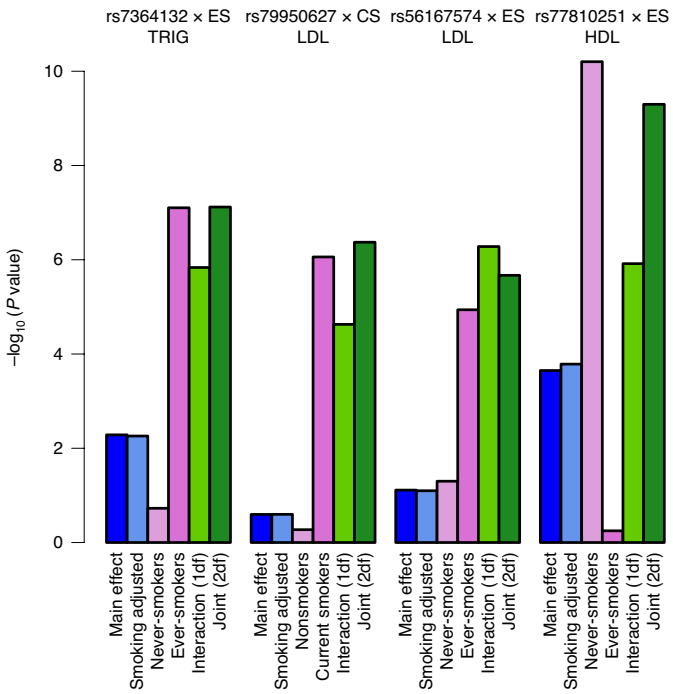


Fig. 3 | Associations observed primarily in one smoking stratum. For select variants for which an association was primarily observed in only one smoking stratum, we compare the *P* values for stage 1 linear association models, including a main-effect model adjusted for age, sex, principal components, 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.

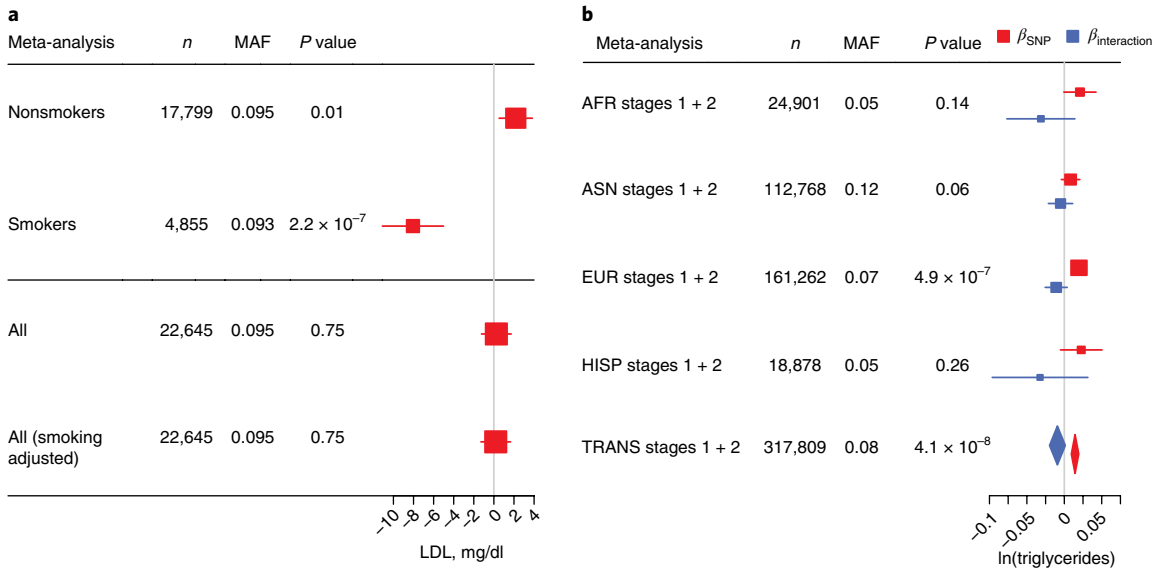


Fig. 4 | Forest plots of select associations. a, Plots showing association between rs73453125 and LDL among AFR individuals in stage 1 (where a series of models was available). Variant β values (95% confidence intervals) and *P* values are drawn from main-effect linear regression models for nonsmokers, 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.

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

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.

nature genetics

SUPPLEMENTARY INFORMATION

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

In the format provided by the authors and unedited.

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

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

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

Supplementary Material

TABLE OF CONTENTS

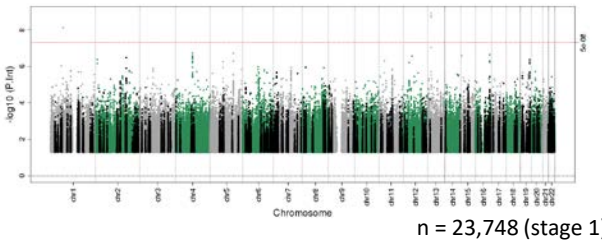
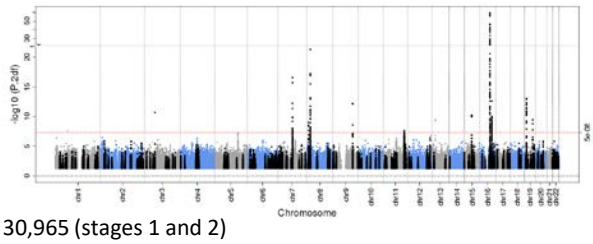
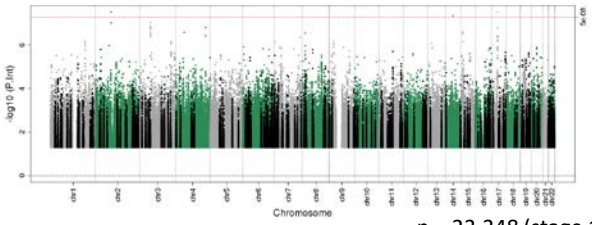
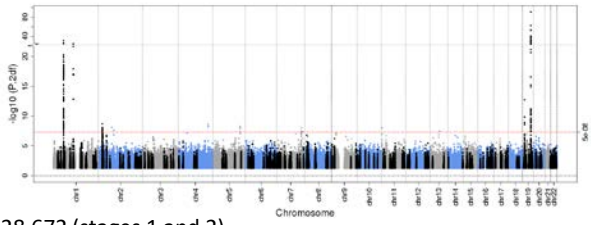
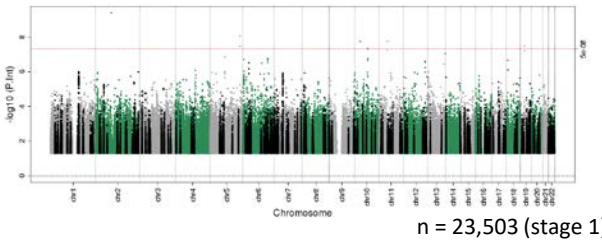
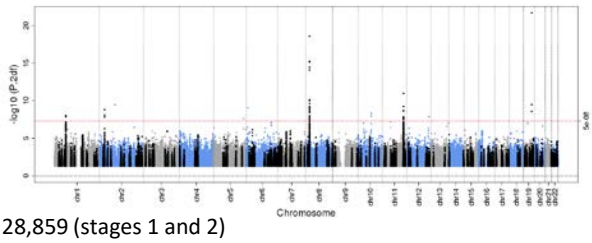
Supplementary Figure 1

Manhattan Plots. Shown are the genome-wide results for linear regression models of lipids traits that include both the variant and a variant × smoking status interaction term, adjusted for age, sex, PCs, study-specific covariates (as necessary) and smoking status. Shown are the p values for the 1df interaction test and the 2df joint test of interaction and main effect. Points shaded in black are within 1 MB of known lipids loci. Plots are drawn from genome-wide stage 1 results with stage 1 and 2 results added where available (i.e. for those variants that were included in follow-up).

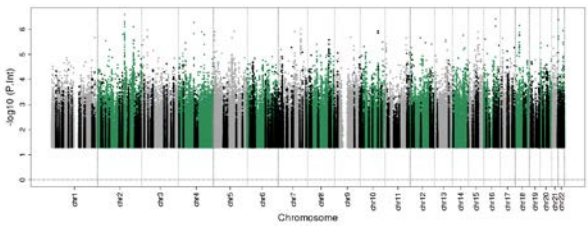
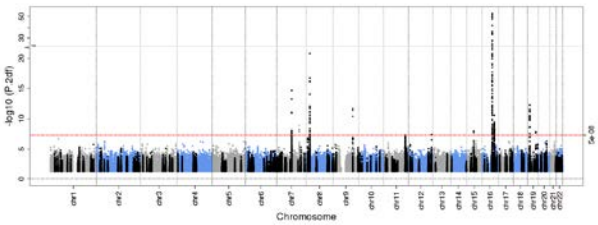
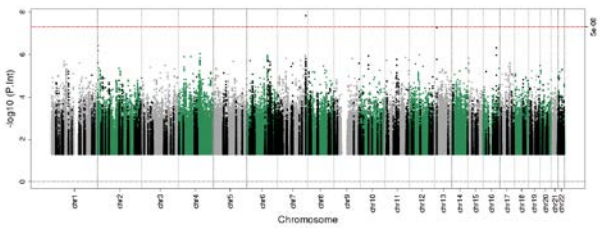
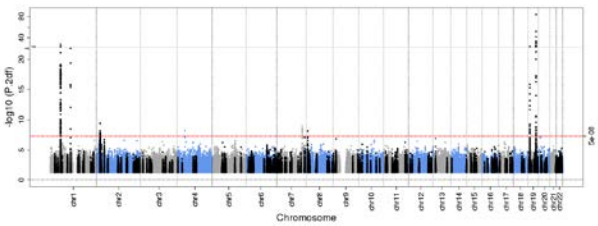
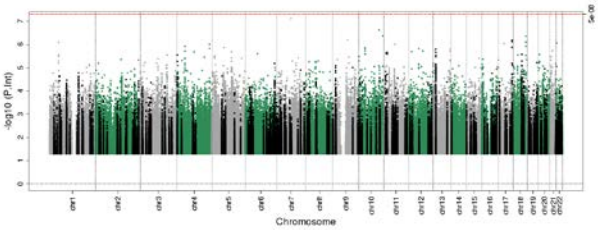
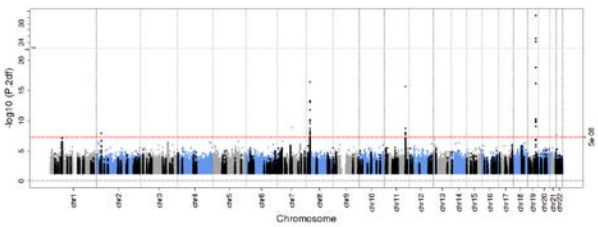
Supplementary Figure 1

African Ancestry

Current Smoking

	1df Test of Interaction	2df Joint Test of Main Effect and Interaction
HDL		
	n = 23,748 (stage 1) and 30,965 (stages 1 and 2)	
LDL		
	n = 23,348 (stage 1) and 28,672 (stages 1 and 2)	
Triglycerides		
	n = 23,503 (stage 1) and 28,859 (stages 1 and 2)	

Ever Smoking

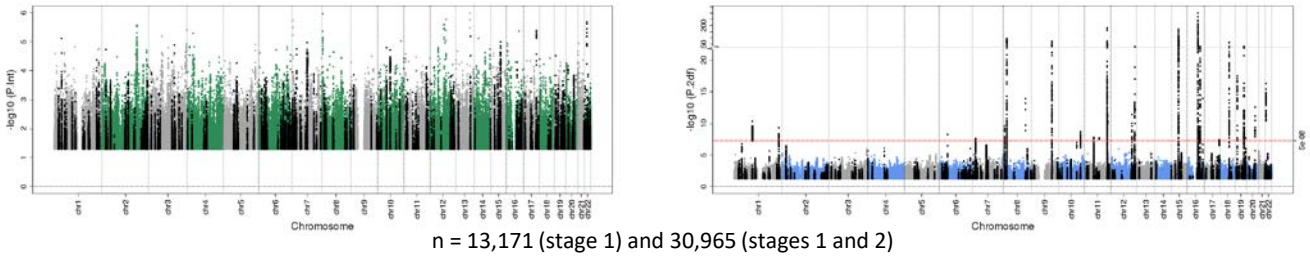
HDL		
	n = 23,753 (stage 1) and 29,365 (stages 1 and 2)	
LDL		
	n = 23,353 (stage 1) and 27,071 (stages 1 and 2)	
Triglycerides		
	n = 23,508 (stage 1) and 27,258 (stages 1 and 2)	

Asian Ancestry

Current Smoking

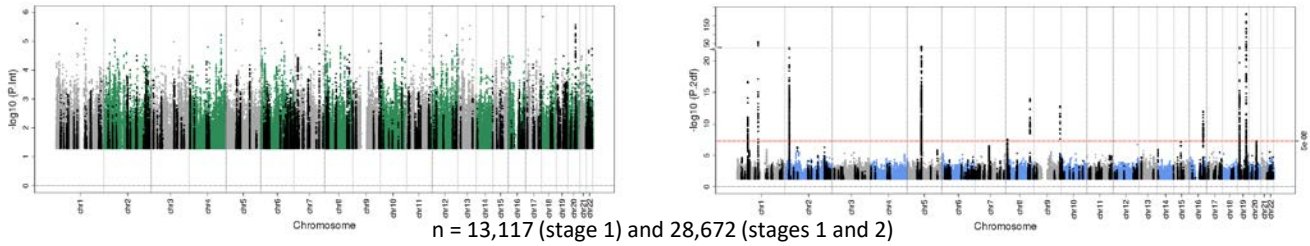
1df Test of Interaction | 2df Joint Test of Main Effect and Interaction

HDL



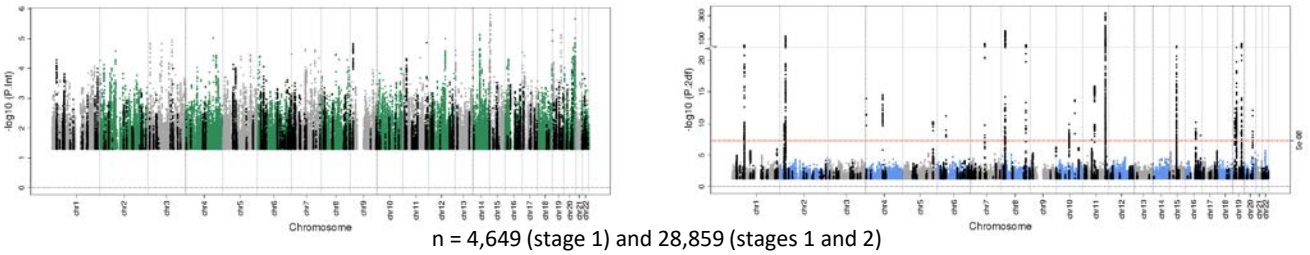
n = 13,171 (stage 1) and 30,965 (stages 1 and 2)

LDL



n = 13,117 (stage 1) and 28,672 (stages 1 and 2)

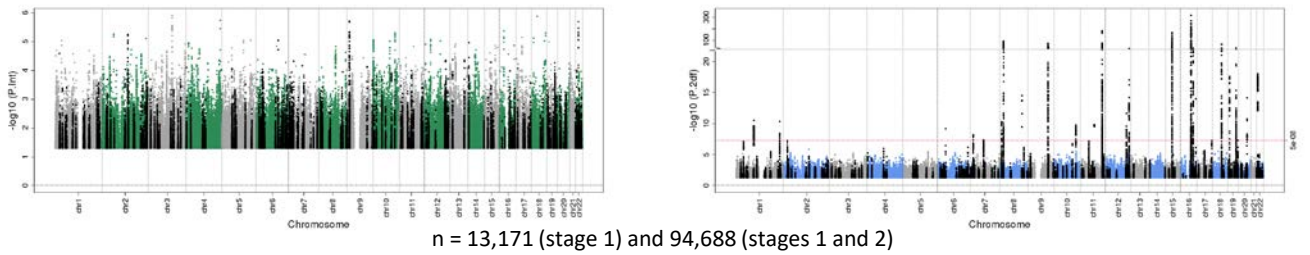
Triglycerides



n = 4,649 (stage 1) and 28,859 (stages 1 and 2)

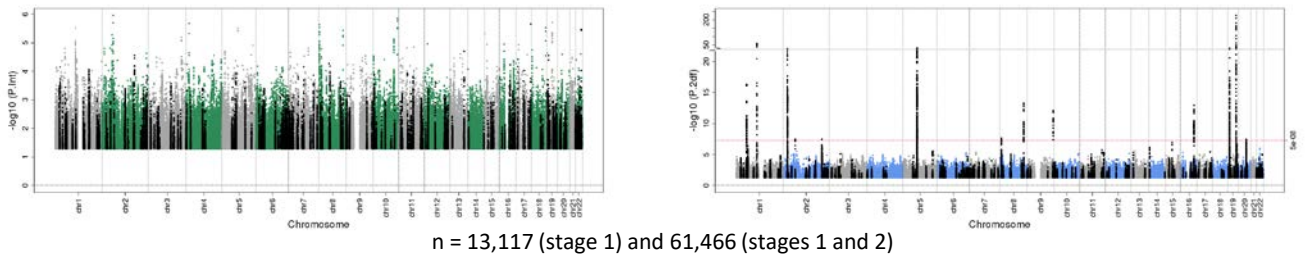
Ever Smoking

HDL



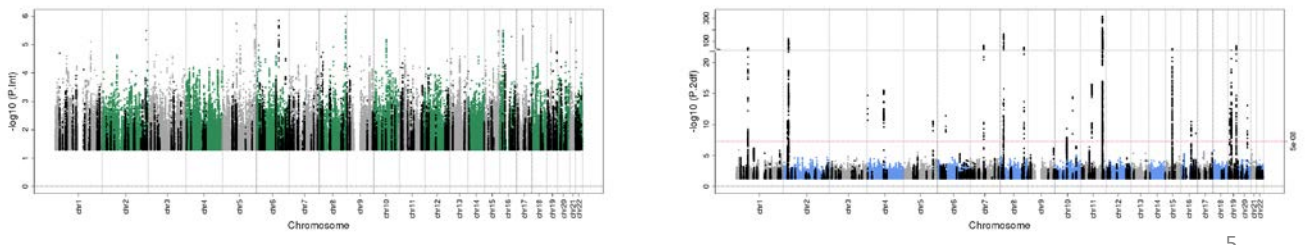
n = 13,171 (stage 1) and 94,688 (stages 1 and 2)

LDL



n = 13,117 (stage 1) and 61,466 (stages 1 and 2)

Triglycerides



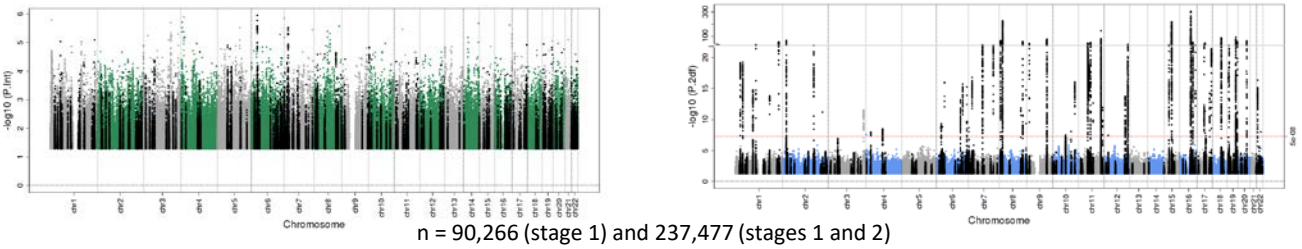
n = 4,649 (stage 1) and 116,570 (stages 1 and 2)

European Ancestry

Current Smoking

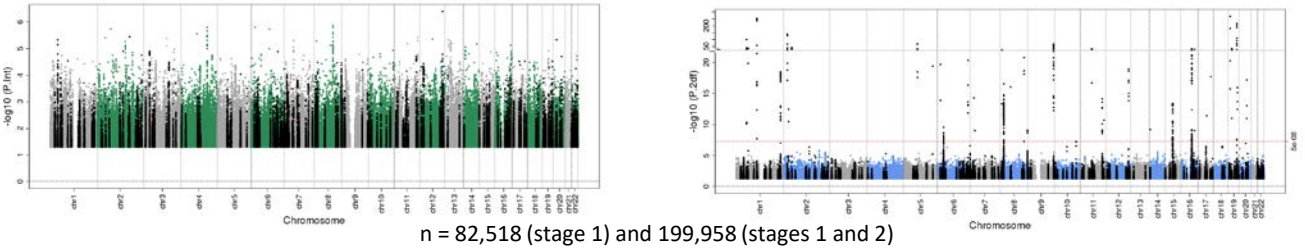
1df Test of Interaction | 2df Joint Test of Main Effect and Interaction

HDL



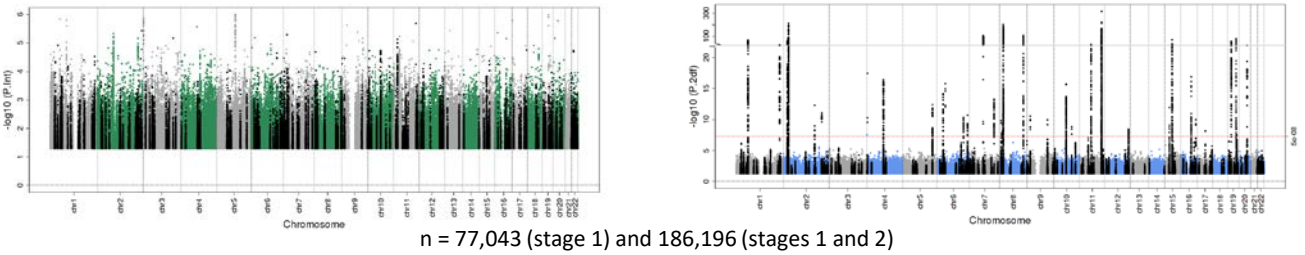
n = 90,266 (stage 1) and 237,477 (stages 1 and 2)

LDL



n = 82,518 (stage 1) and 199,958 (stages 1 and 2)

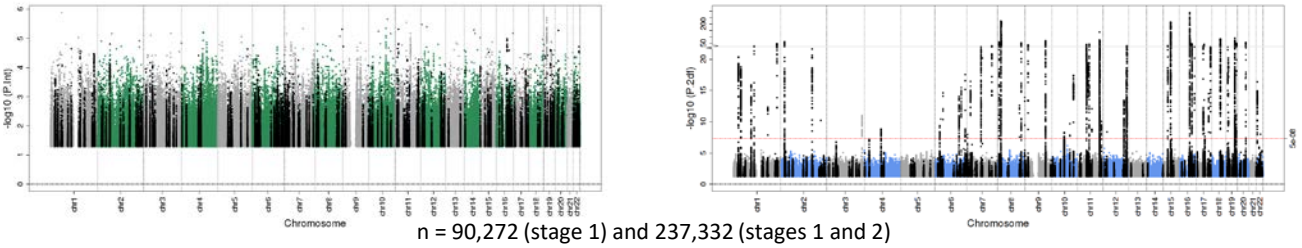
Triglycerides



n = 77,043 (stage 1) and 186,196 (stages 1 and 2)

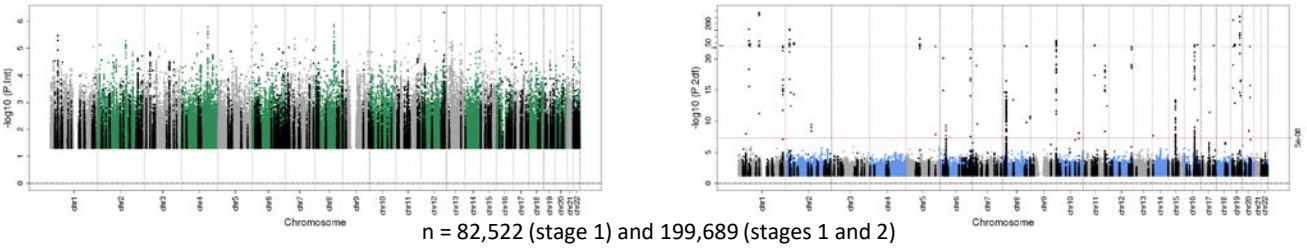
Ever Smoking

HDL



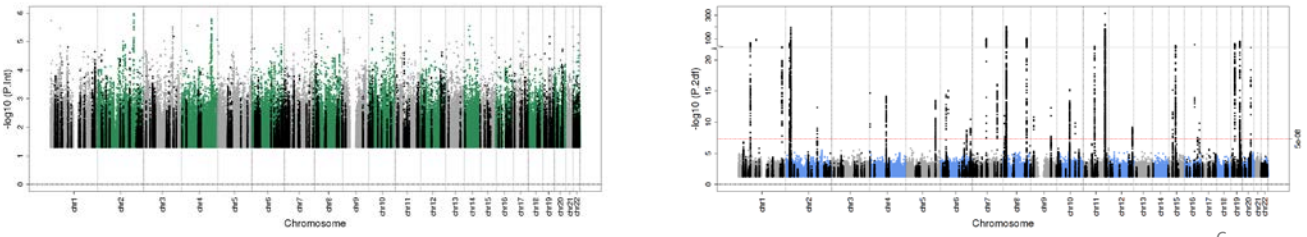
n = 90,272 (stage 1) and 237,332 (stages 1 and 2)

LDL



n = 82,522 (stage 1) and 199,689 (stages 1 and 2)

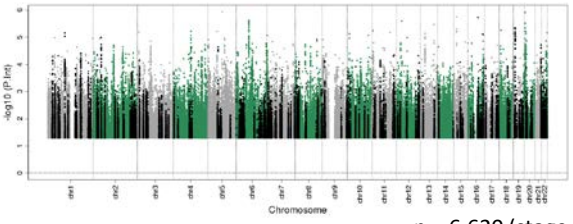
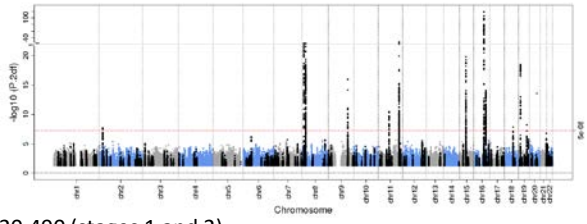
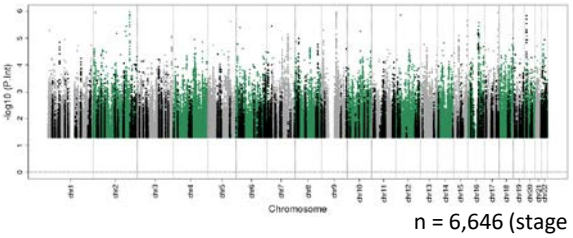
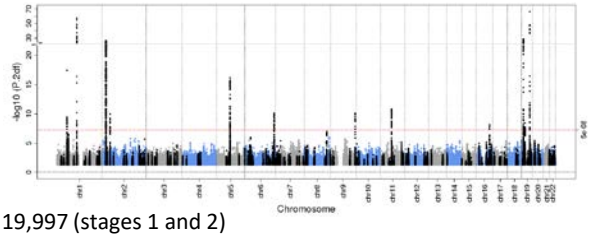
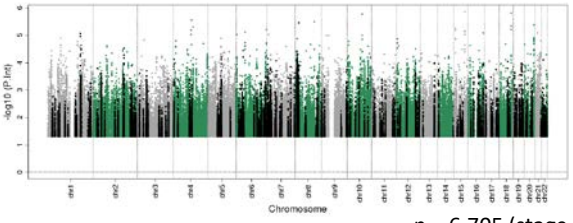
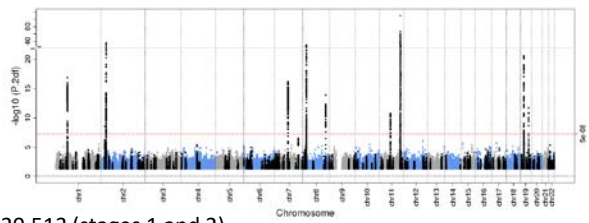
Triglycerides



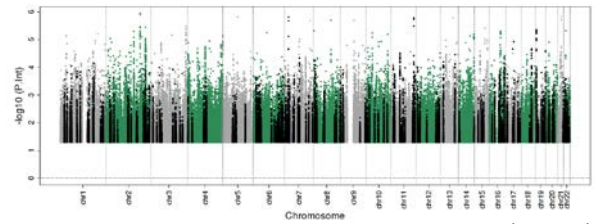
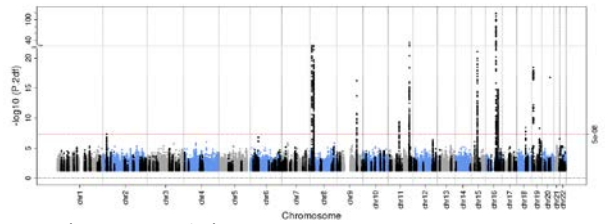
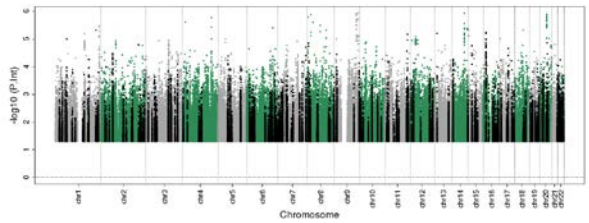
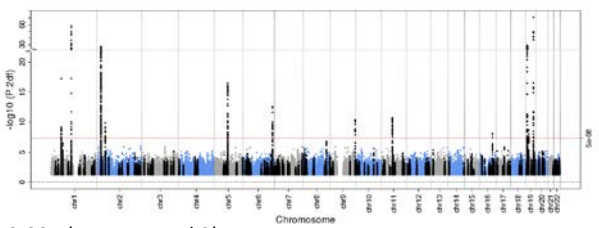
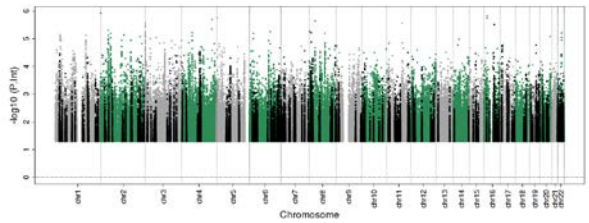
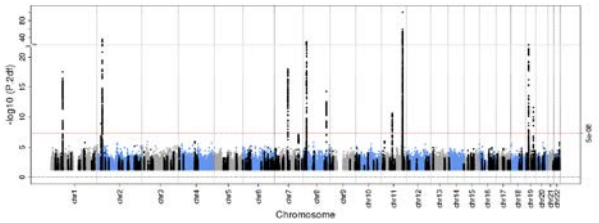
n = 77,047 (stage 1) and 185,638 (stages 1 and 2)

Hispanic Ancestry

Current Smoking

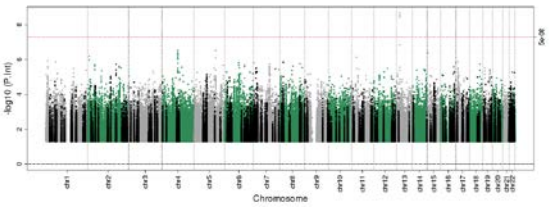
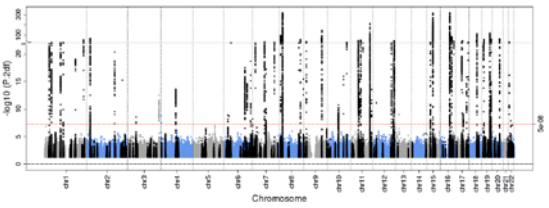
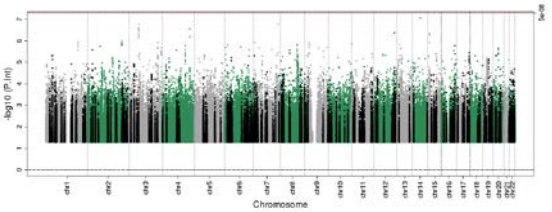
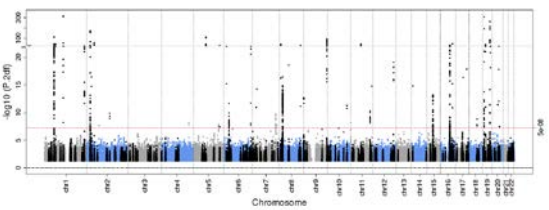
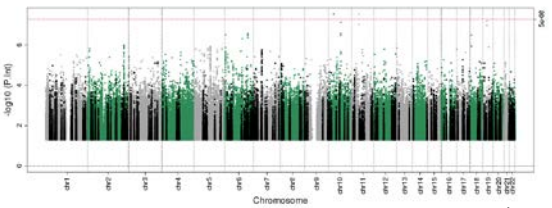
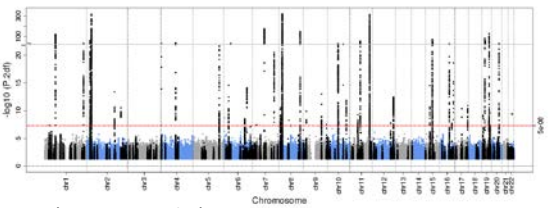
	1df Test of Interaction	2df Joint Test of Main Effect and Interaction
HDL		
	n = 6,620 (stage 1) and 20,490 (stages 1 and 2)	
LDL		
	n = 6,646 (stage 1) and 19,997 (stages 1 and 2)	
Triglycerides		
	n = 6,705 (stage 1) and 20,513 (stages 1 and 2)	

Ever Smoking

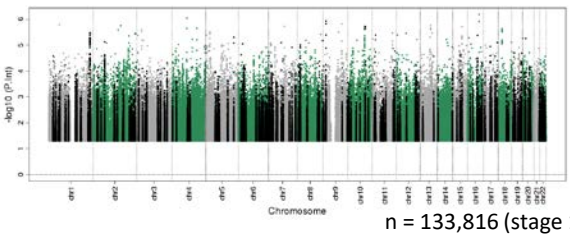
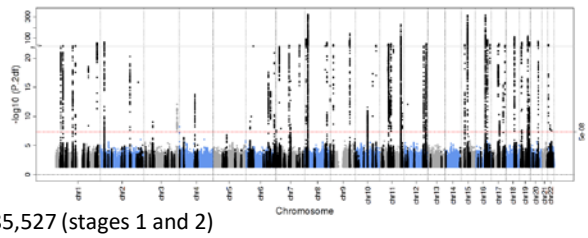
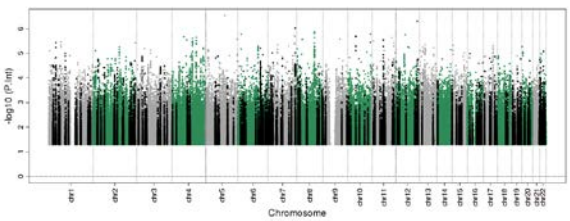
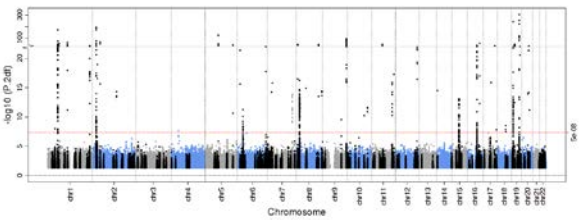
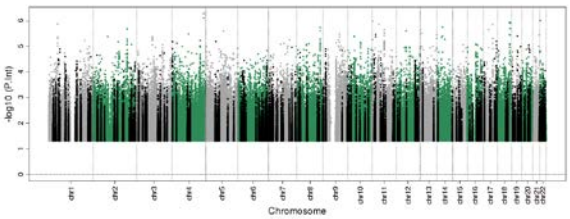
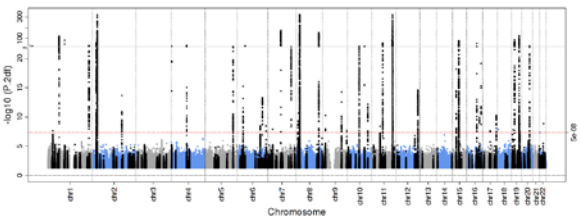
HDL		
	n = 6,620 (stage 1) and 20,490 (stages 1 and 2)	
LDL		
	n = 6,646 (stage 1) and 19,997 (stages 1 and 2)	
Triglycerides		
	n = 6,705 (stage 1) and 20,513 (stages 1 and 2)	

Trans-Ancestry

Current Smoking

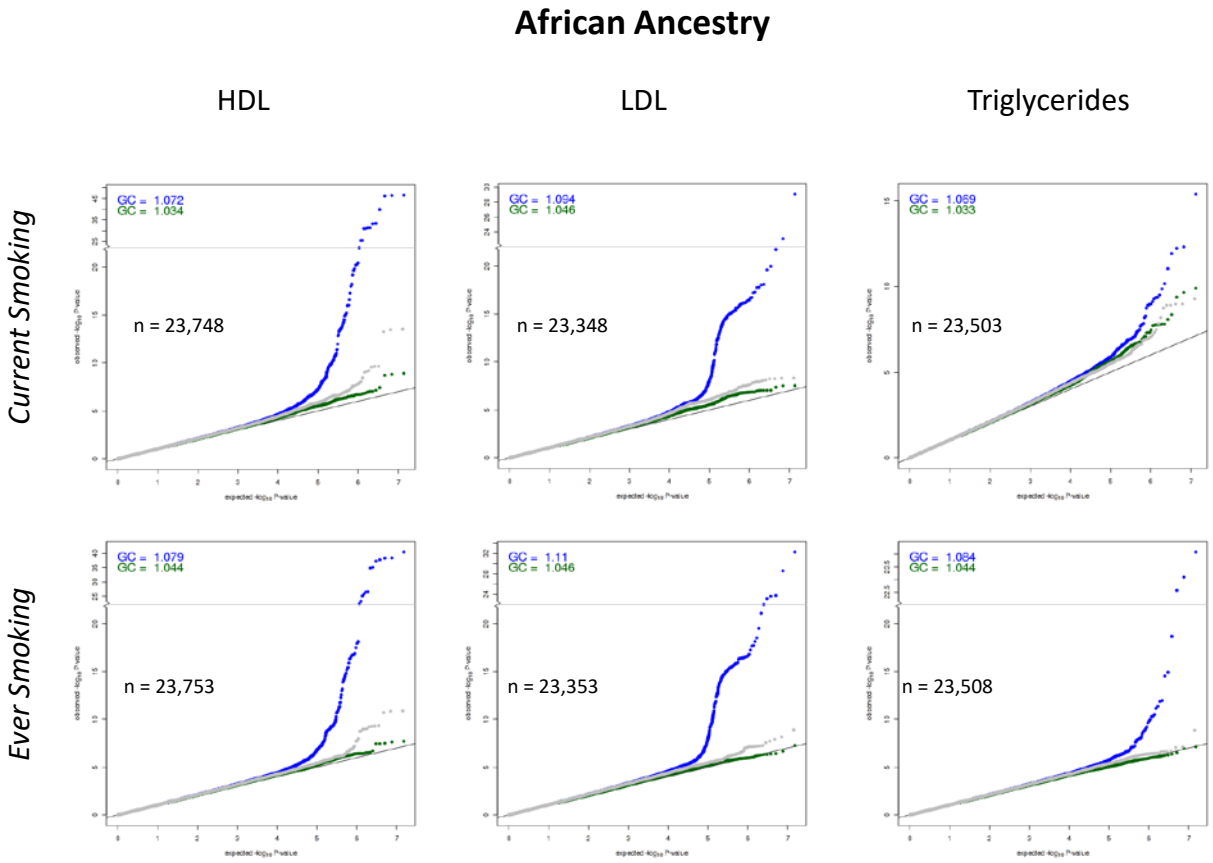
	1df Test of Interaction	2df Joint Test of Main Effect and Interaction
HDL		
	n = 133,805 (stage 1) and 387,272 (stages 1 and 2)	
LDL		
	n = 125,629 (stage 1) and 313,738 (stages 1 and 2)	
Triglycerides		
	n = 111,900 (stage 1) and 353,040 (stages 1 and 2)	

Ever Smoking

HDL		
	n = 133,816 (stage 1) and 385,527 (stages 1 and 2)	
LDL		
	n = 125,638 (stage 1) and 311,868 (stages 1 and 2)	
Triglycerides		
	n = 111,909 (stage 1) and 350,881 (stages 1 and 2)	

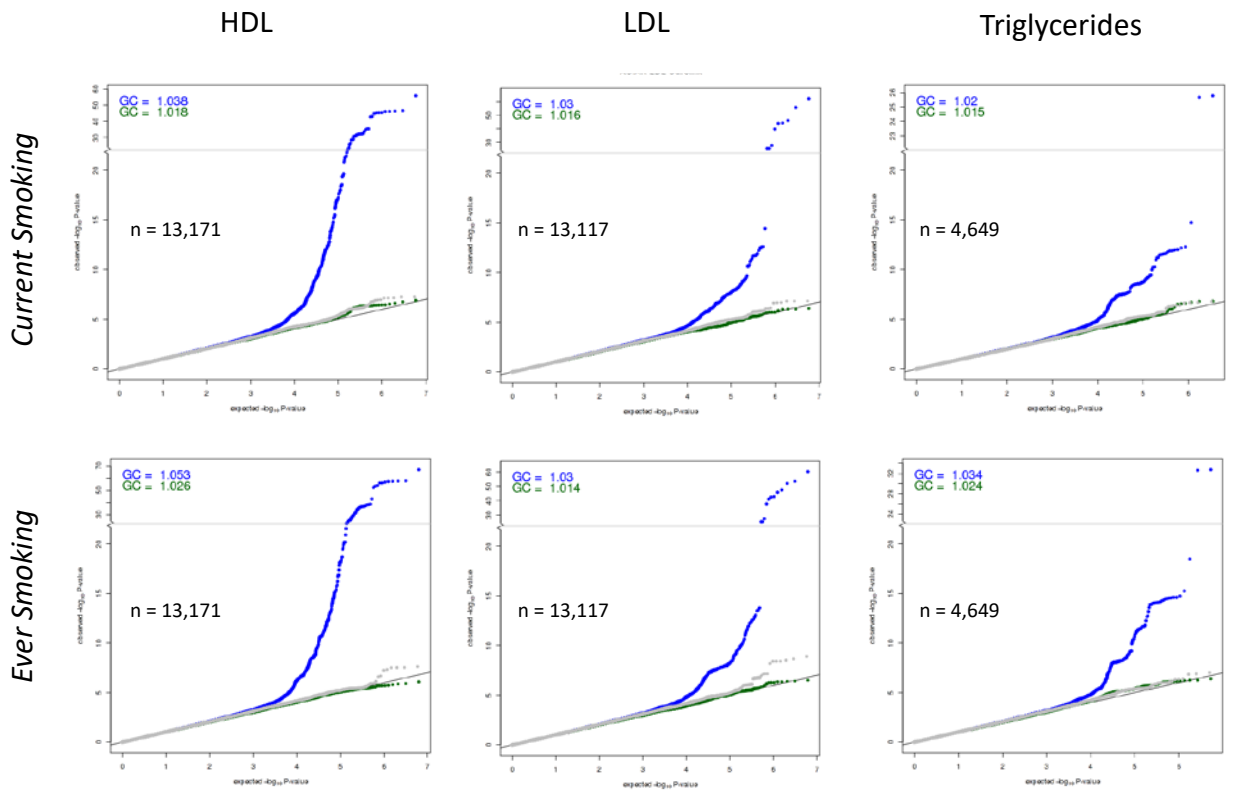
Supplementary Figure 2

QQ Plots. Shown are the p values for the genome-wide stage 1 results for linear regression models of lipids traits that include both the variant and a variant × smoking status interaction term, adjusted for age, sex, PCs, study-specific covariates (as necessary) and smoking status. Shown are the p values for the 2df joint test of interaction and main effect, the 2df joint test after excluding known lipids loci, and the 1df test of interaction. Points shaded in black are within 1 MB of known lipids loci.

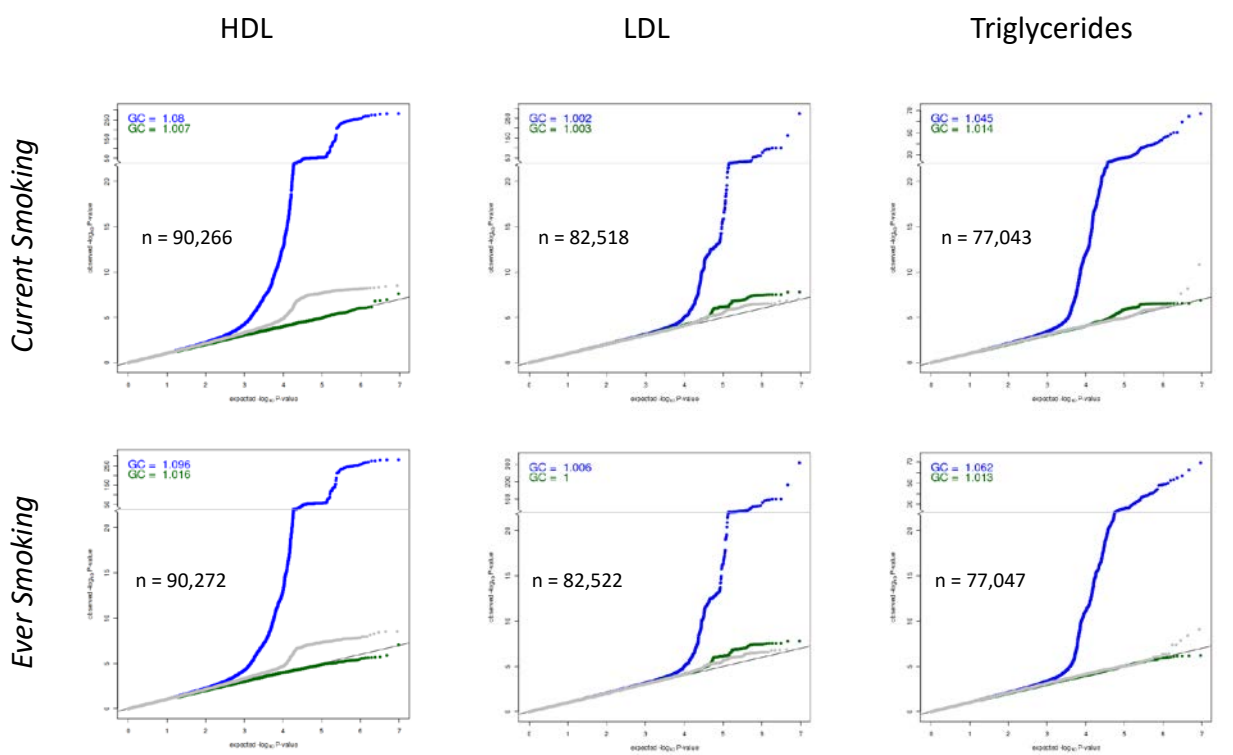


BLUE: 2df test, GRAY: 2df test excluding known loci, GREEN: 1df test

Asian Ancestry

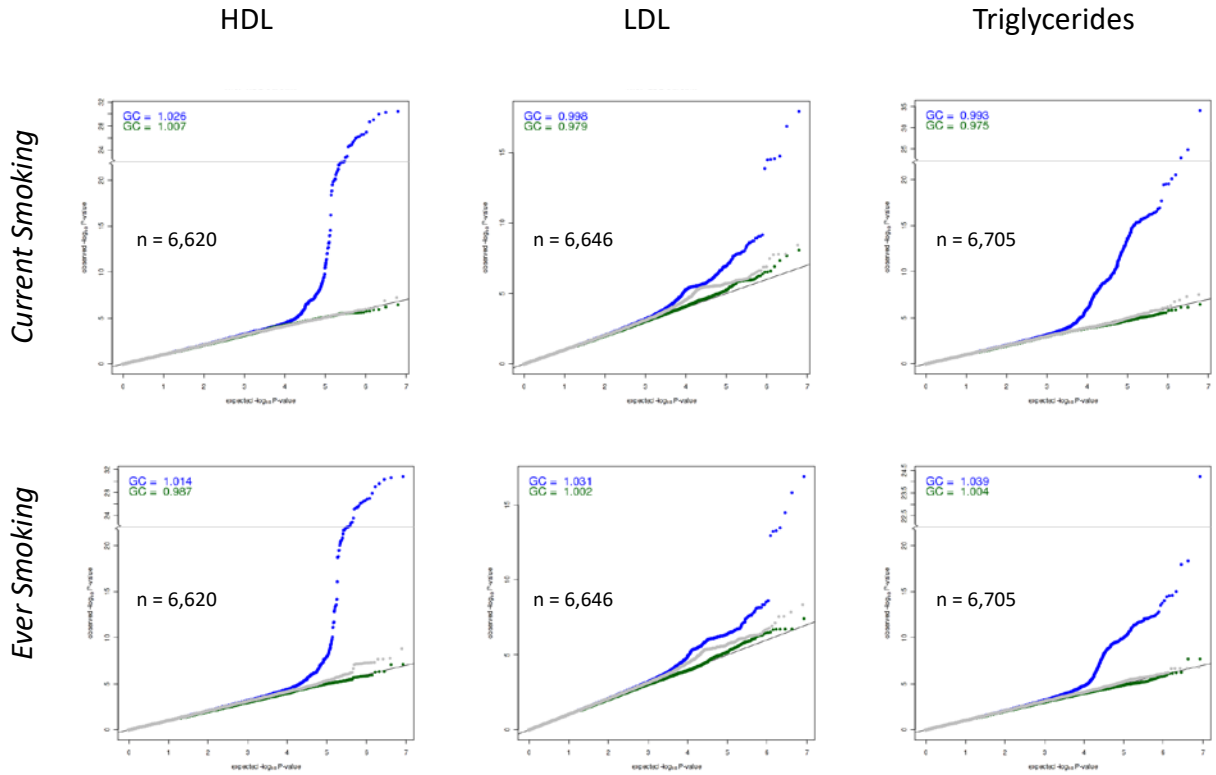


European Ancestry

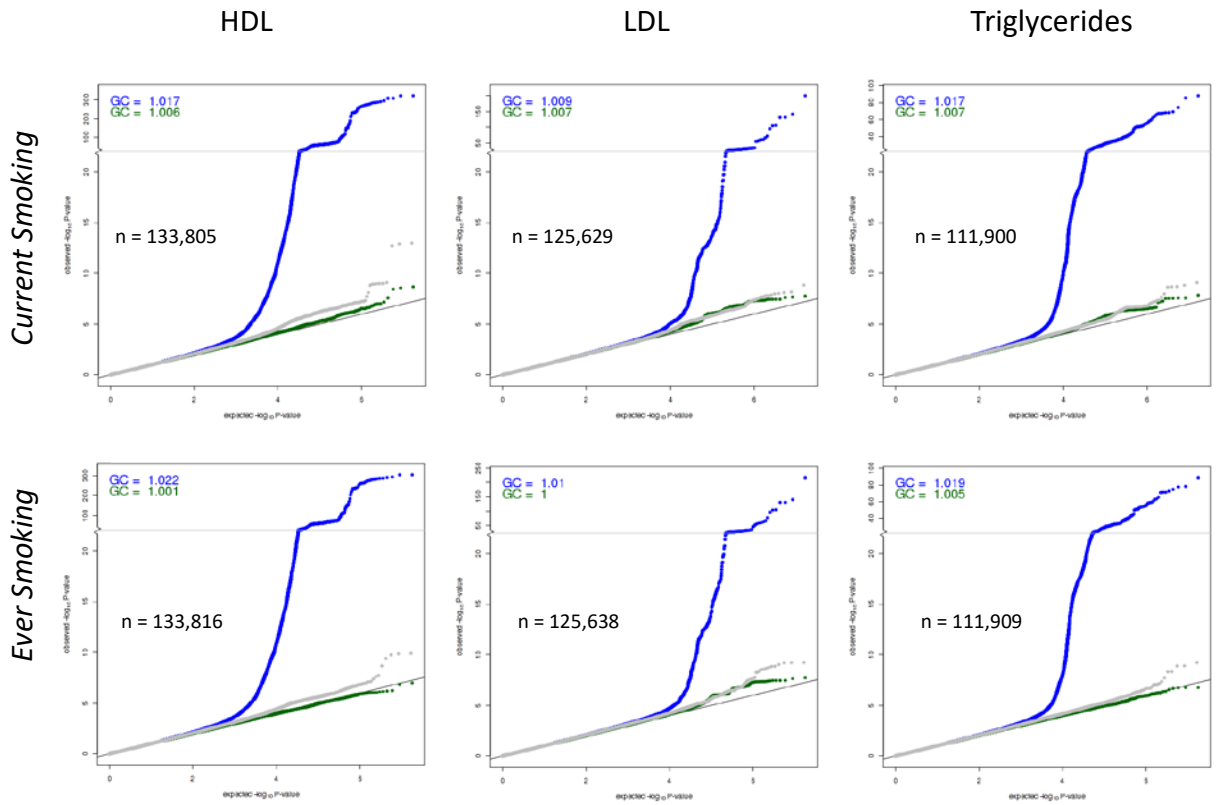


BLUE: 2df test, GRAY: 2df test excluding known loci, GREEN: 1df test

Hispanic Ancestry



Trans-Ancestry

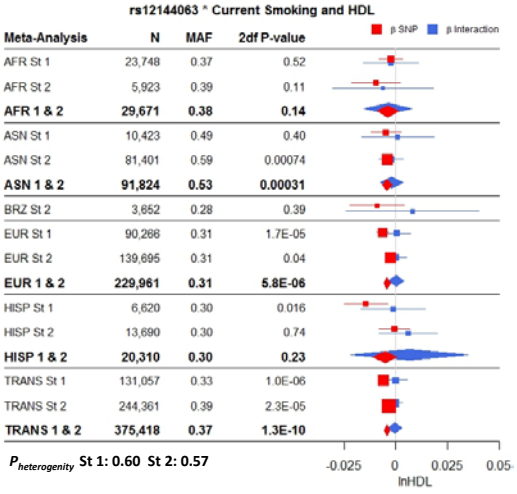


BLUE: 2df test, GRAY: 2df test excluding known loci, GREEN: 1df test

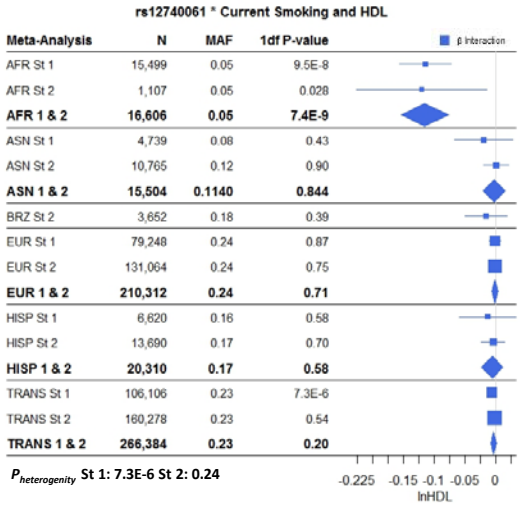
Supplementary Figure 3

Forest Plots of Novel Loci: The results of either the 2df test (β for Interaction and Main Effect shown) or 1df test (only β for Interaction shown), whichever was more statistically significant. 95% confidence intervals included.

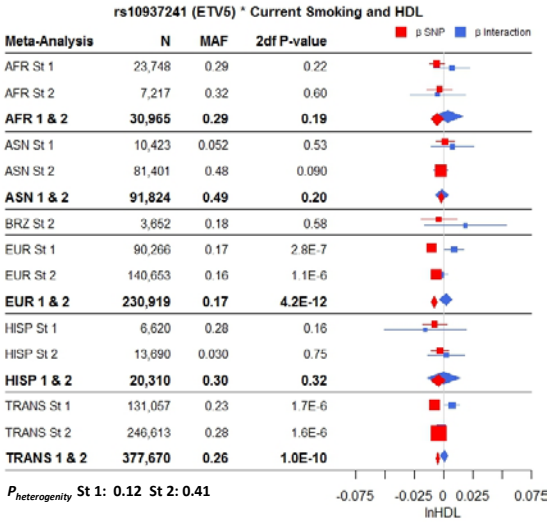
a



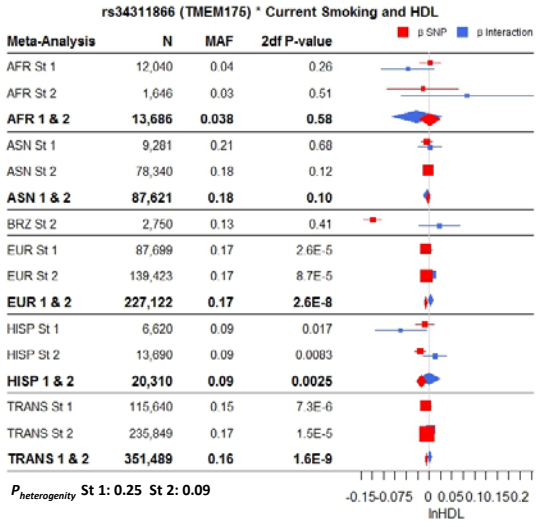
b



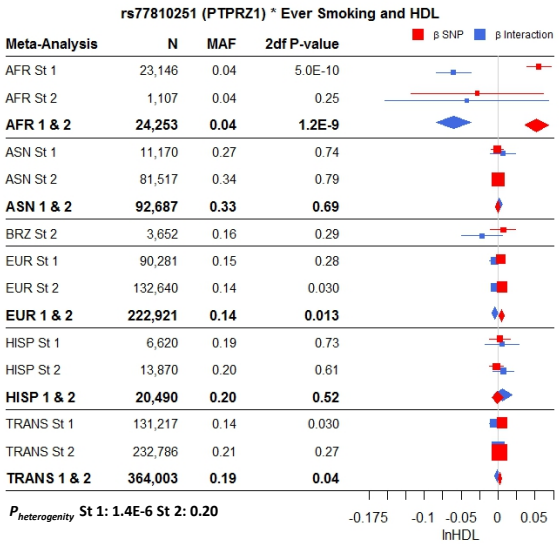
c



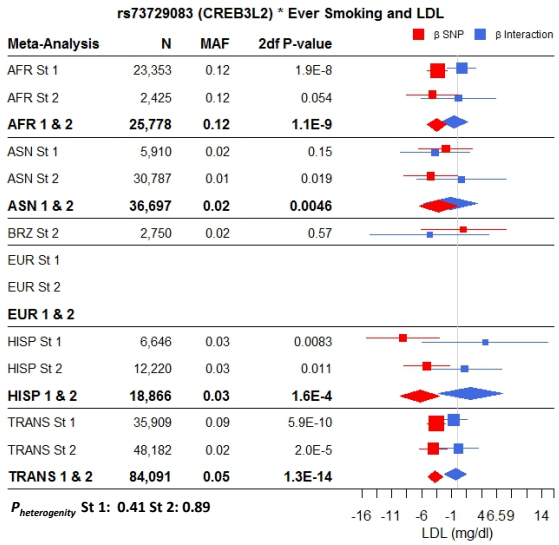
d



e



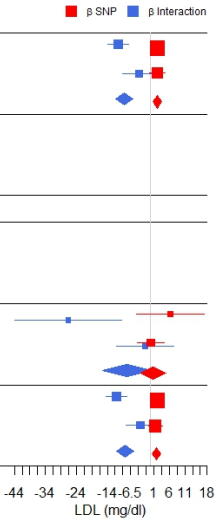
f



g

rs73453125 (CNTNAP2) * Current Smoking and LDL

Meta-Analysis	N	MAF	2df P-value
AFR St 1	22,645	0.09	1.4E-8
AFR St 2	3,675	0.09	0.22
AFR 1 & 2	**26,323**	**0.09**	**4.4E-08**
ASN St 1			
ASN St 2			
ASN 1 & 2			
BRZ St 2			
EUR St 1			
EUR St 2			
EUR 1 & 2			
HISP St 1	2,023	0.03	0.0065
HISP St 2	12,220	0.02	0.91
HISP 1 & 2	**14,243**	**0.02**	**0.15**
TRANS St 1	24,668	0.09	1.4E-9
TRANS St 2	15,595	0.03	0.27
TRANS 1 & 2	**40,566**	**0.07**	**2.0E-08**

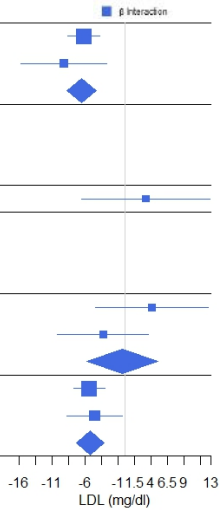


$P_{heterogenity}$ St 1: 0.18 St 2: 0.71

h

rs56167574 * Ever Smoking and LDL

Meta-Analysis	N	MAF	1df P-value
AFR St 1	23,353	0.12	5.3E-7
AFR St 2	2,425	0.11	0.0056
AFR 1 & 2	**25,778**	**0.12**	**1.5E-8**
ASN St 1			
ASN St 2			
ASN 1 & 2			
BRZ St 2	2,750	0.03	0.51
EUR St 1			
EUR St 2			
EUR 1 & 2			
HISP St 1	6,646	0.03	0.34
HISP St 2	12,220	0.02	0.36
HISP 1 & 2	**18,866**	**0.02**	**0.92**
TRANS St 1	29,999	0.11	8.5E-6
TRANS St 2	17,395	0.06	0.037
TRANS 1 & 2	**47,394**	**0.10**	**9.3E-7**

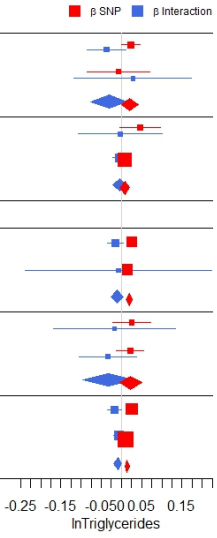


$P_{heterogenity}$ St 1: 0.02 St 2: 0.10

i

rs10101067 * Current Smoking and Triglycerides

Meta-Analysis	N	MAF	2df P-value
AFR St 1	22,188	0.05	0.10
AFR St 2	2,713	0.04	0.93
AFR 1 & 2	**24,901**	**0.05**	**0.14**
ASN St 1	2,761	0.11	0.13
ASN St 2	110,007	0.12	0.11
ASN 1 & 2	**112,768**	**0.12**	**0.062**
BRZ St 2			
EUR St 1	72,064	0.07	1.6E-05
EUR St 2	89,198	0.07	0.012
EUR 1 & 2	**161,262**	**0.07**	**4.9E-07**
HISP St 1	5,250	0.05	0.57
HISP St 2	13,625	0.05	0.44
HISP 1 & 2	**18,878**	**0.05**	**0.26**
TRANS St 1	102,263	0.07	5.7E-7
TRANS St 2	215,546	0.09	0.0018
TRANS 1 & 2	**317,809**	**0.08**	**4.1E-08**

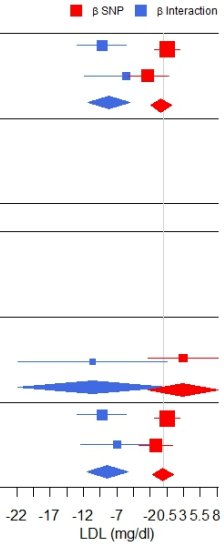


$P_{heterogenity}$ St 1: 0.89 St 2: 0.88

j

rs79950627 * Current Smoking and LDL

Meta-Analysis	N	MAF	2df P-value
AFR St 1	23,348	0.07	4.3E-7
AFR St 2	2,704	0.08	0.0056
AFR 1 & 2	**26,052**	**0.07**	**9.6E-9**
ASN St 1			
ASN St 2			
ASN 1 & 2			
BRZ St 2			
EUR St 1			
EUR St 2			
EUR 1 & 2			
HISP St 1			
HISP St 2	12,220	0.01	0.17
HISP 1 & 2	**12,220**	**0.01**	**0.17**
TRANS St 1	23,348	0.07	4.3E-7
TRANS St 2	14,924	0.02	0.0044
TRANS 1 & 2	**38,272**	**0.05**	**7.2E-9**

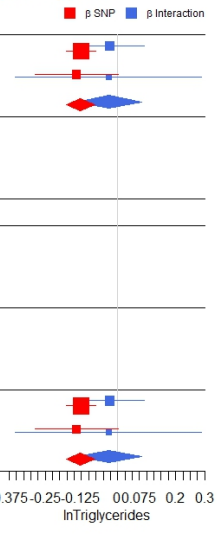


$P_{heterogenity}$ St 1: -- St 2: 0.22

k

rs4758675 (B3GNT4) * Current Smoking and Triglycerides

Meta-Analysis	N	MAF	2df P-value
AFR St 1	11,875	0.02	2.4E-7
AFR St 2	1,107	0.03	0.053
AFR 1 & 2	**12,982**	**0.02**	**1.3E-8**
ASN St 1			
ASN St 2			
ASN 1 & 2			
BRZ St 2			
EUR St 1			
EUR St 2			
EUR 1 & 2			
HISP St 1			
HISP St 2			
HISP 1 & 2			
TRANS St 1	21,875	0.02	2.4E-7
TRANS St 2	1,107	0.03	0.053
TRANS 1 & 2	**12,982**	**0.02**	**1.3E-8**

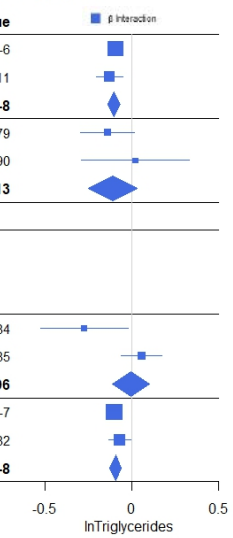


$P_{heterogenity}$ St 1: -- St 2: --

l

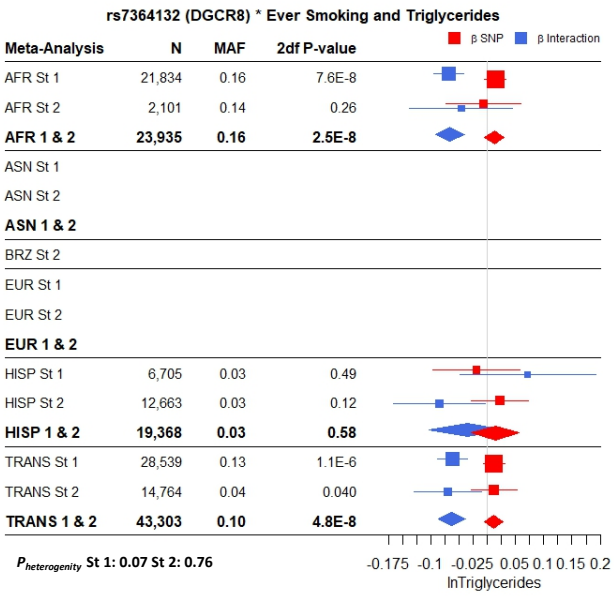
rs60029395 * Current Smoking and Triglycerides

Meta-Analysis	N	MAF	1df P-value
AFR St 1	15,747	0.12	5.4E-6
AFR St 2	3,301	0.16	0.0011
AFR 1 & 2	**19,048**	**0.13**	**3.3E-8**
ASN St 1	1,813	0.03	0.079
ASN St 2	1,417	ASN St 2	0.90
ASN 1 & 2	**3,230**	**0.03**	**0.13**
BRZ St 2			
EUR St 1			
EUR St 2			
EUR 1 & 2			
HISP St 1	2,046	0.04	0.034
HISP St 2	12,663	0.02	0.35
HISP 1 & 2	**14,709**	**0.02**	**0.96**
TRANS St 1	19,606	0.12	4.8E-7
TRANS St 2	17,381	0.12	0.032
TRANS 1 & 2	**36,987**	**0.12**	**6.6E-8**



$P_{heterogenity}$ St 1: 0.34 St 2: 0.03

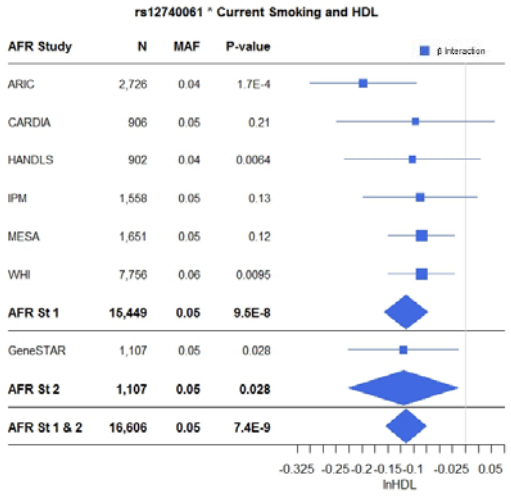
m


rs7364132 (DGCR8) * Ever Smoking and Triglycerides
β SNP
β Interaction
Meta-Analysis N MAF 2df P-value
AFR St 1 21,834 0.16 7.6E-8
AFR St 2 2,101 0.14 0.26
AFR 1 & 2 23,935 0.16 2.5E-8
ASN St 1
ASN St 2
ASN 1 & 2
BRZ St 2
EUR St 1
EUR St 2
EUR 1 & 2
HISP St 1 6,705 0.03 0.49
HISP St 2 12,663 0.03 0.12
HISP 1 & 2 19,368 0.03 0.58
TRANS St 1 28,539 0.13 1.1E-6
TRANS St 2 14,764 0.04 0.040
TRANS 1 & 2 43,303 0.10 4.8E-8
P heterogenity St 1: 0.07 St 2: 0.76
-0.175 -0.1 -0.025 0.05 0.1 0.15 0.2
lnTriglycerides

Supplementary Figure 4

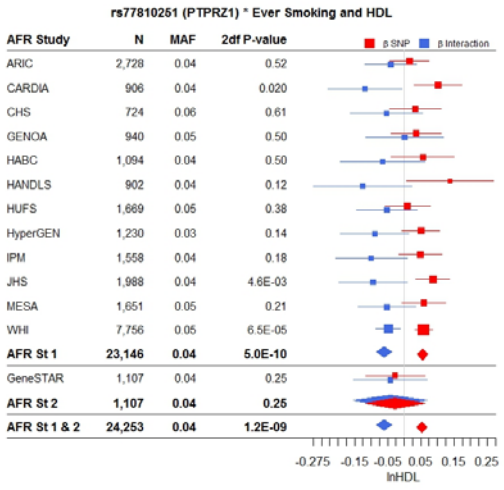
Within Ancestry Forest Plots of Novel Loci: For those associations that were only observed within one ancestry, results for all cohorts in the ancestry in which the association was observed. The results of either the 2df test (β for Interaction and Main Effect shown) or 1df test (only β for Interaction shown), whichever was more statistically significant. 95% confidence intervals included.

a



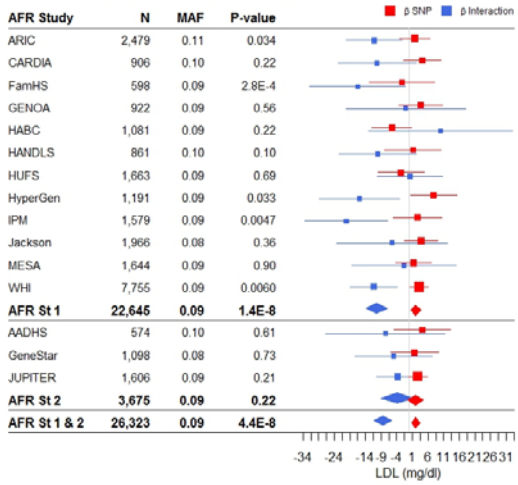
AFR Study	N	MAF	P-value
ARIC	2,726	0.04	1.7E-4
CARDIA	906	0.05	0.21
HANDLS	902	0.04	0.0064
IPM	1,558	0.05	0.13
MESA	1,651	0.05	0.12
WHI	7,756	0.06	0.0095
AFR St 1	**15,449**	**0.05**	**9.5E-8**
GeneSTAR	1,107	0.05	0.028
AFR St 2	**1,107**	**0.05**	**0.028**
AFR St 1 & 2	**16,606**	**0.05**	**7.4E-9**

b



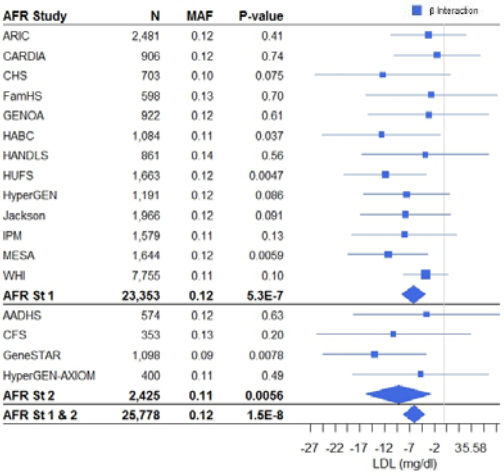
AFR Study	N	MAF	2df P-value
ARIC	2,728	0.04	0.52
CARDIA	906	0.04	0.020
CHS	724	0.06	0.61
GENOA	940	0.05	0.50
HABC	1,094	0.04	0.50
HANDLS	902	0.04	0.12
HUFS	1,669	0.05	0.38
HyperGEN	1,230	0.03	0.14
IPM	1,558	0.04	0.18
JHS	1,988	0.04	4.6E-03
MESA	1,651	0.05	0.21
WHI	7,756	0.05	6.5E-05
AFR St 1	**23,146**	**0.04**	**5.0E-10**
GeneSTAR	1,107	0.04	0.25
AFR St 2	**1,107**	**0.04**	**0.25**
AFR St 1 & 2	**24,253**	**0.04**	**1.2E-09**

c



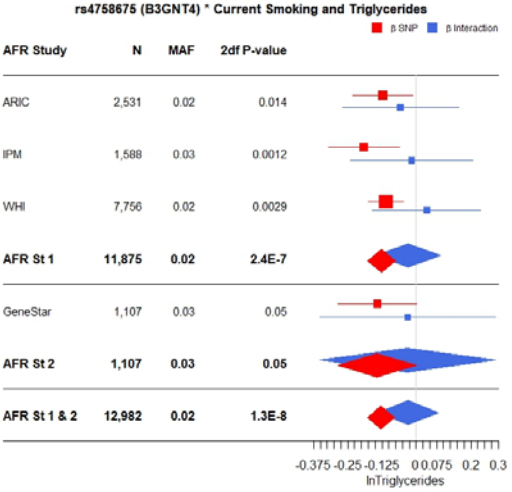
AFR Study	N	MAF	P-value
ARIC	2,479	0.11	0.034
CARDIA	906	0.10	0.22
FamHS	598	0.09	2.8E-4
GENOA	922	0.09	0.56
HABC	1,081	0.09	0.22
HANDLS	861	0.10	0.10
HUFS	1,663	0.09	0.69
HyperGen	1,191	0.09	0.033
IPM	1,579	0.09	0.0047
Jackson	1,966	0.08	0.36
MESA	1,644	0.09	0.90
WHI	7,755	0.09	0.0060
AFR St 1	**22,645**	**0.09**	**1.4E-8**
AADHS	574	0.10	0.61
GeneStar	1,098	0.08	0.73
JUPITER	1,606	0.09	0.21
AFR St 2	**3,675**	**0.09**	**0.22**
AFR St 1 & 2	**26,323**	**0.09**	**4.4E-8**

d



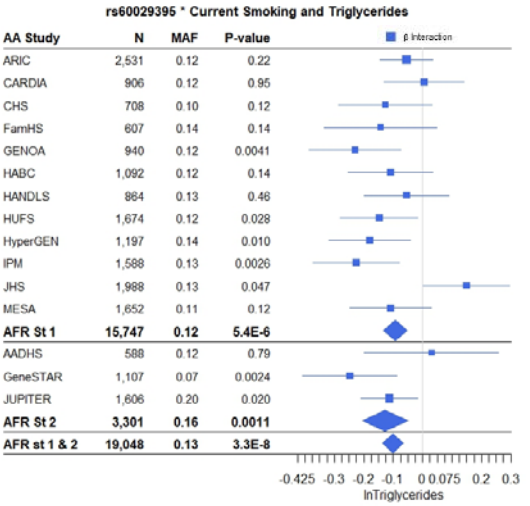
AFR Study	N	MAF	P-value
ARIC	2,481	0.12	0.41
CARDIA	906	0.12	0.74
CHS	703	0.10	0.075
FamHS	598	0.13	0.70
GENOA	922	0.12	0.61
HABC	1,084	0.11	0.037
HANDLS	861	0.14	0.56
HUFS	1,663	0.12	0.0047
HyperGEN	1,191	0.12	0.086
Jackson	1,966	0.12	0.091
IPM	1,579	0.11	0.13
MESA	1,644	0.12	0.0059
WHI	7,755	0.11	0.10
AFR St 1	**23,353**	**0.12**	**5.3E-7**
AADHS	574	0.12	0.63
CFS	353	0.13	0.20
GeneSTAR	1,098	0.09	0.0078
HyperGEN-AXIOM	400	0.11	0.49
AFR St 2	**2,425**	**0.11**	**0.0056**
AFR St 1 & 2	**25,778**	**0.12**	**1.5E-8**

e



AFR Study	N	MAF	2df P-value
ARIC	2,531	0.02	0.014
IPM	1,588	0.03	0.0012
WHI	7,756	0.02	0.0029
AFR St 1	**11,875**	**0.02**	**2.4E-7**
GeneStar	1,107	0.03	0.05
AFR St 2	**1,107**	**0.03**	**0.05**
AFR St 1 & 2	**12,982**	**0.02**	**1.3E-8**

f

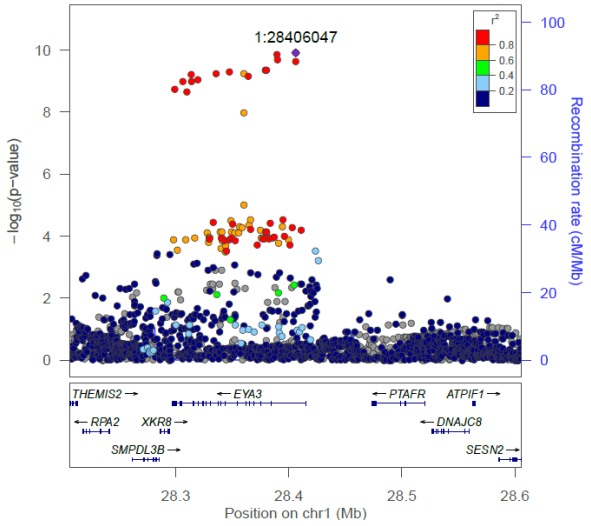


AA Study	N	MAF	P-value
ARIC	2,531	0.12	0.22
CARDIA	906	0.12	0.95
CHS	708	0.10	0.12
FamHS	607	0.14	0.14
GENOA	940	0.12	0.0041
HABC	1,092	0.12	0.14
HANDLS	864	0.13	0.46
HUFS	1,674	0.12	0.028
HyperGEN	1,197	0.14	0.010
IPM	1,588	0.13	0.0026
JHS	1,988	0.13	0.047
MESA	1,652	0.11	0.12
AFR St 1	**15,747**	**0.12**	**5.4E-6**
AADHS	588	0.12	0.79
GeneSTAR	1,107	0.07	0.0024
JUPITER	1,606	0.20	0.020
AFR St 2	**3,301**	**0.16**	**0.0011**
AFR st 1 & 2	**19,048**	**0.13**	**3.3E-8**

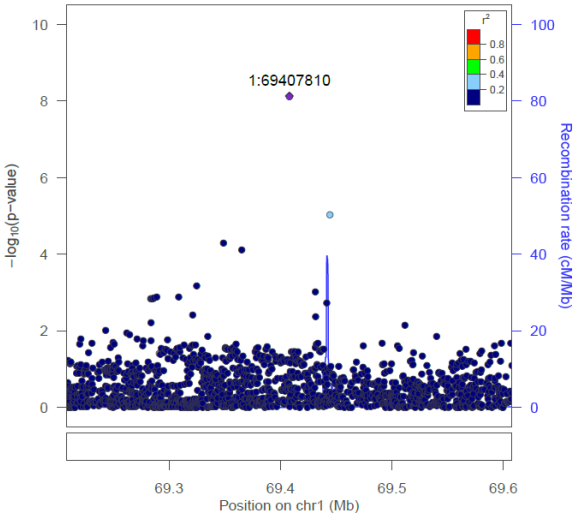
Supplementary Figure 5

Regional Association Plots for Novel Loci: LocusZoom plots showing p-values for stage 1 and 2 combined meta-analyses or stage 1 meta-analyses (if variant not carried forward to stage 2).

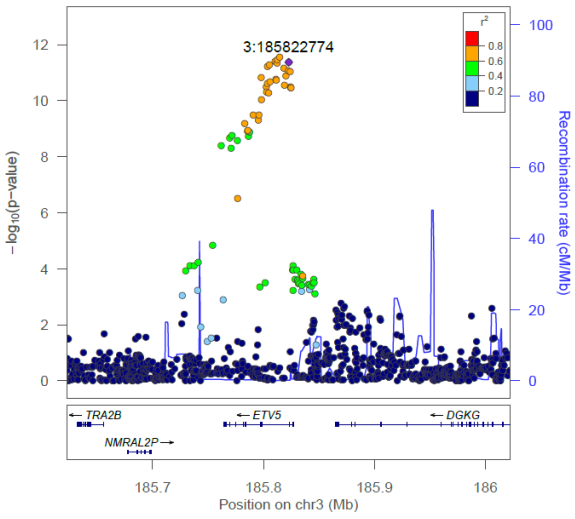
rs12144063 (*EYA3*) × Current Smoking and HDL
TRANS ancestry, 2df tests (n = 375,418)



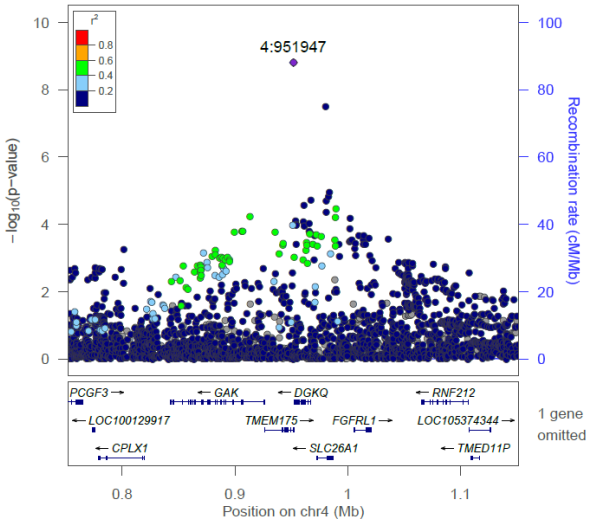
rs12740061 (*LOC105378783*) × Current Smoking and HDL
AFR ancestry, 1df tests (n = 16,606)



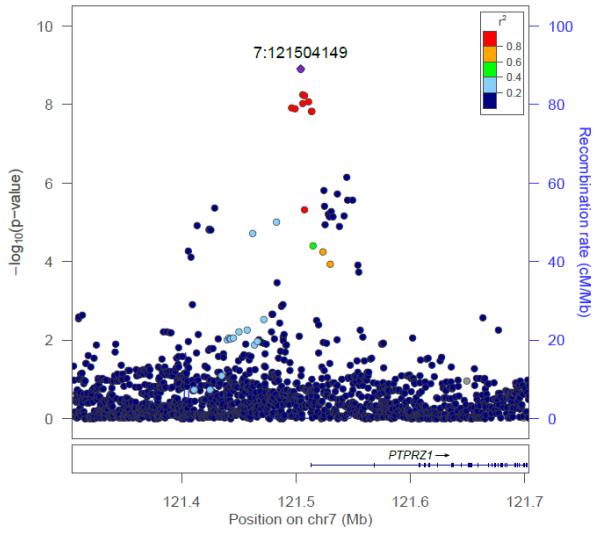
rs10937241 (*ETV5*) × Current Smoking and HDL
EUR ancestry, 2df tests (n = 230,919)



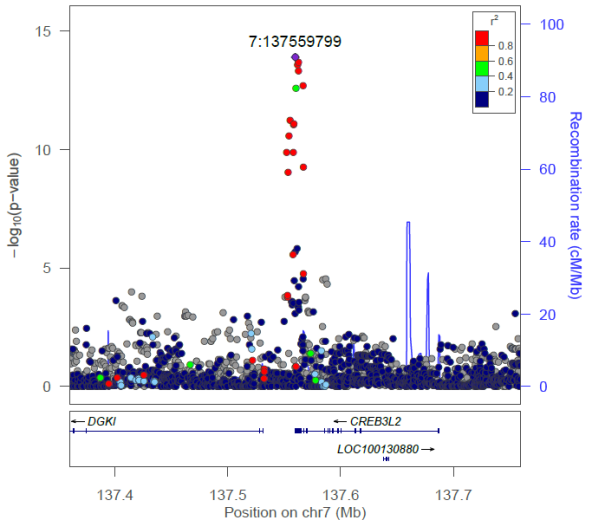
rs34311866 (*TMEM175*) × Current Smoking and HDL
TRANS ancestry, 2df tests (n = 351,489)



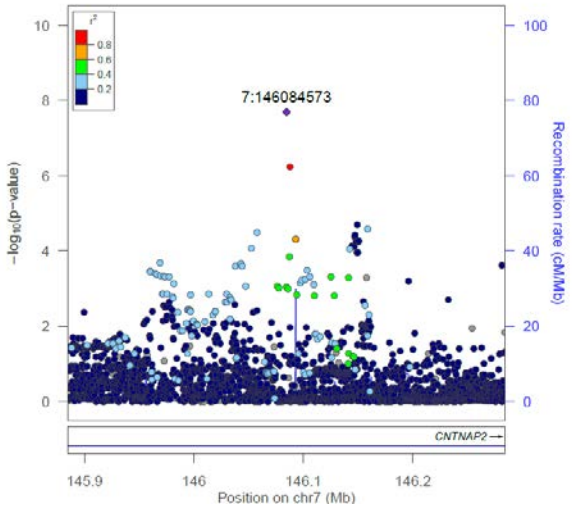
rs77810251 (*PTPRZ1*) × Ever Smoking and HDL
AFR ancestry, 2df tests (n = 24,253)



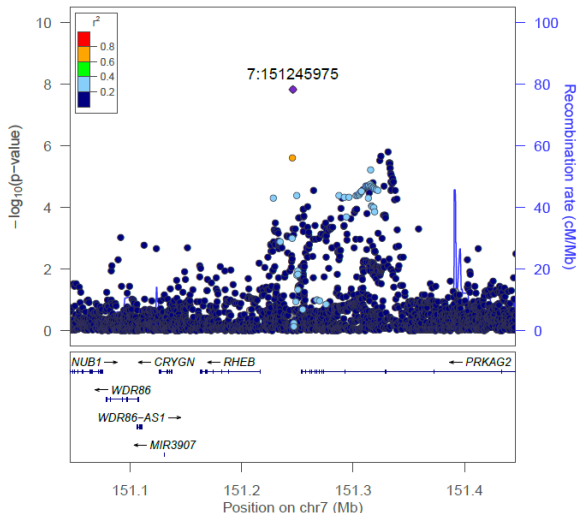
rs73729083 (*CREB3L2*) × Ever Smoking and LDL
TRANS ancestry, 2df tests (n = 84,091)



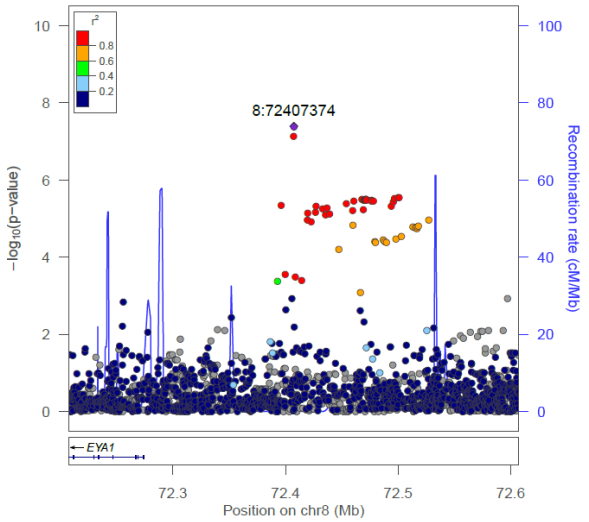
rs73453125 (*CNTNAP2*) × Current Smoking and LDL
TRANS ancestry, 2df tests (n = 40,566)



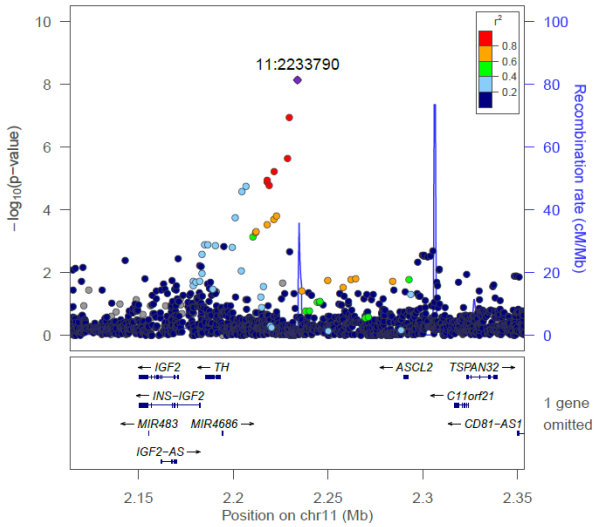
rs56167574 (*PRKAG2*) × Ever Smoking and LDL
AFR ancestry, 1df tests (n = 25,778)



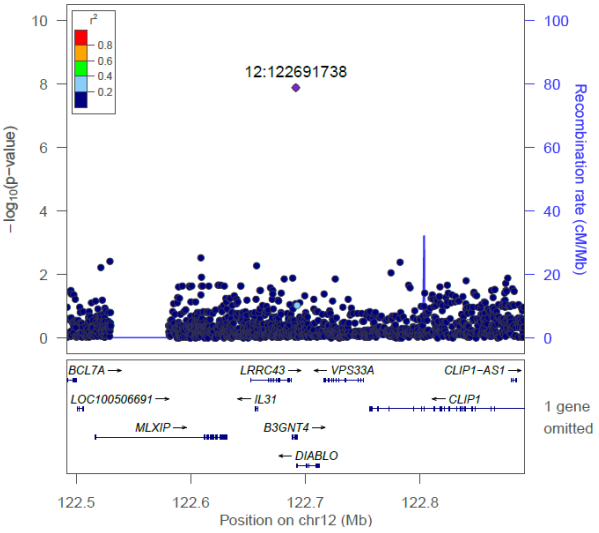
rs10101067 (*EYA1*) × Current Smoking and Triglycerides
TRANS ancestry, 2df tests (n = 317,809)



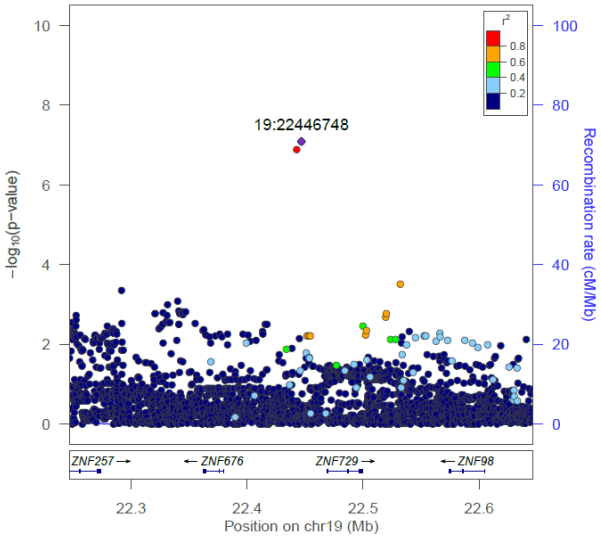
rs79950627 (*MIR4686*) × Current Smoking and LDL
TRANS ancestry, 2df tests (n = 38,272)



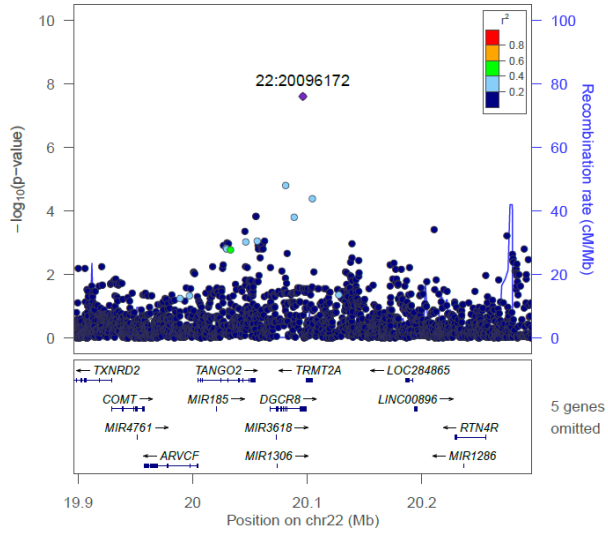
rs4758675 (*B3GNT4*) × Current Smoking and Triglycerides
AFR ancestry, 2df tests (n = 12,982)



rs60029395 (*ZNF729*) × Current Smoking and Triglycerides
AFR ancestry, 1df tests (n = 19,048)

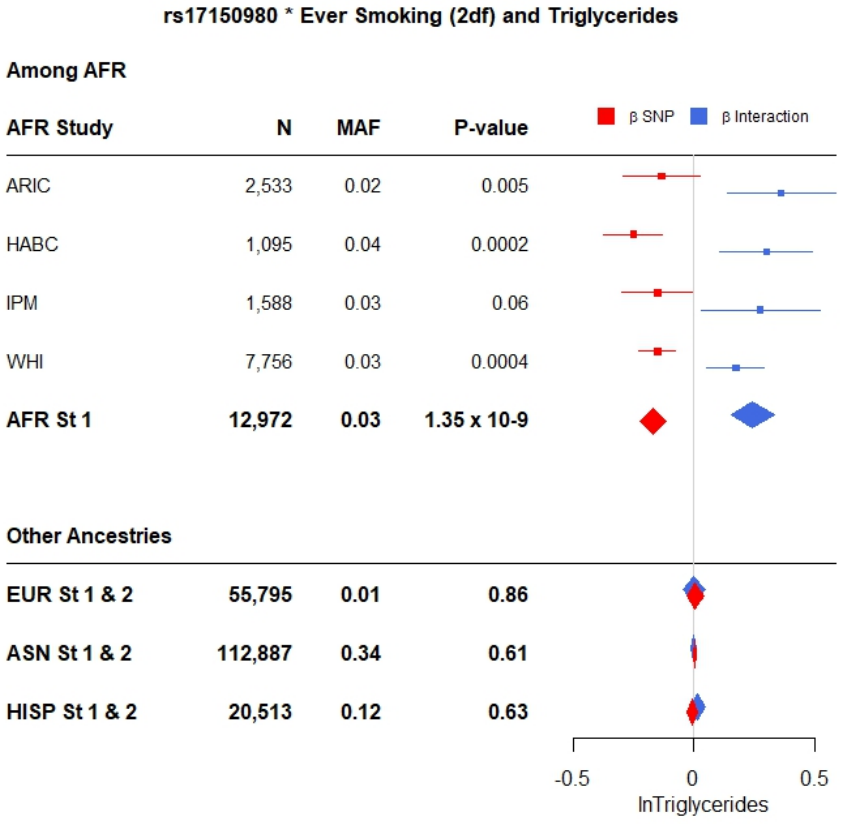


rs7364132 (*DGCR8*) × Ever Smoking and Triglycerides
AFR ancestry, 2df tests (n = 23,935)

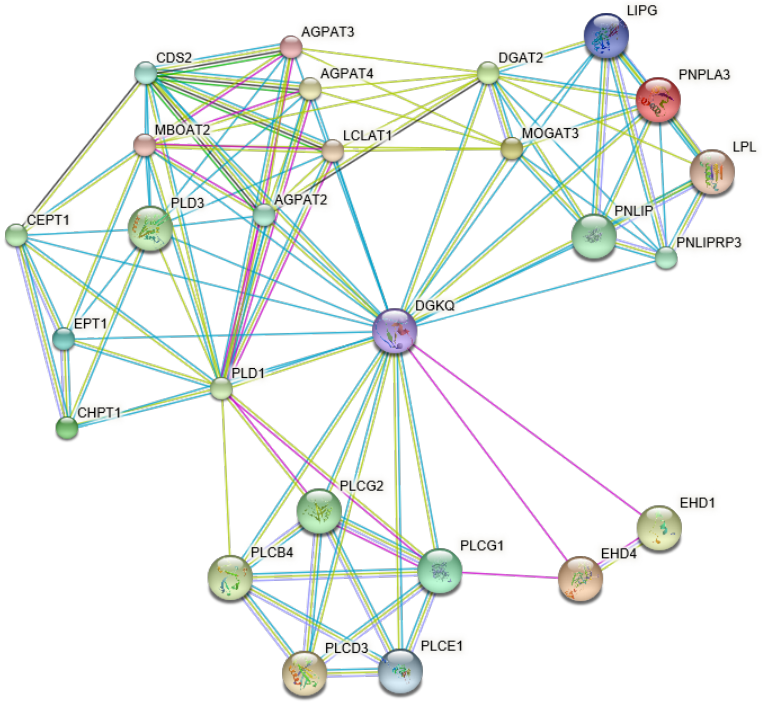


Supplementary Figure 6

rs17150980 × Ever-Smoking on Triglycerides, Stage 1 Meta-Analysis: The results for the 2df test (β for Interaction and Main Effect shown, with 95% confidence intervals) for Stage 1 meta-analyses. This variant was not available in any Stage 2 AFR studies, so follow-up was not possible.



Supplementary Figure 7



CONSORTIA AUTHORS AND AFFILIATIONS:

Consortia

COGENT-Kidney Consortium

Thu H Le[1], Haojia Wu[2], Artur Akbarov[3], Gibran Hemani[4], George Davey Smith[4], Kyle J Gaulton[5], Girish N Nadkarni[6,7], Adan Valladares-Salgado[8], Niels Wacher-Rodarte[9], Josyf C Mychaleckyj[10], Nicole D Dueker[11], Yang Hai[12], Jeffrey Haessler[13], Adrienne M Stilp[14], Gu Zhu[15], James P Cook[16], Johan Arnlov[17,18], Susan H Blanton[11,19], Martin H de Borst[20], Thomas A Buchanan[21], Sylvia Cechova[1], Fadi J Charchar[22-24], Pei-Lun Chu[25], Jeffrey Damman[26], James Eales[3], Ali G Gharavi[27], Vilmantas Giedraitis[28], Andrew C Heath[29], Eli Ipp[30,31], Krzysztof Kiryluk[27], Holly J Kramer[32], Anders Larsson[33], Cecilia M Lindgren[34-36], Pamela AF Madden[29], Grant W Montgomery[37], Leslie J Raffel[38], Ralph L Sacco[39-41], Elena Sanchez[27], Holger Stark[42], Johan Sundstrom[33], Anny H Xiang[43], Aleksandra Zivkovic[42], Lars Lind[33], Erik Ingelsson[44-47], Nicholas G Martin[15], John B Whitfield[15], Jianwen Cai[48], Yukinori Okada[49,50], Koichi Matsuda[51], Tatjana Rundek[39,40], Esteban J Parra[52], Miguel Cruz[8], Maciej Tomaszewski[3,53], and Benjamin D Humphreys[2]

1. Department of Medicine, Division of Nephrology, University of Virginia, Charlottesville, VA 22908, USA.
2. Division of Nephrology, Washington University School of Medicine, St Louis, MO 63110, USA. .
3. Division of Cardiovascular Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester M13 9PT, UK.
4. MRC Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol BS8 1TH, UK.
5. Department of Pediatrics, University of California, San Diego, San Diego, CA 92161, USA.
6. Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
7. Division of Nephrology and Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
8. Unidad de Investigación Médica en Bioquímica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 06720, Mexico. .
9. Unidad de Investigación Médica en Epidemiologia Clinica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 06720, Mexico.
10. Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. .
11. John P Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL 33124, USA. .
12. Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA. .
13. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, USA.
14. Department of Biostatistics, University of Washington, Seattle, WA 98195, USA.
15. Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia.
16. Department of Biostatistics, University of Liverpool L63 3GL, UK.
17. Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Huddinge 141 83, Sweden.
18. School of Health and Social Studies, Dalarna University, Falun 791 88, Sweden.
19. Dr John T Macdonald Department of Human Genetics, University of Miami, Miami, FL 33124, USA.

20. Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, Netherlands. .
21. Department of Medicine, Division of Endocrinology and Diabetes, Keck School of Medicine of USC, Los Angeles, CA 90033, USA.
22. School of Health and Life Sciences, Federation University Australia, Ballarat, VIC 3350, Australia.
23. Department of Cardiovascular Sciences, University of Leicester, Leicester LE1 7RH, UK.
24. Department of Physiology, University of Melbourne, VIC 3010, Australia.
25. Department of Internal Medicine, Fu Jen Catholic University Hospital, School of Medicine, Fu Jen Catholic University, New Taipei City 242, Taiwan.
26. Department of Pathology, Erasmus Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, Netherlands.
27. Department of Medicine, Division of Nephrology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA.
28. Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala 751 85, Sweden.
29. Department of Psychiatry, Washington University in St Louis, St Louis, MO 63110, USA.
30. David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90024, USA.
31. Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Torrance, CA 90502, USA.
32. Department of Medicine and Nephrology, Loyola University Medical Center, Maywood, IL 60153, USA.
33. Department of Medical Sciences, Clinical Epidemiology, Uppsala University, Uppsala 751 85, Sweden.
34. Li Ka Shing Centre for Health Information and Discovery, Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK.
35. Broad Institute of Harvard and MIT, Boston, MA 02142, USA.
36. Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
37. Brisbane Institute for Molecular Bioscience, University of Queensland, St. Lucia, QLD 4072, Australia.
38. Department of Pediatrics, Division of Genetic and Genomic Medicine, University of California, Irvine, Orange, CA 92868, USA.
39. Departments of Neurology and Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL 33136, USA.
40. Evelyn F McKnight Brain Institute, Miller School of Medicine, University of Miami, Miami, FL 33136, USA.
41. Jackson Memorial Hospital, University of Miami, Miami, FL 33136-1096, USA.
42. Institute of Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany.
43. Department of Research and Education, Kaiser Permanente Southern California, Pasadena, CA 91101, USA.
44. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94309, USA.
45. Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94309, USA.
46. Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305, USA. .
47. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 751 85, Sweden. .
48. Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA. .
49. Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan.
50. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Suita 565-0871, Japan.

51. Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan.
52. Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada.
53. Division of Medicine, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK. .

The EPIC-InterAct Consortium

Stephen J Sharp[1], Inês Barroso[2,3], Mark I McCarthy[4-6], Antonio Agudo[7], Guy Fagherazzi[8], Aurelio Barricarte[9-11], Heiner Boeing[12], Miren Dorronsoro[11,13,14], Sara Grioni[15], Rudolf Kaaks[16], Timothy J Key[17], Carmen Navarro[11,18,19], Peter M Nilsson[20], Kim Overvad[21,22], Domenico Palli[23], Salvatore Panico[24], J. Ramón Quirós[25], Olov Rolandsson[26], Carlotta Sacerdote[27,28], María-José Sánchez[11,29,30], Marc Gunter[31], Annemieke MW Spijkerman[32], Anne Tjonneland[33], Rosario Tumino[34,35], Yvonne T van der Schouw[36], Elio Riboli[37]

1. MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom.
2. The Wellcome Trust Sanger Institute, Cambridge, United Kingdom.
3. University of Cambridge Metabolic Research Laboratories, Cambridge, UK.
4. Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, UK.
5. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
6. Oxford NIHR Biomedical Research Centre, Oxford, UK.
7. Catalan Institute of Oncology (ICO), Barcelona, Spain.
8. CESP, University Paris-South, Faculty of Medicine, University Versailles-St Quentin, Inserm U1018, University ParisSaclay, Villejuif, France.
9. Navarre Public Health Institute (ISPN), Pamplona, Spain.
10. Navarra Institute for Health Research (IdiSNA) Pamplona, Spain.
11. CIBER de Epidemiología y Salud Pública (CIBERESP), Spain.
12. German Institute of Human Nutrition Potsdam-Rehbruecke, Germany.
13. Public Health Division of Gipuzkoa, San Sebastian, Spain.
14. Instituto BIO-Donostia, Basque Government, San Sebastian, Spain.
15. Epidemiology and Prevention Unit, Milan, Italy.
16. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
17. University of Oxford, United Kingdom, (20) Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
18. Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Spain.
19. Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
20. Lund University, Malmö, Sweden.
21. Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark.
22. Aalborg University Hospital, Aalborg, Denmark.
23. Cancer Risk Factors and Life-Style Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence, Italy.
24. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy.
25. Public Health Directorate, Asturias, Spain.
26. Umeå University, Umeå, Sweden.
27. Unit of Cancer Epidemiology, Citta' della Salute e della Scienza HospitalUniversity of Turin and Center for Cancer Prevention (CPO), Torino, Italy.
28. Human Genetics Foundation (HuGeF), Torino, Italy.
29. Andalusian School of Public Health, Granada, Spain.
30. Instituto de Investigación Biosanitaria de Granada (Granada.ibs), Granada (Spain).
31. International Agency for Research on Cancer, Lyon, France.

32. National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
33. Danish Cancer Society Research Center, Copenhagen, Denmark.
34. ASP Ragusa, Italy.
35. AIRE - ONLUS, Ragusa, Italy.
36. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.
37. School of Public Health, Imperial College London, UK.

Lifelines Cohort Study

[1]Behrooz Z Alizadeh[1], H Marike Boezen[1], Lude Franke[2], Gerjan Navis[3], Marianne Rots[4], Morris Swertz[2], Bruce HR Wolffenbuttel[5], Cisca Wijmenga[2]

1. Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands
2. Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands
3. Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands
4. Department of Medical Biology, University of Groningen, University Medical Center Groningen, The Netherlands
5. Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands

Understanding Society Scientific Group

Michaela Benzeval[1], Jonathan Burton[1], Nicholas Buck[1], Annette Jäckle[1], Meena Kumari[1], Heather Laurie[1], Peter Lynn[1], Stephen Pudney[1], Birgitta Rabe[1], Dieter Wolke[2]

1. Institute for Social and Economic Research, University of Essex, UK.
2. University of Warwick, UK.

STAGE 1 STUDY DESCRIPTIONS: Brief descriptions are provided below for each of the discovery studies. Unless otherwise noted, the blood draw for serum lipids and the determination of smoking status occurred concurrently (either at the same study visit or within a few months of each other).

AGES (Age Gene/Environment Susceptibility Reykjavik Study): The AGES Reykjavik study originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967. A total of 19,381 people attended, resulting in a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow up and was examined in all stages; another was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGESReykjavik study re-examined 5,764 survivors of the original cohort who had participated before in the Reykjavik Study.

ARIC (Atherosclerosis Risk in Communities): The ARIC study is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals, predominantly European American and African American, aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities. Cohort members completed three additional triennial follow-up examinations,a fifth exam in 2011-2013, and a sixth exam in 2016-2017. The ARIC study has been described in detail previously (The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: Design and objectives. Am J Epidemiol. 1989;129:687-702).

Bio*Me* Biobank (Bio*Me* Biobank of Institute for Personalized Medicine at Mount Sinai): The Bio*Me* Biobank, founded in September 2007, is an ongoing, consented electronic medical record (EMR)-linked bio- and data repository that enrolls participants non-selectively from the Mount Sinai Medical Center patient population. The Bio*Me* Biobank currently (Winter 2015) comprises over 31,000 participants from diverse ancestries characterized by a broad spectrum of (longitudinal) biomedical traits. On average 400 new participants are consented each month. Bio*Me* participants represent the broad ancestral, ethnic and socioeconomic diversity with a distinct and population-specific disease burden, characteristic of Northern Manhattan communities served by Mount Sinai Hospital. Enrolled participants consent to be followed throughout their clinical care (past, present, and future) at Mount Sinai in real-time, integrating their genomic information with their electronic health record for discovery research and clinical care implementation. Bio*Me* participants are predominantly of African, Hispanic/Latino, and European ancestry. Participants who self-identify as Hispanic/Latino further report to be of Puerto Rican (39%), Dominican (23%), Central/South American (17%), Mexican (5%) or other Hispanic (16%) ancestry. More than 40% of European ancestry participants are genetically determined to be of Ashkenazi Jewish ancestry.

The IRB-approved Bio*Me* Biobank consent permits use of samples and de-identified linkable past, present and future clinical information from EMRs; re-contacting participants for enrollment in future research; unlimited duration of storage, and access to clinical information from the entire medical records, as well as local and external sharing of specimens and data.

The Bio*Me* Biobank has a longitudinal design as participants consent to make any EMR data from past (dating back as far as 2003), present and future inpatient or outpatient encounters available for research. The median number of clinical encounters per participant is 21, reflecting predominant enrollment of participants with common chronic conditions from primary care facilities. Mount Sinai's system-wide Epic EMR implementation captures a full spectrum of biomedical phenotypes, including clinical outcomes, covariate and exposure data. This clinical information is complemented by detailed information on ancestry, residence history, familial medical history, education, socio-economic status, physical activity, smoking, alcohol use, and weight history being collected in a systematic manner by interview-based questionnaire at time of enrollment. Phenotype harmonization and validation is critical to facilitate consortium-wide analyses. By applying advanced medical

informatics and data mining tools, high-quality and validated phenotype data can be culled from Mount Sinai's Epic EMR. Fully-implemented phenotype algorithms include; T2D, CKD, CAD, lipid disorders, peripheral artery disease, resistant hypertension, blood cell traits, abdominal aortic aneurism, venous thromboembolism among others (see also Phenotype KnowledgeBase (PheKB) of the eMERGE Network (http://emerge.mc.vanderbilt.edu/emerge-network).

A total of 14,017 participants have been genotyped for both GWAS (11,150 Illumina OmniExpress BeadChip, 2,867 Affymetrix Human SNP Array 6.0) *and* ExomeChip (Illumina HumanExome v1.0 BeadChip) arrays funded by institutional sources. An additional 16,000 Bio*Me* participants are scheduled for genotyping using the Illumina MEGA Chip (by April 2015), funded by NHGRI through our PAGEII grant (U01HG007417) (n=12,500) and through institutional funds (n=3,500).

CARDIA (Coronary Artery Risk Development in Young Adults): CARDIA is a prospective multicenter study with 5,115 adults Caucasian and African American participants of the age group 18-30 years, recruited from four centers at the baseline examination in 1985-1986. The recruitment was done from the total community in Birmingham, AL, from selected census tracts in Chicago, IL and Minneapolis, MN; and from the Kaiser Permanente health plan membership in Oakland, CA. The details of the study design for the CARDIA study have been previously published.(1) Eight examinations have been completed since initiation of the study, respectively in the years 0, 2, 5, 7, 10, 15, 20, 25, and 30. Written informed consent was obtained from participants at each examination and all study protocols were approved by the institutional review boards of the participating institutions. Age and race were self-reported using standardized questionnaires, as were use of cholesterol-lowering medication, and smoking status (current, former, or never). All participants were asked to fast for 12 hours before each clinic visit. Lipid measures were performed on plasma blood samples drawn from the antecubital vein and stored at −70°C until analyzed. Plasma total cholesterol, HDL, and triglyceride levels were measured using enzymatic methods (2); HDL levels were measured after dextran-sulfate-magnesium precipitation of other lipoproteins.(3) LDL levels were estimated with the Friedewald equation for individuals with fasting triglyceride values less than 400 mg/dL.(4) The test-retest correlations for total cholesterol, HDL, LDL, and triglycerides were 0.98 to 0.99.(5)

1. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hully SB, Jacobs DR Jr., Liu K, Savage PJ. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41:1105–1116.

2. Warnick GR. Enzymatic methods for quantification of lipoprotein lipids. Methods Enzymol. 1986;129:101–123.

3. Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg+2 precipitation procedure for quantitation of high-density lipoprotein cholesterol. Clin Chem. 1982;28:1379–1388.

4. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18:499–502.

5. Gross M, Steffes M, Jacobs DR Jr., Yu X, Lewis L, Lewis CE, Loria CM. Plasma F2-isoprostanes and coronary artery calcification: the CARDIA Study. Clin Chem. 2005;51:125–131.

CHS (Cardiovascular Health Study):

CHS (Cardiovascular Health Study): CHS is a population-based cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted across four field centers (1). The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional predominately African-American cohort of 687 persons was enrolled in 1992-

93 for a total sample of 5,888. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. European ancestry participants were excluded from the GWAS study sample due to prevalent coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke, or transient ischemic attack at baseline. After QC, genotyping was successful for 3271 European ancestry and 823 African-American participants. CHS was approved by institutional review committees at each site and individuals in the present analysis gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

1. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991; 1:263-76.

CROATIA-Korcula: The CROATIA-Korcula study is a family-based, cross-sectional study in the isolated island of Korcula that included 965 examinees aged 18-95. Blood samples were collected in 2007 along with many clinical and biochemical measures and lifestyle and health questionnaires.

CROATIA-Vis: The CROATIA-Vis study is a family-based, cross-sectional study in the isolated island of Vis that included 1,056 examinees aged 8-93. Blood samples were collected in 2003 and 2004 along with many clinical and biochemical measures and lifestyle and health questionnaires.

ERF (Erasmus Rucphen Family study): Erasmus Rucphen Family is a family based study that includes inhabitants of a genetically isolated community in the South-West of the Netherlands[1-2], studied as part of the Genetic Research in Isolated Population (GRIP) program. The goal of the study is to identify the risk factors in the development of complex disorders. Study population includes approximately 3,000 individuals who are living descendants of 22 couples who lived in the isolate between 1850 and 1900 and had at least six children baptized in the community church. All data were collected between 2002 and 2005. All participants gave informed consent, and the Medical Ethics Committee of the Erasmus University Medical Centre approved the study.

1. Aulchenko YS, Heutink P, Mackay I et al: Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 2004; 12: 527–534.

2. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS: The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 2005; 69:288–295.

FamHS (Family Heart Study): The NHLBI FamHS study design, collection of phenotypes and covariates as well as clinical examination have been previously described (https://dsgweb.wustl.edu/fhscc/; PMID: 8651220). In brief, the FamHS recruited 1,200 families (approximately 6,000 individuals), half randomly sampled, and half selected because of an excess of coronary heart disease (CHD) or risk factor abnormalities as compared with age- and sex-specific population rates. The participants were sampled from four population-based parent studies: the Framingham Heart Study, the Utah Family Tree Study, and two centers for the Atherosclerosis Risk in Communities study (ARIC: Minneapolis, and Forsyth County, NC). These individuals attended a clinic exam (1994-1996) and a broad range of phenotypes were assessed in the general domains of CHD, atherosclerosis, cardiac and vascular function, inflammation and hemostasis, lipids and lipoproteins, blood pressure, diabetes and insulin resistance, pulmonary function, diet, education, socioeconomic status, habitual behavior, physical activity, anthropometry, medical history and medication use. Approximately 8 years later, study participants belonging to the largest pedigrees were invited for a second clinical exam (2002-04). The most important CHD risk factors were measured again, including lipids, parameters of glucose metabolism, blood pressure, anthropometry, and several biochemical and hematologic markers. In addition, a computed tomography examination provided measures of coronary and aortic calcification, and abdominal and liver fat burden. Medical history and medication use was updated. A total of 2,756 European ancestry subjects in 510 extended random and high CHD risk families

were studied. Also, 633 African ancestry subjects were recruited at ARIC field center at the University of Alabama in Birmingham. Informed consent was obtained from all participants.

FHS (Framingham Heart Study): FHS began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort (mean age 37 years; 52% women) consisted of 5,124 children and spouses of children of the original cohort. A third generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled in 2002-2005 and are seen every 4 to 8 years. Details of study designs for the three cohorts are summarized elsewhere. At each clinic visit, a medical history was obtained with a focus on cardiovascular content, and participants underwent a physical examination including measurement of height and weight from which BMI was calculated.. For this study, lipid measurements were used from the first exam of the 2nd generation (1971-1975) and the 3rd generation (2002-2005) cohorts. Fasting levels of total cholesterol, high density lipoprotein cholesterol, and triglycerides were measured using standard enzymatic methods in accordance with LRC protocols. LDL cholesterol was calculated using the Friedewald formula. Current smoking and ever smoking data were also recorded at the first exam of each cohort and used in analyses.

GENOA (Genetic Epidemiology Network of Arteriopathy): GENOA is one of four networks in the NHLBI Family-Blood Pressure Program (FBPP; PubMed PMIDs 11799070, 15121494). GENOA's long-term objective is to elucidate the genetics of target organ complications of hypertension, including both atherosclerotic and arteriolosclerotic complications involving the heart, brain, kidneys, and peripheral arteries. The longitudinal GENOA Study recruited European-American and African-American sibships with at least 2 individuals with clinically diagnosed essential hypertension before age 60 years. All other members of the sibship were invited to participate regardless of their hypertension status. Participants were diagnosed with hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg based on the second and third readings at the time of their clinic visit. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. During the first exam (1995-2000), 1,583 European Americans from Rochester, MN and 1,854 African Americans from Jackson, MS were examined. Between 2000 and 2005, 1,241 of the European Americans and 1,482 of the African Americans returned for a second examination. Because African-American probands for GENOA were recruited through the Atherosclerosis Risk in Communities (ARIC) Jackson field center participants, we excluded ARIC participants from analyses.

GenSalt (Genetic Epidemiology Network of Salt Sensitivity): GenSalt is a multi-center, family based study designed to identify, through dietary sodium and potassium intervention, salt-sensitivity susceptibility genes which may underlie essential hypertension in rural Han Chinese families. Approximately 629 families with at least one 'proband' with high blood pressure were recruited and tested for a wide variety of physiological, metabolic and biochemical measures at baseline and at multiple times during the 3-week intervention. The intervention consisted of one week on a low sodium diet, followed by one week on a high sodium diet, and finally one week on a high sodium diet with a potassium supplement.

HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span): HANDLS is a community-based, longitudinal epidemiologic study examining the influences of race and socioeconomic status (SES) on the development of age-related health disparities among a sample of socioeconomically diverse African Americans and whites. This unique study will assess over a 20-year period physical parameters and also evaluate genetic, biologic, demographic, and psychosocial, parameters of African American and white participants in higher and lower SES to understand the driving factors behind persistent black-white health disparities in overall longevity, cardiovascular disease, and cognitive decline. The study recruited 3,722 participants from Baltimore, MD with a mean age of 47.7 years, 2,200 African Americans and 1,522 whites, with 41% reporting household incomes below the 125% poverty delimiter.

Genotyping was done on a subset of self-reporting African American participants by the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health (NIH). A larger genotyping effort included a small subset of self-reporting European ancestry samples. This research was supported by the Intramural Research Program of the NIH, NIA and the National Center on Minority Health and Health Disparities.

Health ABC (Health, Aging, and Body Composition): Cohort description: The Health ABC study is a prospective cohort study investigating the associations between body composition, weight-related health conditions, and incident functional limitation in older adults. Health ABC enrolled well-functioning, community-dwelling black (n=1281) and white (n=1794) men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of white and all black Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Participants have undergone annual exams and semi-annual phone interviews. The current study sample consists of 1559 white participants who attended the second exam in 1998-1999 with available genotyping data.

Genotyping: Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1M-Duo BeadChip system. Samples were excluded from the dataset for the reasons of sample failure, genotypic sex mismatch, and first-degree relative of an included individual based on genotype data. Genotyping was successful in 1663 Caucasians. Analysis was restricted to SNPs with minor allele frequency $\geq$ 1%, call rate $\geq$97% and HWE p$\geq$10-6. Genotypes were available on 914,263 high quality SNPs for imputation based on the HapMap CEU (release 22, build 36) using the MACH software (version 1.0.16). A total of 2,543,888 imputed SNPs were analyzed for association with vitamin D levels.

Association analysis: Linear regression models were used to generate cohort-specific residuals of naturally log transformed vitamin D levels adjusted for age, sex, BMI and season defined as summer (June-August), fall (September-November), winter (December to February) and spring (March to May) standardized to have mean 0 and variance of 1. Association between the additively coded SNP genotypes and the vitamin D residuals standardized was assessed using linear regression models. For imputed SNPs, expected number of minor alleles (i.e. dosage) was used in assessing association with the vitamin D residuals.

HERITAGE (Health, Risk Factors, Exercise Training and Genetics): The HERITAGE is the only known family-based study of exercise intervention to evaluate the role of genes and sequence variants involved in the response to a physically active lifestyle. The current study is based on the data collected at baseline of the study from 99 White families (244 males, 255 females). All subjects were required to be sedentary and free of chronic diseases at baseline. There are over 18 trait domains (e.g. dietary, lipids and lipoproteins, glucose and insulin metabolism [fasting and IVGTT], steroids, body composition and body fat distribution, cardiorespiratory fitness), for a grand total of over one thousand variables. Moreover, most of the outcome traits were measured twice on two separate days both at baseline and after exercise training was completed. Marker data include a genome-wide linkage scan and GWAS, in addition to a large number of candidate genes.

HUFS (Howard University Family Study): HUFS followed a population-based selection strategy designed to be representative of African American families living in the Washington, DC metropolitan area. The major objectives of the HUFS were to study the genetic and environmental basis of common complex diseases including hypertension, obesity and associated phenotypes. Participants were sought through door-to-door canvassing, advertisements in local print media and at health fairs and other community gatherings. In order to maximize the utility of this cohort for the study of multiple common traits, families were not ascertained based on any phenotype. During a clinical examination, demographic information was collected by interview.

HyperGEN (Hypertension Genetic Epidemiology Network): HyperGEN is a family-based study that looks at the genetic causes of hypertension and related conditions in EA and AA subjects. HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample.

HyperGEN has collected data on 2,471 Caucasian-American subjects and 2,300 African-American subjects, from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah.

GS:SFHS (Generation Scotland: Scottish Family Health Study): The Generation Scotland (www.generationscotland.org) Scottish Family Health Study (GS:SFHS) is a family-based genetic epidemiology cohort with DNA, other biological samples (serum, urine and cryopreserved whole blood) and socio-demographic and clinical data from approximately 24,000 volunteers, aged 18-98 years, in ~7,000 family groups. An important feature of GS:SFHS is the breadth of phenotype information, including detailed data on cognitive function, personality traits and mental health. Although data collection was cross-sectional, GS:SFHS becomes a longitudinal cohort as a result of the ability to link to routine NHS data, using the community health index (CHI) number.

JHS (Jackson Heart Study): The Jackson Heart Study is a longitudinal, community-based observational cohort study investigating the role of environmental and genetic factors in the development of cardiovascular disease in African Americans. Between 2000 and 2004, a total of 5301 participants were recruited from a tri-county area (Hinds, Madison, and Rankin Counties) that encompasses Jackson, MS. Details of the design and recruitment for the Jackson Heart Study cohort has been previously published.1-3 Briefly, approximately 30% of participants were former members of the Atherosclerosis Risk in Communities (ARIC) study. The remainder were recruited by either 1) random selection from the Accudata list, 2) commercial listing, 3) a constrained volunteer sample, in which recruitment was distributed among defined demographic cells in proportions designed to mirror those in the overall population, or through the Jackson Heart Study Family Study.

1. Wyatt SB, Diekelmann N, Henderson F, Andrew ME, Billingsley G, Felder SH et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. Ethn Dis 2003; 13(4):438-455.

2. Taylor HA, Jr., Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn Dis 2005; 15:S6-17.

3. Fuqua SR, Wyatt SB, Andrew ME, et al. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. Ethn Dis 2005; 15:S6-29.

MESA (Multi-Ethnic Study of Atherosclerosis): The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol. 2002 Nov 1;156(9):871-81. PubMed PMID: 12397006.

NEO (The Netherlands Epidemiology of Obesity study): The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panel. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. In random subsamples of participants, magnetic resonance imaging of abdominal fat, pulse wave velocity of the aorta, heart, and brain, magnetic resonance spectroscopy of the liver, indirect calorimetry, dual energy X-ray absorptiometry, or accelerometry measurements were performed. The collection of data started in September 2008 and completed at the end of September 2012. Participants are currently being followed for the incidence of obesity-related diseases and mortality.

RS (Rotterdam Study): The Rotterdam Study is a prospective, population-based cohort study among individuals living in the well-defined Ommoord district in the city of Rotterdam in The Netherlands[1]. The aim of the study is to determine the occurrence of cardiovascular, neurological, ophthalmic, endocrine, hepatic, respiratory, and psychiatric diseases in elderly people. The cohort was initially defined in 1990 among approximately 7,900 persons, aged 55 years and older, who underwent a home interview and extensive physical examination at the baseline and during follow-up rounds every 3-4 years (RS-I). Cohort was extended in 2000/2001 (RS-II, 3,011 individuals aged 55 years and older) and 2006/2008 (RS-III, 3,932 subjects, aged 45 and older). Written informed consent was obtained from all participants and the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam, approved the study. For RS-II and RS-III, smoking determination and lipids measurements occurred concurrently; for RS-I, however, smoking status was determined during the initial study visit, while the blood draw for lipids measurements was conducted during a third follow-up visit to the study center.

1. Ikram, M.A., Brusselle, G.G.O., Murad, S.D., Duijn, C.M. van, Franco, O.H., Goedegebure, A., Klaver, C.C.W., Nijsten, T.E.C., Peeters, R.P., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vernooij, M.W., Hofman, A. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol 2017 doi:10.1007/s10654-017-0321-4

SCHS-CHD (Singapore Chinese Health Study - Coronary Heart Disease): SCHS-CHD is a case-control study of coronary heart disease that was nested within the Singapore Chinese Health Study (SCHS), a prospective cohort study of 63,257 Singaporean Chinese men and women aged 45-74 years living in Singapore. We selected cases and controls from participants that provided blood samples and were free of coronary heart disease and stroke at the time of blood collection (N=24,454). Cases (N=760) had acute myocardial infarction (AMI) or died of coronary heart disease. AMI was identified through the Singapore Myocardial Infarction Registry or through the nationwide hospital discharge database followed by confirmation of AMI by cardiologists' review of medical records using the Multi-Ethnic Study of Atherosclerosis criteria (available at: http://www.mesa-nhlbi.org/manuals.aspx). Coronary heart disease deaths were identified through the Singapore Registry of Births and Deaths (ICD9 410-414 as first stated cause of death). Matched controls (N=1,491) were selected using a risk-set sampling strategy. Controls were participants who were alive and free of coronary heart disease at the time of the diagnosis or death of the index cases and were matched for age, sex, dialect group, year of recruitment and date of blood collection. In-person interviews and phlebotomy were conducted before the onset of disease and non-fasting venous blood was stored at -80^0C for extraction of DNA and blood biochemistry.

SEED (Singapore Epidemiology of Eye Diseases): SiMES (Singapore Malay Eye Study): SiMES is a population-based cross-sectional epidemiological study of 3,280 individuals from one of the three major ethnic groups residing in Singapore. In summary, 5,600 individuals have been selected by an age-stratified sampling strategy. Among these 4,168 individuals are eligible for this study. 3,280 individuals finally participated in the

study (78.7% response rate). All subjects were Malay and aged 40-80 years [PMID: 17365815; 25953847]. Non-fasting lipid levels were measured by an automated autoanalyzer (ADVIA 2400, Bayer Diagnostics). **SINDI (Singapore Indian Eye Study)**: SINDI is a population-based, cross-sectional study of Asian Indian adults aged 40–80 years residing in the Southwestern part of Singapore. Age stratified random sampling was used to select 6,350 eligible participants, of which 3,400 participated in the study (75.6% response rate). Detailed methodology has been published [PMID: 19995197; 25953847]. Non-fasting lipid levels were measured by an automated autoanalyzer (Beckman Coulter Unicel DxC 800). **SCES (Singapore Chinese Eye Study):** SCES is a population-based, cross-sectional study of Chinese adults aged 40–80 years residing in the Southwestern part of Singapore. Age stratified random sampling was used to select 6,350 eligible participants, of which 3,353 participated in the study (72.8% response rate). Detailed methodology has been published [PMID: 19995197; 25953847]. Non-fasting lipid levels were measured by an automated autoanalyzer (Beckman Coulter Unicel DxC 800).

SP2 (Singapore Prospective Study Program): SP2 is a population-based study of diabetes and cardiovascular disease in Singapore. It first surveyed subjects (Chinese, Malay and Indian) from four cross-sectional studies that were conducted in Singapore between 1982 and 1998. Subjects were between the ages of 24-95 years and represented a random sample of the Singapore population. Subjects were re-visited between 2003 and 2007. Among the 10,747 individuals who were eligible, 5,157 subjects completed a questionnaire and the subsequent clinical examinations [PMID: 19406920]. Data from this re-visit were utilized for this study. Fasting HDL-C, TC and TG were measured by an automated analyzer autoanalyzer (ADVIA 2400, Bayer Diagnostics). LDL-C was calculated from Friedewald formula [PMID: 4337382].

WGHS (Women's Genome Health Study): WGHS is a prospective cohort of female North American health care professionals representing participants in the Women's Health Study (WHS) trial who provided a blood sample at baseline and consent for blood-based analyses (PMID: 18070814). Participants in the WHS were 45 years or older at enrollment and free of cardiovascular disease, cancer or other major chronic illness. The current data are derived from 23,294 WGHS participants for whom whole genome genotype information was available at the time of analysis and for whom self-reported European ancestry could be confirmed by multidimensional scaling analysis of 1,443 ancestry informative markers in PLINK v. 1.06. At baseline, BP and lifestyle habits related to smoking, consumption of alcohol, and physical activity as well as other general clinical information were ascertained by a self-reported questionnaire, an approach which has been validated in the WGHS demographic, namely female health care professionals.

WHI (Women's Health Initiative): WHI is a long-term national health study that focuses on strategies for preventing common diseases such as heart disease, cancer and fracture in postmenopausal women. A total of 161,838 women aged 50–79 years old were recruited from 40 clinical centers in the US between 1993 and 1998. WHI consists of an observational study, two clinical trials of postmenopausal hormone therapy (HT, estrogen alone or estrogen plus progestin), a calcium and vitamin D supplement trial, and a dietary modification trial [1]. Study recruitment and exclusion criteria have been described previously [1]. Recruitment was done through mass mailing to age-eligible women obtained from voter registration, driver's license and Health Care Financing Administration or other insurance list, with emphasis on recruitment of minorities and older women [2]. Exclusions included participation in other randomized trials, predicted survival < 3 years, alcoholism, drug dependency, mental illness and dementia. For the CT, women were ineligible if they had a systolic BP > 200 mm Hg or diastolic BP > 105 mm Hg, a history of hypertriglyceridemia or breast cancer. Study protocols and consent forms were approved by the IRB at all participating institutions. Socio-demographic characteristics, lifestyle, medical history and self-reported medications were collected using standardized questionnaires at the screening visit. Physical measures of height, weight and blood pressure were measured at a baseline clinical visit [2]. The genome wide association study (GWAS) non-overlapping samples are composed of a case-control study (WHI Genomics and Randomized Trials Network – GARNET, which included all coronary heart disease, stroke, venous thromboembolic events and selected diabetes cases that happened during the active intervention phase in the WHI HT clinical trials and aged matched controls), women selected to be "representative" of the HT trial (mostly

younger white HT subjects that were also enrolled in the WHI memory study - WHIMS) and the WHI SNP Health Association Resource (WHI SHARe), a randomly selected sample of 8,515 African American and 3,642 Hispanic women from WHI. GWAS was performed using Affymetrix 6.0 (WHI-SHARe), HumanOmniExpressExome-8v1_B (WHIMS), Illumina HumanOmni1-Quad v1-0 B (GARNET). Extensive quality control (QC) of the GWAS data included alignment ("flipping") to the same reference panel, imputation to the 1000G data (using the recent reference panel - v3.20101123), identification of genetically related individuals, and computations of principal components (PCs) using methods developed by Price et al. (using EIGENSOFT software 53), and finally the comparison with self-reported ethnicity. After QC and exclusions from analysis protocol, the number of women included in analysis is 4,423 whites for GARNET, 5,202 white for WHIMS, 7,919 for SHARe African American and 3,377 for SHARe Hispanics.

1. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, Rossouw JE. The women's health initiative recruitment methods and results. Ann Epidemiol. 2003;13:S18-77

2. Design of the women's health initiative clinical trial and observational study. The women's health initiative study group. Control Clin Trials. 1998;19:61-109

STAGE 2 STUDY DESCRIPTIONS: Brief descriptions are provided below for each of the replication studies/cohorts. Unless otherwise noted, the blood draw for serum lipids and the determination of smoking status occurred concurrently (either at the same study visit or within a few months of each other).

AA-DHS (African American Diabetes Heart Study): AA-DHS objectives are to improve understanding of ethnic differences in CAC and CP in populations of African and European ancestry. The AA-DHS consists of self-reported African Americans with T2D recruited from two Wake Forest School of Medicine (WFSM) studies: the family-based Diabetes Heart Study (DHS) and unrelated individuals in the AA-DHS. DHS is a cross-sectional study of European American and African American families with siblings concordant for T2D. AA-DHS started after DHS and enrolled unrelated African Americans. The AA-DHS GWAS utilized the Illumina 5M chip with imputation to 1,000 Genomes.

Airwave (The Airwave Health Monitoring Study): The Airwave Health Monitoring Study (22) was established to evaluate possible health risks associated with use of TETRA, a digital communication system used by police forces and other emergency services in Great Britain since 2001. The study has been broadened to investigate more generally the health of the work force. From 2004, participants from each force who agreed to participate were enrolled either with an enrolment questionnaire or a comprehensive health screening performed locally. This includes questionnaire, 7-day food diaries, anthropometry, measurements of cardiovascular and cognitive function, blood chemistry, coagulation and hematology. By March 2015, the study had recruited 53,606 participants, of whom 45,433 had attended the health screening, and 14,002 have genotype data (1000G imputed).

Ref: Elliott, P. et al. The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. Environ Res 134, 280-5 (2014).

ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial): ASCOT is a randomised control clinical trial investigating the cardiac outcomes of blood pressure lowering and lipid lowering treatments. Of 19,342 hypertensive patients (40–79 years of age with at least three other cardiovascular risk factors) who were randomized to one of two antihypertensive regimens in ASCOT (atenolol, Beta-Blocker vs amlodipine, Calcium-Channel-Blocker), 10,305 patients with non-fasting total cholesterol concentrations of 6.5 mmol/l or less (measured at the non-fasting screening visit) had been randomly assigned additional atorvastatin 10 mg or placebo. Only a proportion of United Kingdom, Irish, Sweden, Norway, Finland and Denmark consented to contribute DNA and participate in genetic studies. PMID 11685901

Baependi Heart Study (Brazil): The Baependi Heart Study, is an ongoing family-based cohort conducted in a rural town of the state of Minas Gerais. The study has enrolled approximate 2,200 individuals (over 10% of the town's adult population) and 10-year follow up period of longitudinal data. Briefly, probands were selected at random across 11 out of the 12 census districts in Baependi. After enrolment, the proband's first-degree (parents, siblings, and offspring), second-degree (half-siblings, grandparents/grandchildren, uncles/aunts, nephews/nieces, and double cousins), and third-degree (first cousins, great uncles/aunts, and great nephews/nieces) relatives, and his/her respective spouse's relatives resident both within Baependi (municipal and rural area) and surrounding towns were invited to participate. Only individuals age 18 and older were eligible to participate in the study. The study is conducted from a clinic/office in an easily accessible sector of the town, where the questionnaires were completed. A broad range of phenotypes ranging from cardiovascular, neurocognitive, psychiatric, imaging, physiologic and several layers of endophenotypes like metabolomics and lipidomics have been collected throughout the years Details about follow-up visits and available data can be found in the cohort profile paper (PMID: 18430212). DNA samples were genotyped using the Affymetrix 6.0 genechip. After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

BBJ (Biobank Japan Project): The Biobank Japan (BBJ) Project was established in 2003 with the aim of the implementation of personalized medicine as a leading project of Ministry of Education, Culture, Sports, Science and Technology (MEXT). In collaboration with twelve cooperating institutes, the BBJ has recruited a total of 200,000 people, suffering from at least one of the 47 target common diseases, in the first phase (5-year period). BBJ has collected biospecimens including DNA and serum as well as various clinical and lifestyle information through interview or medical records by using standardized questionnaire. All participants gave written informed consent to this project and this study was approved by ethical committees of RIKEN and participating institutes.

BES (Beijing Eye Study): Beijing Eye Study is a population-based study that assess the associated and risk factors of ocular and general diseases in China population. The study was initialized in 2001, collected data from 4439 subjects aged ≥ 40 years from seven communities in Beijing area, where three of the communities were located in rural districts and four were located in urban districts. BES was followed-up in 2006, with 3251 of the original subjects participated, and in 2011, with 2695 subjects returned for the follow-up examination. At the examinations in 2006 and 2011, trained research staffs asked the subjects questions from a standard questionnaire providing information on family status, level of education, income, quality of life, psychic depression, physical activity, and known major systemic diseases. Fasting blood samples were taken for measurement of blood lipids, glucose, and glycosylated hemoglobin. Individuals were classified as self-reported non-smokers or self-reported current smokers. Alcohol consumption habits based on number of drinks per day were collected. All variables used in analyses were taken from examinations in 2006 or in 2011. The BES subjects were genotyped on two arrays, Illumina Human610-Quad (N = 832) and Illumina OmniExpress (N = 814).

BRIGHT (British Genetics of Hypertension): Participants of the BRIGHT Study are recruited from the Medical Research Council General Practice Framework and other primary care practices in the UK. Each case had a history of hypertension diagnosed prior to 60 years of age with confirmed blood pressure recordings corresponding to seated levels >150/100mmHg (1 reading) or mean of 3 readings >145/95 mmHg. BRIGHT is focused on recruitment of hypertensive individuals with BMI<30. Sample selection for GWAS was based on DNA availability and quantity. PMID 12826435

CAGE-Amagasaki (Cardio-metabolic Genome Epidemiology Network, Amagasaki Study): The Amagasaki Study (CAGE-Amagasaki) is an ongoing population-based cohort study of 5,743 individuals (3,435 males and 2,310 females), aged >18 years and recruited for a baseline examination between September 2002 to August 2003. Participants were interviewed by trained personnel to obtain information on medical and lifestyle variables, and consented to provide DNA for genotyping of molecular variants to investigate genetic susceptibility for so-called lifestyle-related diseases such as hypertension and cardiovascular disorder.

CFS (Cleveland Family Study): The Cleveland Family Study (CFS) is a family-based, longitudinal study designed to characterize the genetic and non-genetic risk factors for sleep apnea. In total, 2534 individuals (46% African American) from 352 families were studied on up to 4 occasions over a period of 16 years (1990-2006). The initial aim of the study was to quantify the familial aggregation of sleep apnea. 632 African Americans were genotyped on the Affymetrix array 6.0 platform through the CARe Consortium with suitable genotying quality control. A further 122 African-Americans had genotyping based on the Illumina OmniExpress + Exome platform. Genomes were imputed separately for each chip based on a 1000 Genomes Project Phase 3 Version 5 cosmopolitan template using SHAPEIT and IMPUTE2..

DESIR (Data from an Epidemiological Study on the Insulin Resistance): The DESIR cohort study aims to: describe and understand the relations between the abnormalities of the syndrome, their evolution, according to age and sex; search for risk factors of insulin resistance, in particular factors associated with the environment, lifestyle and genetic markers; quantify the links between the syndrome and both cardiovascular disease and diabetes; evaluate the frequency of the syndrome in terms of its consequences on public health.

DFTJ (Dongfeng-Tongji Cohort Study): The DFTJ-cohort study includes 27,009 retired employees from a state-owned automobile enterprise in China. This study was launched in 2008 and will be followed up every 5 years. In 2013 we conducted the first follow-up. By using semi-structural questionnaire and health examination, those having cancer or severe diseases were excluded. Fasting blood samples and detailed epidemiology data were collected. The main goal of the cohort was to identify the environmental and genetic risk factors and the gene-environment interactions on chronic diseases, and to find novel biomarkers for chronic disease and mortality prediction. Finally, 1,461 included in the present study with GWAS data. All of the participants wrote informed consent and the ethical committees in the Tongji Medical College approved this research project. Detailed information has been described in elsewhere(1).

QC criteria and imputation methods:

We did the GWAS scan on the DFTJ-cohort with Affymetrix Genome-Wide Human SNP Array 6.0 chips. In total, we genotyped 906,703 SNPs among 1,461 subjects. After stringent QC filtering, SNPs with MAF < 0.01, Hardy-Weinberg Equilibrium (HWE) < 0.0001, and SNP call rate < 95% were excluded. Individuals with call rates < 95% were also not included for further analysis. In total, we retained 1,452 subjects with 658,288 autosomal SNPs for statistical analyses, with an overall call rate of 99.68%. We used MACH 1.0 software to impute untyped SNPs using the LD information from the HapMap phase II database (CHB+JPT as a reference set (2007-08_rel22, released 2007-03-02). Imputed SNPs with high genotype information content (Rsq > 0.3 for MACH) were kept for the further association analysis.

Reference

1) Wang, F., Zhu, J., Yao, P., Li, X., He, M., Liu, Y., Yuan, J., Chen, W., Zhou, L., Min, X. et al. (2012) Cohort profile: The Dongfeng-Tongji cohort study of retired workers. International journal of epidemiology.

DHS (Diabetes Heart Study): The Diabetes Heart Study (DHS) is an ongoing family-based cohort study investigating the epidemiology and genetics of cardiovascular disease (CVD) in a population-based sample. The DHS recruited T2D-affected siblings without advanced renal insufficiency from 1998 through 2005 in western North Carolina. DHS has collected genetic data on 1,220 self-described European American (EA) individuals from 475 families. Genotyping was completed using an Affymetrix Genome-Wide Human SNP Array 5.0 with imputation of 1,000 Genomes project SNPs from this array using IMPUTE2 and the Phase I v2, cosmopolitan (integrated) reference panel, build 37.

DR's EXTRA (Dose Responses to Exercise Training): The Dose-Responses to Exercise Training (DR's EXTRA) Study is a 4-year RCT on the effects of regular physical exercise and healthy diet on endothelial function, atherosclerosis and cognition in a randomly selected population sample (n=3000) of Eastern Finnish men and women, identified from the national population register, aged 55-74 years. Of the eligible sample, 1410 individuals were randomized into one of the 6 groups: aerobic exercise, resistance exercise, diet, combined aerobic exercise and diet, combined resistance exercise and diet, or reference group following baseline assessments. During the four year intervention the drop-out rate was 15%.

EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)): The Estonian Biobank is the population-based biobank of the Estonian Genome Center at the University of Tartu (www.biobank.ee; EGCUT). The entire project is conducted according to the Estonian Gene Research Act and all of the participants have signed the broad informed consent. The cohort size is up to 51535 individuals from 18 years of age and up, which closely reflects the age, sex and geographical distribution of the Estonian population. All of the subjects are recruited randomly by general practitioners and physicians in hospitals. A Computer Assisted Personal interview is filled within 1-2 hours at a doctor's office, which includes personal, genealogical, educational, occupational history and lifestyle data. Anthropometric measurements, blood pressure and resting heart rate are measured and venous blood taken during the visit. Medical history and current health status is recorded according to ICD-10 codes.

EPIC (European Prospective Investigation into Cancer and Nutrition)-Norfolk: The European Prospective Investigation of Cancer (EPIC)-Norfolk study is a population-based cohort study established to study the links diet, lifestyle factors and cancer and other health outcomes. Participants are men and women who were aged between 40 and 79 when they joined the study and who lived in Norwich, UK and the surrounding towns and rural areas. They have been contributing information about their diet, lifestyle and health through questionnaires and health checks over two decades. The Norwich Local Research Ethics Committee granted ethical approval for the study. All participants gave written informed consent.

The EPIC-InterAct Case-Cohort Study: EPIC- InterAct is a type 2 diabetes case-cohort study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) study. EPIC was initiated in the late 1980s and involves collaboration between 23 research institutions across Europe in 10 countries. The majority of EPIC cohorts were recruited from the general population, with some exceptions. French cohorts included women who were members of a health insurance scheme for school and university employees; Turin and Ragusa (Italy) and the Spanish centres included some blood donors. Participants from Utrecht (Netherlands) and Florence (Italy) were recruited via a breast cancer screening program. The majority of participants recruited by the EPIC Oxford (UK) centre consisted of vegetarian and "health conscious" volunteers from England, Wales, Scotland, and Northern Ireland. EPIC-InterAct sampled a random sub-cohort and all individuals who subsequently developed incident T2DM over follow up from the full cohort of participants in EPIC who provided blood samples at baseline in Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden and the UK. Smoking status was determined at baseline and lipid measurements at baseline were undertaken in a centralized laboratory on all participants in the case-cohort study.

FENLAND (The Fenland Study): The Fenland study is a population-based cohort study that uses objective measures of disease expoure to investigate the influence of diet, lifestyle and genetic factors on the development of diabetes and obesity. The volunteers are recruited from general practice lists in and around Cambridgeshire (Cambridge, Ely, and Wisbech) in the United Kingdom from birth cohorts from 1950–1975.

FUSION (Finland-United States Investigation of NIDDM Genetics): The Finland-United States Investigation of NIDDM Genetics (FUSON) study is a long-term effort to identify genetic variants that predispose to type 2 diabetes (T2D) or that impact the variability of T2D-related quantitative traits. The FUSION GWAS sample consists of 1,161 Finnish T2D cases and 1,174 Finnish normal glucose-tolerant (NGT) controls (Scott et al. Science 2007). Cases are defined by fasting plasma glucose ≥ 7.0 mmol/l or 2-h plasma glucose ≥ 11.1 mmol/l, by report of diabetes medication use, or based on medical record review. 789 FUSION cases each reported at least one T2D sibling; 372 Finrisk 2002 T2D cases came from a Finnish population-based risk factor survey. NGT controls are defined by fasting glucose < 6.1 mmol/l and 2-h glucose < 7.8 mmol/l. FUSION controls include 119 subjects from Vantaa, Finland who were NGT at ages 65 and 70 years, 304 NGT spouses from FUSION families, and 651 Finrisk 2002 subjects. The controls were approximately frequency matched to the cases by age, sex, and birth province. Smoking and alcohol data are only available in the FUSION subset of our GWAS samples.

Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316, 1341–1345, 2007.

GeneSTAR (Genetic Studies of Atherosclerosis Risk): GeneSTAR is a family-based prospective study of more than 4000 participants begun in 1983 to determine phenotypic and genetic causes of premature cardiovascular disease. Families were identified from 1983-2006 from probands with a premature coronary disease event prior to 60 years of age who were identified at the time of hospitalization in any of 10 hospitals in the Baltimore, Maryland area. Their apparently healthy 30-59 year old siblings without known coronary disease were recruited and screened between 1983 and 2006. From 2003-2006, adult offspring over 21 years of age of all participating

siblings and probands, as well as the coparents of the offspring were recruited and screened. Genotyping was performed in 3,232 participants on the Illumina 1Mv1_c platform.

GLACIER (Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk): The Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER) Study is nested within the Västerbotten Health Survey, which is part of the Northern Sweden Health and Disease Study, a population-based prospective cohort study from northern Sweden. Participants were genotyped with Illumina CardioMetaboChip array. This array contains ~200,000 variants, the majority being common variants. Systolic and diastolic blood pressures were measured once following a period of five minutes rest with the participant in the supine position using a mercury-gauge sphygmomanometer. Analysis of serum lipids (HDL-C, triglycerides and total cholesterol) were undertaken at the Department of Clinical Chemistry at Umeå University Hospital using routine methods. LDL-C was determined using the Friedewald formula. All participants completed a detailed, optically readable, health and lifestyle questionnaire including questions about smoking status and alcohol intake (FFQ). Cohort description - PMID: 25396097

GRAPHIC (Genetic Regulation of Arterial Pressure of Humans in the Community): The GRAPHIC Study comprises 2024 individuals from 520 nuclear families recruited from the general population in Leicestershire, UK between 2003-2005 for the purpose of investigating the genetic determinants of blood pressure and related cardiovascular traits. A detailed medical history was obtained from study subjects by standardized questionnaires and clinical examination was performed by research nurses following standard procedures. Measurements obtained included height, weight, waist-hip ratio, clinic and ambulatory blood pressure and a 12-lead ECG.

HCHS/SOL (Hispanic Community Health Study/ Study of Latinos): The HCHS/SOL is a community-based cohort study of 16,415 self-identified Hispanic/Latino persons aged 18–74 years and selected from households in predefined census-block groups across four US field centers (in Chicago, Miami, the Bronx, and San Diego). The census-block groups were chosen to provide diversity among cohort participants with regard to socioeconomic status and national origin or background. The HCHS/SOL cohort includes participants who self-identified as having a Hispanic/Latino background; the largest groups are Central American (n = 1,730), Cuban (n = 2,348), Dominican (n = 1,460), Mexican (n = 6,471), Puerto Rican (n = 2,728), and South American (n = 1,068). The HCHS/SOL baseline clinical examination occurred between 2008 and 2011 and included comprehensive biological, behavioral, and sociodemographic assessments. Consenting HCHS/SOL subjects were genotyped at Illumina on the HCHS/SOL custom 15041502 B3 array. The custom array comprised the Illumina Omni 2.5M array (HumanOmni2.5-8v.1-1) ancestry-informative markers, known GWAS hits and drug absorption, distribution, metabolism, and excretion (ADME) markers, and additional custom content including ~150,000 SNPs selected from the CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California), and PUR (Puerto Rican in Puerto Rico) samples in the 1000Genomes phase 1 data to capture a greater amount of Amerindian genetic variation. QA/QC procedures yielded a total of 12,803 unique study participants for imputation and downstream association analyses.

HRS (Health & Retirement Study): The Health and Retirement Study (HRS) is a longitudinal survey of a representative sample of Americans over the age of 50. The current sample is over 26,000 persons in 17,000 households. Respondents are interviewed every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006, 2008, and 2010. Respondents were removed if they had missing genotype or phenotype data.

Juster, F. T., Suzman, R. (1995). An Overview of the Health and Retirement Study, Journal of Human Resources 30:Suppl: S7-S56.

Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). Int. J. Epidemiol. 2014; 43 (2): 576-585. PMID: 24671021

Crimmins, E.M., Guyer H., Langa K.M., Ofstedal M.B., Wallace R.B., and Weir D.R. (2008). Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study. HRS Documentation Report DR-011. http://hrsonline.isr.umich.edu/sitedocs/userg/dr-011.pdf

HyperGEN-AXIOM (Hypertension Genetic Epidemiology Network): HyperGEN is a family-based study that investigates the genetic causes of hypertension and related conditions in EA and AA subjects. HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample. HyperGEN has collected data on 2,471 Caucasian-American subjects and 2,300 African-American subjects, from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah. HyperGEN participates as a discovery study using GWAS available in a large subset of the samples. The remaining AA subjects without GWAS data were genotyped on the Affymetrix Axiom chip as part of a HyperGEN admixture mapping ancillary study. After excluding subjects already included in the original HyperGEN (or with family members included), this subset of approximately 450 AA subjects are included in the HyperGEN-AXIOM study which participates in replications.

Shetty PB, Tang H, Tayo BO, Morrison AC, Hanis CL, Rao DC, Young JH, Fox ER, Boerwinkle E, Cooper RS, Risch NJ, Zhu X; Candidate Gene Association Resource (CARe) Consortium. Variants in CXADR and F2RL1 are associated with blood pressure and obesity in African-Americans in regions identified through admixture mapping. J Hypertens. 2012 Oct;30(10):1970-6. PMID:22914544

INGI-CARL & INGI-FVG (Italian Network Genetic Isolates): INGI-FVG and INGI-CARL studies include samples coming from isolated populations and belong to the Italian Network of Genetic Isolates (INGI). The Carlantino cohort (INGI-CARL) is a population-based study including approximately 1000 samples from an isolated village of Southern Italy. INGI-CARL examined about 1000 subjects between 1998 and 2005 coming from a small village of the South of Italy situated in the extreme northern part of Puglia Region, while INGI-FVG involved about 1700 subjects between 2008 and 2011 coming from six different villages located in the North-East of Italy in Friuli Venezia Giulia region. A questionnaire was administered to each participant to obtain socio-demographic information, as well as data on professional activity, family history, eating habits and lifestyle, such as smoking, coffee and alcohol consumption, physical activity. Furthermore, a medical screening, including anamnesis, blood pressure, drugs and clinical chemistry evaluation (blood count and different biochemical parameters, such as lipids) were made. All participants gave their written informed consent.

IRAS (Insulin Resistance Atherosclerosis Study): The Insulin Resistance Atherosclerosis Study (IRAS) was an epidemiologic cohort study designed to examine the relationship between insulin resistance and carotid atherosclerosis across a range of glucose tolerance. Individuals of self-reported Mexican-American ethnicity were recruited in San Antonio, TX and San Luis Valley, CO. Recruitment was balanced across age and glucose tolerance status. Inclusion of IRAS data is limited to 194 normoglycemic individuals with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genome Integrated Reference Panel (phase I).

IRAS Family Study (Insulin Resistance Atherosclerosis Study): The IRASFS was a family study designed to examine the genetic and epidemiologic basis of glucose homeostasis traits and abdominal adiposity. Briefly, self-reported Mexican American pedigrees were recruited in San Antonio, TX and San Luis Valley, CO. Probands with large families were recruited from the initial non-family-based IRAS, which was modestly enriched for impaired glucose tolerance and T2D. Inclusion of IRASFS data is limited to 1040 normoglycemic individuals in 88 pedigrees with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genome Integrated Reference Panel (phase I).

JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin): Genetic analysis was performed in a sub-population from JUPITER (Justification for the Use of statins in Prevention: an Intervention Trial Evaluating Rosuvastatin), an international, randomized, placebo-controlled trial of rosuvastatin (20mg/day) in the primary prevention of cardiovascular disease conducted among apparently healthy men and women with LDL-C < 130 mg/dL and hsCRP ≥ 2 mg/L (PMIDs: 18997196, 22331829). Individuals with diabetes or triglyceride concentration >500mg/dL were excluded. The present analysis includes only individuals who provided consent for genetic analysis, had successfully collected genotype information, and who had either verified European or verified South African black ancestry.

KORA (Cooperative Health Research in the Augsburg Region): The KORA study is a series of independent population-based epidemiological surveys of participants living in the region of Augsburg, Southern Germany. All survey participants are residents of German nationality identified through the registration office and were examined in 1994/95 (KORA S3) and 1999/2001 (KORA S4). In the KORA S3 and S4 studies 4,856 and 4,261 subjects have been examined implying response rates of 75% and 67%, respectively. 3,006 subjects participated in a 10-year follow-up examination of S3 in 2004/05 (KORA F3), and 3080 of S4 in 2006/2008 (KORA F4). The age range of the participants was 25 to 74 years at recruitment. Informed consent has been given by all participants. The study has been approved by the local ethics committee. Individuals for genotyping in KORA F3 and KORA F4 were randomly selected and these genotypes are taken for the analysis of the phenotypes in KORA S3 and KORA S4.

LBC1936 (Lothian Birth Cohort 1936): LBC1936 consists of 1091 (548 male) relatively healthy individuals who underwent cognitive and medical testing at a mean age of 69.6 years (SD = 0.8). They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland.[1]

(1) Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol 2012;41:1576-1584.

Lifelines (Netherlands Biobank): Lifelines (https://lifelines.nl/) is a multi-disciplinary prospective population-based cohort study using a unique three-generation design to examine the health and health-related behaviors of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity. In addition, the Lifelines project comprises a number of cross-sectional sub-studies which investigate specific age-related conditions. These include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function and depression, and musculoskeletal conditions. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and over 130,000 participants had been included by April 2013. At the baseline examination, the participants in the study were asked to fill in a questionnaire (on paper or online) before the first visit. During the first and second visit, the first or second part of the questionnaire, respectively, are checked for completeness, a number of investigations are conducted, and blood and urine samples are taken. Lifelines is a facility that is open for all researchers. Information on application and data access procedure is summarized on www.lifelines.nl.

Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol. 2014 Dec 14.

LLFS (The Long Life Family Study): LLFS is a family-based cohort study, including four clinical centers: Boston University Medical Center in Boston, MA, USA, Columbia University College of Physicians and Surgeons in New York City, NY, USA, the University of Pittsburgh in Pittsburgh PA, USA, and University of

Southern Denmark, Denmark. The study characteristics, recruitment, eligibility and enrollment have been previously described (Pedersen et al., 2006, PMID: 17150149; Sebastiani et al., 2009, PMID: 19910380; Newman et al., 2011, PMID: 21258136). In brief, the LLFS was designed to determine genetic, behavioral, and environmental factors related to families of exceptionally healthy, elderly individuals. Phase 1 was conducted between 2006 and 2009 recruiting 4,953 individuals from 539 families. The probands were at least 79 years old in the USA centers, and 90 years old or above in Denmark. The families were selected to participate in the study based on The Family Longevity Selection Score (FLoSS) (Sebastiani et al., 2009, PMID: 19910380), a score generated according to birth-year cohort survival probabilities of the proband and siblings; probands and their families with FLoSS score of 7 or higher, at least one living sibling, and at least one living offspring (minimum family size of 3), who were able to give informed consent and willing to participate were recruited. The individuals were genotyped using ~2.3 million SNPs from the Illumina Omni chip, and then imputed on phased 1000 Genomes with Cosmopolitan data as a reference using MACH and MINIMAC. After excluding participants with 80 years and older, ~3,200 individuals have been included in the analyses for replication.

LOLIPOP (London Life Sciences Prospective Population Study): LOLIPOP is a population based prospective study of about 28K Indian Asian and European men and women, recruited from the lists of 58 General Practitioners in West London, United Kingdom between 2003 and 2008 [1]. Indian Asians had all four grandparents born on the Indian subcontinent. Europeans were of self-reported white ancestry. At enrolment all participants completed an interviewer-administered questionnaire for demographic data, medical history, and smoking and alcohol drinking habits. Anthropometric data were collected and blood pressure measured using an Omron 705CP with the mean of three measurements recorded. Blood samples were collected for the measurement of lipid profile after an overnight fasting of at least 8 hours. Aliquots of whole blood were stored at -80C for extraction of genomic DNA. The LOLIPOP study is approved by the local Research Ethics Committees and all participants provided written informed consent.

Loyola GxE (Kingston Gene-by-environment; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Kingston GxE cohort was obtained from a survey conducted in Kingston, Jamaica as part of a larger project to examine gene by environment interactions in the determination of blood pressure among adults 25-74 years [PMID: 9103091]. The principal criterion for eligibility was a body mass index in either the top or bottom third of BMI for the Jamaican population. Participants were identified principally from the records of the Heart Foundation of Jamaica, a non-governmental organization based in Kingston, which provides low-cost screening services (height and weight, blood pressure, glucose, cholesterol) to the general public. Other participants were identified from among participants in family studies of blood pressure at the Tropical Metabolism Research Unit (TMRU) and from among staff members at the University of the West Indies, Mona.

Loyola SPT (Spanish Town; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): Participants were recruited from Spanish Town, a stable, residential urban area neighboring the capital city of Kingston, Jamaica as part of the ICSHIB [PMID: 9103091]. A stratified random sampling scheme was used to recruit adult males and females aged 25–74 years from the general population. Spanish Town was chosen because its demographic make-up was broadly representative of Jamaica as a whole.

METSIM (Metabolic Syndrome In Men): The METSIM Study includes 10,197 men, aged from 45 to 73 years at recruitment, randomly selected from the population register of the Kuopio town, Eastern Finland, and examined in 2005-2010 (Stancakova A, et al. Diabetes 2009). The aim of the study is to investigate genetic and non-genetic factors associated with type 2 diabetes and cardiovascular disease and its risk factors.

Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M: Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6416 Finnish men. Diabetes 58:1212-1221, 2009.

NESDA (Netherlands Study of Depression and Anxiety): NESDA is a multi-center study designed to examine the long-term course and consequences of depressive and anxiety disorders (http://www.nesda.nl). NESDA included both individuals with depressive and/or anxiety disorders and controls without psychiatric conditions. Inclusion criteria were age 18-65 years and self-reported western European ancestry while exclusion criteria were not being fluent in Dutch and having a primary diagnosis of another psychiatric condition (psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use disorder).

OBA (French obese cases): Study of the genetics of obesity in adults.

Pelotas Birth Cohort Study (The 1982 Pelotas Birth Cohort Study, Brazil): The maternity hospitals in Pelotas, a southern Brazilian city (current population ~330,000), were visited daily in the year of 1982. The 5,914 liveborns whose families lived in the urban area were examined and their mothers interviewed. Information was obtained for more than 99% of the livebirths. These subjects have been followed-up at the following mean ages: 11.3 months (all children born from January to Abril 1982; n=1457), 19.4 months (entire cohort; n=4934), 43.1 months (entire cohort; n=4742), 13.1 years (random subsample; n=715), 14.7 years (systematic subsample; n=1076); 18.2 (male cohorts attending to compulsory Army recruitment examination; n=2250), 18.9 (systematic subsample; n=1031), 22.8 years (entire cohort; n=4297) and 30.2 years (entire cohort; n=3701). Details about follow-up visits and available data can be found in the two Cohort Profile papers (PMID: 16373375 and 25733577). DNA samples (collected at the mean age of 22.8 years) were genotyped for ~2.5 million of SNPs using the Illumina HumanOmni2.5-8v1 array (which includes autosomal, X and Y chromosomes, and mitochondrial variants). After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

PREVEND (The Prevention of REnal and Vascular ENd stage Disease study): The PREVEND study is an ongoing prospective study investigating the natural course of increased levels of urinary albumin excretion and its relation to renal and cardiovascular disease. Inhabitants 28 to 75 years of age (n=85,421) in the city of Groningen, The Netherlands, were asked to complete a short questionnaire, 47% responded, and individuals were then selected with a urinary albumin concentration of at least 10 mg/L (n = 7,768) and a randomly selected control group with a urinary albumin concentration less than 10 mg/L (n = 3,395). Details of the protocol have been described elsewhere (Hillege HL et al. Circulation 2002;106:1777-82).

PROCARDIS (Precocious Coronary Artery Disease): The PROCARDIS (European collaborative study of the genetics of precocious coronary artery disease) study is a multi-centre case-control study in which CAD cases and controls were recruited from the United Kingdom, Italy, Sweden and Germany. Cases were defined as symptomatic CAD before age 66 years and 80% of cases also had a sibling in whom CAD had been diagnosed before age 66 years. CAD was defined as clinically documented evidence of myocardial infarction (MI) (80%), coronary artery bypass graft (CABG) (10%), acute coronary syndrome (ACS) (6%), coronary angioplasty (CA) (1%) or stable angina (hospitalization for angina or documented obstructive coronary disease) (3%). The cases included 2,136 cases who were half or full siblings. PROCARDIS controls had no personal or sibling history of CAD before age 66 years.

RHS (Ragama Health Study): The Ragama Health Study (RHS) is a population-based study of South Asian men and women aged 35-64yrs living in the Ragama Medical Officer of Health (MOH) area, near Colombo, Sri Lanka.* Consenting adults attended a clinic after a 12-h fast with available health records, and were interviewed by trained personnel to obtain information on medical, sociodemographic, and lifestyle variables. A 10-mL sample of venous blood was obtained from each subject. The concurrent study was performed in two tea plantation estates in the Lindula MOH area, near Nuwara Eliya (180 km from Colombo), to investigate the gene-environment interaction in a community with differing lifestyles (e.g., physical activity and diet).. The RHS is a collaborative effort between the Faculty of Medicine, University of Kelaniya and the National Center for Global Health and Medicine, Japan.

*Reference: Dassanayake, A.S. et al. Prevalence and risk factors for non-alcoholic fatty liver disease among adults in an urban Sri Lankan population. J Gastroenterol Hepatol 24, 1284-8 (2009).

SHEEP (Stockholm Heart Epidemiology Project): The SHEEP is a population based case-control study of risk factors for first episode of acute myocardial infarction. The study base comprised all Swedish citizens resident in the Stockholm county 1992-1994 who were 45-70 years of age and were free of previous clinically diagnosed myocardial infarction.

Cases were identified using three different sources: 1) coronary units and internal medicine wards for acute care in all Stockholm hospitals; 2) the National Patient Register; and 3) death certificates. For the present study, only cases who survived at least 28 days were considered (n=1213).

First time incident myocardial infarction cases (n=1213) were identified during a 2-year period (1992-1993) for men and during a 3-year period (1992-1994) for women. Controls (n=1561) were randomly recruited from the study population continuously over time within 2 days of the case occurrence and matched to cases on age (5-years interval), sex and hospital catchment area using computerized registers of the population of Stockholm. Five control candidates were sampled simultaneously to be able to replace potential non-respondent controls. Occasionally, because of late response of the initial control, both the first and alternative controls were considered resulting in the inclusion of more controls than cases. Postal questionnaires covering a wide range of exposure areas including occupational exposures, life style factors, social factors and health related factors were distributed to the participants. Clinical investigations were performed at least three months after myocardial infarction of cases and their matched controls. The investigations included blood samplings under fasting conditions with collection of whole blood for DNA extraction, serum and plasma. A biobank was established containing DNA, serum and plasma.

Exposure information based on both the questionnaire and biological data from the health examination was available for 78% of the male and 67% of the female non-fatal cases; the corresponding figures for their controls were 68% and 64%.

SHIP (Study of Health in Pomerania): The Study of Health In Pomerania (SHIP) is a prospective longitudinal population-based cohort study in Mecklenburg-West Pomerania assessing the prevalence and incidence of common diseases and their risk factors (PMID: 20167617). SHIP encompasses the two independent cohorts SHIP and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study centre for a computer-assisted personal interviews and extensive physical examinations. The study protocol was approved by the medical ethics committee of the University of Greifswald. Oral and written informed consents was obtained from each of the study participants

Genome-wide SNP-typing was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 or the Illumina Human Omni 2.5 array (SHIP-TREND samples). Array processing was carried out in accordance with the manufacturer's standard recommendations. Genotypes were determined using GenomeStudio Genotyping Module v1.0 (GenCall) for SHIP-TREND and the Birdseed2 clustering algorithm for SHIP. Imputation of genotypes in SHIP and SHIP-TREND was performed with the software IMPUTE v2.2.2 based on 1000 Genomes release March 2012.

SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study): The Shanghai Women's Health Study (SWHS) is an ongoing population-based cohort study of approximately 75,000 women who were aged 40-70 years at study enrollment and resided in in urban Shanghai, China; 56,832 (75.8%) provided a blood

samples. Recruitment for the SWHS was initiated in 1997 and completed in 2000. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), menstrual history, residential history, occupational history, and family history of cancer. Included in the current project were 892 women who had GWAS data and lipids data.

The Shanghai Men's Health Study (SMHS) is an ongoing population-based cohort study of 61,480 Chinese men who were aged between 40 and 74 years, were free of cancer at enrollment, and lived in urban Shanghai, China; 45,766 (74.4%) provided a blood samples. Recruitment for the SMHS was initiated in 2002 and completed in 2006. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), residential history, occupational history, and family history of cancer. Included in the current project were 298 men who had GWAS data and lipids data.

Genotyping and imputation: Genomic DNA was extracted from buffy coats by using a Qiagen DNA purification kit (Valencia, CA) or Puregene DNA purification kit (Minneapolis, MN) according to the manufacturers' instructions and then used for genotyping assays. The GWAS genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6.0) platform or Illumina 660, following manufacturers' protocols. After sample quality control, we exclude SNPs with 1) MAF <0.01; 2) call rate <95%; 2) bad genotyping cluster; and 3) concordance rate <95% among duplicated QC samples. Genotypes were imputed using the program MACH (http://www.sph.umich.edu/csg/abecasis/MACH/download/), which determines the probable distribution of missing genotypes conditional on a set of known haplotypes, while simultaneously estimating the fine-scale recombination map. Phased autosome SNP data from HapMap Phase II Asians (release 22) were used as the reference. To test for associations between the imputed SNP data with BMI, linear regression (additive model) was used, in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty of genotype imputation (http://www.sph.umich.edu/csg/abecasis/MACH/download/).

The lipid profiles were measured at Vanderbilt Lipid Laboratory. Total cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides (TG) were measured using an ACE Clinical Chemistry System (Alfa Wassermann, Inc, West Caldwell, NJ). Low-density lipoprotein (LDL) cholesterol levels were calculated by using the Friedwald equation. The levels of LDL cholesterol were directly measured using an ACE Clinical Chemistry System for subjects with TG levels ≥ 400 mg/dL. Fasting status was defined as an interval between the last meal and blood draw of 8 hours or longer.

TAICHI-G: The TaiChi consortium consists of 7 studies that collaborated initially in a large scale metabochip study, and became an ongoing consortium for studies of cardiometabolic disease in the Chinese population in Taiwan. The seven studies included the following: 1) HALST (Healthy Aging Longitudinal Study in Taiwan), a population based epidemiologic study of older adults living in all major geographic regions of Taiwan established by the Taiwan National Health Research Institutes (NHRI); 2) SAPPHIRe (Stanford-Asian Pacific Program in Hypertension and Insulin Resistance), a family based study established in 1995 with an initial goal of identifying major genetic loci underlying hypertension and insulin resistance in East Asian populations, with Taiwan subjects participating in the TaiChi consortium; 3) TCAGEN (Taiwan Coronary Artery Disease GENetic), a cohort study that that enrolled patients undergoing coronary angiography or percutaneous intervention at the National Taiwan University Hospital (NTUH) in the setting of either stable angina pectoris or prior myocardial infarction; 4) TACT (TAiwan Coronary and Transcatheter intervention), a cohort study enrolled patients with angina pectoris and objective documentation of myocardial ischemia who underwent diagnostic coronary angiography and/or revascularization any time after October 2000 at the National Taiwan University Hospital (NTUH) (similar to TCAGEN but recruitment was independent of TCAGEN); 5) Taiwan DRAGON (Taiwan Diabetes and RelAted Genetic COmplicatioN), acohort study of Type 2 diabetes at Taichung Veterans General Hospital (Taichung VGH) in Taiwan, with participants including individuals with either newly diagnosed or established diabetes

(subjects with hyperglycemia who did not meet diagnostic criteria for Type 2 DM were not included); 6) TCAD (Taichung CAD study), includes patients with a variety of cardiovascular diseases who received care at the Taichung Veterans General Hospital (Taichung VGH), i.e. specifically individuals who were hospitalized for diagnostic and interventional coronary angiography examinations and treatment; 7) TUDR (Taiwan US Diabetic Retinopathy) enrolled subjects with Type 2 diabetes who received care at Taichung Veteran General Hospital (Taichung VGH), and a small number of subjects from Taipei Tri-Service General Hospital (TSGH); TUDR subjects underwent a complete ophthalmic and fundus examination to carefully document the presence and extent of retinopathy. From these 7 studies, samples for over 1,800 subjects were selected based on completeness of standard metabolic phenotyping and knowledge of cardiac disease status, to undergo GWAS genotyping with an Illumina human-omni 'chip' specific for Asian population (Illumina, San Diego, CA; cat. No. 20004337), hence TAICHI-G.

THRV (Taiwan study of Hypertensives Rare Variants): THRV proposed to identify rare and low frequency genetic variants for blood pressure and hypertension through whole exome sequencing of a subset of highly enriched Taiwan Chinese hypertensive families and as many matched controls. The Taiwan Chinese families (approximately N=1,200 subjects) were previously recruited as part of the NHLBI-sponsored SAPPHIRe Network which is part of the Family Blood Pressure Program (FBPP). The SAPPHIRe families were recruited to have multiple hypertensive sibs and some of them also included one normotensive/hypotensive sib. The matched controls (N=1,200) were selected from the large population-based HALST Study and a Hospital-based population, both in Taipei, Taiwan.

TRAILS (Tracking Adolescents' Individual Lives Survey): TRAILS is a prospective cohort study of Dutch adolescents and young adults, with bi- or triennial measurements from age 11 onwards, which started in 2001. TRAILS consists of a general population and a clinical cohort (https://www.trails.nl/en/home). In the population cohort, six assessment waves have been completed to date, at mean ages 11.1 (SD = 0.6), 13.6 (SD = 0.5), 16.3 (SD = 0.7), 19.1 (SD = 0.6), 22.3 (SD = 0.6), and 25.8 (SD = 0.6). Data for the present study were collected in the population cohort only, during the third assessment wave. The study was approved by the Dutch Central Committee on Research Involving Human Subjects.

TUDR (Taiwan-US Diabetic Retinopathy): 2009 to present, is a cohort that enrolled subjects with Type 2 diabetes receiving care at Taichung Veteran General Hospital (Taichung VGH), and a small number of subjects from Taipei Tri-Service General Hospital. All TUDR subjects underwent a complete ophthalmic and fundus examination to carefully document the presence and extent of retinopathy.

TWINGENE (TwinGene of the Swedish Twin Registry): The aim of the TwinGene project has been to systematically transform the oldest cohorts of the Swedish Twin Registry (STR) into a molecular-genetic resource. Beginning in 2004, about 200 twins were contacted each month until the data collection was completed in 2008. A total of 21 500 twins were contacted where of 12 600 participated. Invitations to the study contained information of the study and its purpose. Along with the invitations consent forms and health questionnaire were sent to the subjects. When the signed consent forms where returned, the subjects were sent blood sampling equipment and asked to contact a local health facility for blood sampling. The study population was recruited among twins participating in the Screening Across the Lifespan Twin Study (SALT) which was a telephone interview study conducted in 1998-2002. Other inclusion criteria were that both twins in the pair had to be alive and living in Sweden. Subjects were excluded from the study if they preciously declined participation in future studies or if they had been enrolled in other STR DNA sampling projects. The subjects were asked to make an appointment for a health check-up at their local health-care facility on the morning Monday to Thursday and not the day before a national holiday, this to ensure that the sample would reach the KI biobank the following morning by overnight mail. The subjects were instructed to fast from 20.00 the previous night. By venipuncture a total of 50 ml of blood was drawn from each subject. Tubes with serum and blood for biobanking as well as for clinical chemistry tests were sent to KI by overnight mail. One 7ml EDTA tube of whole blood is stored in -80°C while

a second 7ml EDTA tube of blood is used for DNA extraction using Puregene extraction kit (Gentra systems, Minneapolis, USA). After excluding subjects in which the DNA concentration in the stock-solution was below 20ng/µl as well as subset of 302 female monozygous twin pairs participating in a previous genome wide effort DNA from 9896 individual subjects was sent to SNP&SEQ Technology Platform Uppsala, Sweden for genome wide genotyping with Illumina OmniExpress bead chip (all available dizygous twins + one twin from each available MZ twin pair). For this project, smoking status was determined during SALT (1998-2002), while blood draw for lipids measurement was conducted between 2004 and 2008.

UKHLS (Understanding Society / The UK Household Longitudinal Study): The United Kingdom Household Longitudinal Study, also known as Understanding Society (https://www.understandingsociety.ac.uk) is a longitudinal panel survey of 40.000 UK households (England, Scotland, Wales and Northern Ireland) representative of the UK population. Participants are surveyed annually since 2009 and contribute information relating to their socioeconomic circumstances, attitudes, and behaviours via a computer assisted interview. The study includes phenotypical data for a representative sample of participants for a wide range of social and economic indicators as well as a biological sample collection encompassing biometric, physiological, biochemical, and haematological measurements and self-reported medical history and medication use. The United Kingdom Household Longitudinal Study has been approved by the University of Essex Ethics Committee and informed consent was obtained from every participant. For this project, smoking status was determined at wave 2, while blood draws were conducted 5 months after wave 2 and 3 interviews; thus smoking status and lipid measurements were between 5 and 17 months apart.

YFS (The Cardiovascular Risk in Young Finns Study): The YFS is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-11 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent.

ADDITIONAL ACKNOWLEDGMENTS: This work utilized the computational resources of the NIH HPC Biowulf cluster. AEJ was supported in part by the American Heart Association (13POST16500011) and NIH (2T32HL007055, K99HL130580). J.R.O'C. was supported in part by a grant from the NIH (U01 HL137181).

STAGE 1 STUDY ACKNOWLEDGMENTS:

AGES (Age Gene/Environment Susceptibility Reykjavik Study): This study has been funded by NIH contract N01-AG012100, HSSN271201200022C, the NIA Intramural Research Program, an Intramural Research Program Award (ZIAEY000401) from the National Eye Institute, an award from the National Institute on Deafness and Other Communication Disorders (NIDCD) Division of Scientific Programs (IAA Y2-DC_1004-02), Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.

ARIC (Atherosclerosis Risk in Communities) study: The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

Bio*Me* Biobank (Bio*Me* Biobank of Institute for Personalized Medicine at Mount Sinai): The Mount Sinai IPM Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies.

CARDIA (Coronary Artery Risk Development in Young Adults): The CARDIA Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed and approved by CARDIA for scientific content.

CHS (Cardiovascular Health Study): This Cardiovascular Health Study (CHS) research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, HHSN268201800001C N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, U01HL130114, R01HL087652, R01HL105756, R01HL103612, R01HL085251, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern

California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CROATIA-Korcula: We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools and the Croatian Institute for Public Health. We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the participants. The SNP genotyping for the CROATIA-Korcula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. CROATIA-Korcula (CR-Korcula) was funded by the Medical Research Council UK, The Croatian Ministry of Science, Education and Sports (grant 216-1080315-0302), the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947), the Croatian Science Foundation (grant 8875) and the Centre of Competencies for Integrative Treatment, Prevention and Rehabilitation using TMS. JEH is now an NIH Visiting Fellow (Post-Doctoral), NHLBI Population Sciences Branch, The Framingham Heart Study, but completed this work at the Medical Research Council Human Genetics Unit at the University of Edinburgh.

CROATIA-Vis: We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health. The SNP genotyping for the CROATIA-Vis cohort was performed in the core genotyping laboratory of the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, Scotland. CROATIA-Vis (CR-Vis) was funded by the Medical Research Council UK, The Croatian Ministry of Science, Education and Sports (grant 216-1080315-0302) and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947).

ERF (Erasmus Rucphen Family study): The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). The ERF study was further supported by ENGAGE consortium and CMSB. High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). ERF was further supported by the ZonMw grant (project 91111025). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, P. Snijders for his help in data collection and E.M. van Leeuwen for genetic imputation. The ERF study was supported by CardioVasculair Onderzoek Nederland (CVON2012-03), Netherlands Heart Foundation and the innovation programme Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) under the grant agreement No 645740 as part of the Personalized pREvention of Chronic DIseases (PRECeDI) project.

FamHS (Family Heart Study): The FamHS is funded by R01HL118305 and R01HL117078 NHLBI grants, and 5R01DK07568102 and 5R01DK089256 NIDDK grant.

FHS (Framingham Heart Study): This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract Nos. N01-HC-25195 and HHSN268201500001I) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for

Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. This research was also supported by grant R01 HL118305 (PI: D.C. Rao). Training grant T32 GM 074905 supported the contribution of Elise Lim.

GENOA (Genetic Epidemiology Network of Arteriopathy): Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL119443, HL118305, HL054464, HL054457, HL054481, HL071917 and HL087660) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (Stephen T. Turner, MD, Mariza de Andrade PhD, Julie Cunningham, PhD). We thank Eric Boerwinkle, PhD and Megan L. Grove from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA for their help with genotyping. We would also like to thank the families that participated in the GENOA study.

GenSalt (Genetic Epidemiology Network of Salt Sensitivity): The Genetic Epidemiology Network of Salt Sensitivity is supported by research grants (U01HL072507, R01HL087263, and R01HL090682) from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

GS:SFHS: Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89). We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span): The Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol number 09-AG-N248). Data analyses for the HANDLS study utilized the high-performance computational resources of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov; http://hpc.nih.gov)).

Health ABC (Health, Aging, and Body Composition): Health ABC was funded by the National Institutes of Aging. This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The GWAS was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

HERITAGE (Health, Risk Factors, Exercise Training and Genetics): The HERITAGE Family Study was supported by National Heart, Lung, and Blood Institute grant HL-45670.

HUFS (Howard University Family Study): The Howard University Family Study was supported by National Institutes of Health grants S06GM008016-320107 to Charles Rotimi and S06GM008016-380111 to Adebowale Adeyemo. We thank the participants of the study, for which enrollment was carried out at the Howard University General Clinical Research Center, supported by National Institutes of Health grant 2M01RR010284. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of the National Institutes of Health. This research was supported in part by the Intramural Research Program of 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). Genotyping support was provided by the Coriell Institute for Medical Research.

HyperGEN (Hypertension Genetic Epidemiology Network): The hypertension network is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah: (Network Coordinating Center, Field Center, and Molecular Genetics Lab); Univ. of Alabama at Birmingham: (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin: (Echo Genotyping Lab); Boston University: (Field Center); University of Minnesota: (Field Center and Biochemistry Lab); University of North Carolina: (Field Center); Washington University: (Data Coordinating Center); Weil Cornell Medical College: (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators: http://www.biostat.wustl.edu/hypergen/Acknowledge.html

JHS (Jackson Heart Study): The Jackson Heart Study is supported by contracts HSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute on Minority Health and Health Disparities. The authors acknowledge the Jackson Heart Study team institutions (University of Mississippi Medical Center, Jackson State University and Tougaloo College) and participants for their long-term commitment that continues to improve our understanding of the genetic epidemiology of cardiovascular and other chronic diseases among African Americans.

MESA (Multi-Ethnic Study of Atherosclerosis): This research was supported by the Multi-Ethnic Study of Atherosclerosis (MESA) contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, by grant HL071205 and by UL1-DR-001079 from NCRR . Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. This publication was partially developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

NEO (The Netherlands Epidemiology of Obesity study): The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Petra Noordijk, Pat van Beelen and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023). Diana van Heemst was supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757).

RS (Rotterdam Study): The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The Rotterdam study was supported by CardioVasculair Onderzoek Nederland (CVON2012-03), Netherlands Heart Foundation and the innovation programme Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) under the grant agreement No 645740 as part of the Personalized pREvention of Chronic DIseases (PRECeDI) project.

The generation and management of GWAS genotype data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, Marjolein Peters and Carolina Medina-Gomez for their help in creating the GWAS database, and Karol Estrada, Yurii Aulchenko and Carolina Medina-Gomez for the creation and analysis of imputed data.

SCHS-CHD (Singapore Chinese Health Study - Coronary Heart Disease): The Singapore Chinese Health Study is supported by the National Institutes of Health, USA (RO1 CA144034 and UM1 CA182876), the nested case-control study of myocardial infarction by the Singapore National Medical Research Council (NMRC 1270/2010) and genotyping by the HUJ-CREATE Programme of the National Research Foundation, Singapore (Project Number 370062002).

SCES (Singapore Chinese Eye Study), SiMES (Singapore Malay Eye Study), (SINDI) Singapore Indian Eye Study: The Singapore Malay Eye Study (SiMES), the Singapore Indian Eye Study (SINDI), and the Singapore Chinese Eye Study (SCES) are supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616). Ching-Yu Cheng is supported by an award from NMRC (CSA/033/2012). The Singapore Tissue Network and the Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore provided services for tissue archival and genotyping, respectively. **SP2 (Singapore Prospective Study Program):** SP2 is supported by the individual research grant and clinician scientist award schemes from the National Medical Research Council and the Biomedical Research Councils of Singapore.

WGHS (Women's Genome Health Study): The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), with collaborative scientific support and funding for genotyping provided by Amgen.

WHI (Women's Health Initiative): The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

STAGE 2 STUDY ACKNOWLEDGMENTS:

AA-DHS (African American Diabetes Heart Study): The investigators acknowledge the cooperation of our Diabetes Heart Study (DHS) and AA-DHS participants. This work was supported by NIH R01 DK071891, R01 HL092301 and the General Clinical Research Center of Wake Forest School of Medicine M01-RR-07122.

Airwave (The Airwave Health Monitoring Study): We thank all participants in the Airwave Health Monitoring Study. The study is funded by the Home Office (Grant number 780-TETRA) with additional support from the National Institute for Health Research (NIHR), Imperial College Healthcare NHS Trust (ICHNT) and Imperial College Biomedical Research Centre (BRC). The study has ethical approval from the National Health Service Multi-site Research Ethics Committee (MREC/13/NW/0588). This work used computing resources provided by the MRC- funded UK MEDical Bioinformatics partnership programme (UK MED-BIO) (MR/L01632X/1). P.E. would like to acknowledge support from the Medical Research Council and Public Health England for the MRC-PHE Centre for Environment and Health (MR/L01341X/1) and from the NIHR NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141).

ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial): The ASCOT study was supported by Pfizer, New York, NY, USA for the ASCOT study and the collection of the ASCOT DNA repository; by Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. In particular we thank Clare Muckian and David Toomey for their help in DNA extraction, storage, and handling. Genotyping was funded by the CNG, MRC and the National Institutes of Health Research (NIHR). We would also like to acknowledge the Barts and The London Genome Centre staff for genotyping. This work forms part of the research programme of the NIHR Cardiovascular Biomedical Research Unit at Barts and The London, QMUL. H.R.W, M.J.C and P.B.M. wishes to acknowledge the NIHR Cardiovascular Biomedical Research Unit at Barts and The London, Queen Mary University of London, UK for support.

Baependi Heart Study (Brazil): The Baependi Heart Study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (Grant 2013/17368-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Hospital Samaritano Society (Grant 25000.180.664/2011-35), through Ministry of Health to Support Program Institutional Development of the Unified Health System (SUS-PROADI).

BBJ (Biobank Japan Project): BioBank Japan project is supported by the Japan Agency for Medical Research and Development and by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government.

BES (Beijing Eye Study): BES was supported by the National Key Laboratory Fund, Beijing, China.

BRIGHT (British Genetics of Hypertension): This work was supported by the Medical Research Council of Great Britain (grant number G9521010D) and the British Heart Foundation (grant number PG/02/128). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work forms part of the research program of the National Institutes of Health Research (NIHR Cardiovascular Biomedical Research) Cardiovascular Biomedical Unit at Barts and The London, QMUL.

CAGE-Amagasaki (Cardio-metabolic Genome Epidemiology Network, Amagasaki Study): The CAGE Network studies were supported by grants for the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency; the Program for Promotion of Fundamental Studies in Health Sciences, National Institute of Biomedical Innovation Organization (NIBIO); and the Grant of National Center for Global Health and Medicine (NCGM).

CFS (Cleveland Family Study): The CFS was supported by the National Institutes of Health, the National Heart, Lung, Blood Institute grant HL113338, R01HL098433, HL46380.

DESIR (Data from an Epidemiological Study on the Insulin Resistance): The DESIR Study Group is composed of Inserm-U1018 (Paris: B. Balkau, P. Ducimetière, E. Eschwège), Inserm-U367 (Paris: F. Alhenc-Gelas), CHU d'Angers (A. Girault), Bichat Hospital (Paris: F. Fumeron, M. Marre, R. Roussel), CHU de Rennes (F. Bonnet), CNRS UMR-8199 (Lille: A. Bonnefond, P. Froguel), Medical Examination Services (Alençon, Angers, Blois, Caen, Chartres, Chateauroux, Cholet, LeMans, Orléans and Tours), Research Institute for General Medicine (J. Cogneau), the general practitioners of the region and the Cross- Regional Institute for Health (C. Born, E. Caces, M. Cailleau, N. Copin, J.G. Moreau, F. Rakotozafy, J. Tichet, S. Vol).

The DESIR study was supported by Inserm contracts with CNAMTS, Lilly, Novartis Pharma and Sanofi-aventis, and by Inserm (Réseaux en Santé Publique, Interactions entre les déterminants de la santé, Cohortes Santé TGIR 2008), the Association Diabète Risque Vasculaire, the Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Société Francophone du Diabète, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche and Topcon.

DFTJ (Dongfeng-Tongji Cohort Study): This work was supported by grants from the National Basic Research Program grant (2011CB503800), the Programme of Introducing Talents of Discipline, the grants from the National Natural Science Foundation (grant NSFC-81473051, 81522040 and 81230069), and the Program for the New Century Excellent Talents in University (NCET-11-0169).

DHS (Diabetes Heart Study): The authors thank the investigators, staff, and participants of the DHS for their valuable contributions. This study was supported by the National Institutes of Health through HL67348 and HL092301.

DR's EXTRA (Dose Responses to Exercise Training): The study was supported by grants from Ministry of Education and Culture of Finland (722 and 627; 2004-2010); Academy of Finland (102318, 104943, 123885, 211119); European Commission FP6 Integrated Project (EXGENESIS), LSHM-CT-2004-005272; City of Kuopio; Juho Vainio Foundation; Finnish Diabetes Association; Finnish Foundation for Cardiovascular Research; Kuopio University Hospital; Päivikki and Sakari Sohlberg Foundation; Social Insurance Institution of Finland 4/26/2010.

EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)): This study was supported by EU H2020 grants 692145, 676550, 654248, Estonian Research Council Grant IUT20-60 and PUT1660, NIASC, EIT – Health and NIH-BMI Grant No: 2R01DK075787-06A1 and EU through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012 GENTRANSMED).

EPIC (European Prospective Investigation into Cancer and Nutrition)-Norfolk: The genetics work in the EPIC Norfolk Study is funded by the Medical Research Council (MC_PC_13048) and the study as a whole has received funding from the Medical Research Council and Cancer Research, United Kingdom, British Heart Foundation, the Medical Research Council, the Ministry of Agriculture, Fisheries and Food, and the Europe against Cancer Programme of the Commission of the European Communities. We thank all EPIC participants and staff for their contribution to the study.

The EPIC-InterAct Case-Cohort Study: We thank all EPIC participants and staff for their contribution to the study. The InterAct study received funding from the European Union (Integrated Project LSHM-CT-2006-037197 in the Framework Programme 6 of the European Community). We thank staff from the Technical, Field Epidemiology and Data Functional Group Teams of the Medical Research Council Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping and data handling work. We specifically thank Sarah Dawson for coordinating the sample provision for biomarker

measurements, Abigail Britten for coordinating DNA sample provision and genotyping of candidate markers, Nicola Kerrison, Chris Gillson and Abigail Britten for data provision and genotyping quality control, Matt Sims for writing the technical laboratory specification for the intermediate pathway biomarker measurements and for overseeing the laboratory work.

FENLAND (The Fenland Study): The Fenland Study is funded by the Wellcome Trust and the Medical Research Council (MC_U106179471 and MC_UU_12015/1). We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We further acknowledge support from the Medical Research Council (MC_PC_13046MC_UU_12015/1).

FUSION (Finland-United States Investigation of NIDDM Genetics): The FUSION study was supported by DK093757, DK072193, DK062370 and ZIA-HG000024.

Genotyping was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine.

GeneSTAR (Genetic Studies of Atherosclerosis Risk): [for the smoking/lipids and smoking/BP analyses] GeneSTAR was supported by National Institutes of Health grants from the National Heart, Lung, and Blood Institute (HL49762, HL59684, HL58625, HL071025, U01 HL72518, and HL087698), National Institute of Nursing Research (NR0224103), and by a grant from the National Center for Research Resources to the Johns Hopkins General Clinical Research Center (M01-RR000052).

GLACIER (Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk): We thank the participants, health professionals and data managers involved in the Västerbottens Intervention Project. We are also grateful to the staff of the Northern Sweden Biobank for preparing materials and to K Enqvist and T Johansson (Västerbottens County Council, Umeå, Sweden) for DNA preparation. The current study was supported by Novo Nordisk (PWF), the Swedish Research Council (PWF), the Swedish Heart Lung Foundation (PWF), the European Research Council (PWF), and the Skåne Health Authority (PWF).

GRAPHIC (Genetic Regulation of Arterial Pressure of Humans in the Community): The GRAPHIC Study was funded by the British Heart Foundation (BHF/RG/2000004). This work falls under the portfolio of research supported by the NIHR Leicester Cardiovascular Biomedical Research Unit. CPN and NJS are funded by the BHF and NJS is a NIHR Senior Investigator.

HCHS/SOL (Hispanic Community Health Study/ Study of Latinos): The baseline examination of HCHS/SOL was supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements additionally contributed funding to HCHS/SOL. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping was also supported by National Center for Advancing Translational Sciences UL1TR000124 and NIDDK DK063491 to the Southern California Diabetes Endocrinology Research Center. This research was also supported in part by the Intramural Research Program of the NIDDK, contract no. HHSB268201200054C, and Illumina.

HRS (Health & Retirement Study): HRS is supported by the National Institute on Aging (NIA U01AG009740 and R03 AG046389). Genotyping was funded separately by NIA (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.

HyperGEN-AXIOM (Hypertension Genetic Epidemiology Network): The study was support by the National Institutes of Health, the National Heart, Lung, Blood Institute grant HL086718.

INGI-CARL (Italian Network Genetic Isolates): This study was partially supported by Regione FVG (L.26.2008) and Italian Ministry of Health (GR-2011-02349604).

INGI-FVG (Italian Network Genetic Isolates): This study was partially supported by Regione FVG (L.26.2008) and Italian Ministry of Health (GR-2011-02349604).

IRAS (Insulin Resistance Atherosclerosis Study): The IRAS is supported by the National Heart Lung Institute (HL047887, HL047889, HL047890, and HL47902). Genotyping for this study was supported by the GUARDIAN Consortium with grant support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

IRAS Family Study (Insulin Resistance Atherosclerosis Study): The IRASFS is supported by the National Heart Lung and Blood Institute (HL060944, HL061019, and HL060919). Genotyping for this study was supported by the GUARDIAN Consortium with grant support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin): Support for genotype data collection and collaborative genetic analysis in JUPITER was provided by Astra-Zeneca.

KORA (Cooperative Health Research in the Augsburg Region): The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

LBC1936 (Lothian Birth Cohort 1936): We thank the LBC1936 cohort participants and team members who contributed to these studies. Phenotype collection was supported by Age UK (The Disconnected Mind project). Genotyping was funded by the BBSRC (BB/F019394/1). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged.

LifeLines (Netherlands Biobank): The Lifelines Cohort Study, and generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation.

The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

LLFS (The Long Life Family Study): The study is supported by the National Institute on Aging (NIA) grant U01AG023746.

LOLIPOP (London Life Sciences Prospective Population Study): The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381) the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143) and H2020 programs (iHealth-T2D, 643774). We acknowledge support of the MRC-PHE Centre for Environment and Health, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the Imperial College Healthcare NHS Trust, the NHS, the NIHR or the Department of Health. We thank the participants and research staff who made the study possible. John C. Chambers is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017).

Loyola GxE (Kingston Gene-by-environment; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Loyola GxE project was supported by NIH Grant R01HL53353.

Loyola SPT (Spanish Town; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Loyola SPT project was supported by NIH Grant R01HL53353.

METSIM (Metabolic Syndrome In Men): The METSIM study was supported by the Academy of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the European Community (HEALTH-F2-2007 201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and ZIA- HG000024. Genotyping was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine.

NESDA (Netherlands Study of Depression and Anxiety): Funding was obtained from the Netherlands Organization for Scientific Research (Geestkracht program grant 10-000-1002); the Center for Medical Systems Biology (CSMB, NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL), VU University's Institutes for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, University Medical Center Groningen, Leiden University Medical Center, National Institutes of Health (NIH, R01D0042157-01A, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health.Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by NWO.

OBA (French obese cases): The obese French adults were recruited by the laboratory "Integrated Genomics and Metabolic Diseases Modeling" (UMR 8199 CNRS / Université de Lille 2 / Institut Pasteur de Lille) of Pr. Philippe Froguel.

Pelotas Birth Cohort Study (The 1982 Pelotas Birth Cohort Study, Brazil): The 1982 Pelotas Birth Cohort Study is conducted by the Postgraduate Program in Epidemiology at Universidade Federal de Pelotas with the collaboration of the Brazilian Public Health Association (ABRASCO). From 2004 to 2013, the Wellcome Trust supported the study. The International Development Research Center, World Health Organization, Overseas Development Administration, European Union, National Support Program for Centers of Excellence (PRONEX),

the Brazilian National Research Council (CNPq), and the Brazilian Ministry of Health supported previous phases of the study.

Genotyping of 1982 Pelotas Birth Cohort Study participants was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil).

PREVEND (The Prevention of REnal and Vascular ENd stage Disease study): PREVEND genetics is supported by the Dutch Kidney Foundation (Grant E033), the EU project grant GENECURE (FP-6 LSHM CT 2006 037697), the National Institutes of Health (grant 2R01LM010098), The Netherlands organization for health research and development (NWO-Groot grant 175.010.2007.006, NWO VENI grant 916.761.70, ZonMw grant 90.700.441).

PROCARDIS (Precocious Coronary Artery Disease): PROCARDIS was supported by the European Community Sixth Framework Program (LSHM-CT- 2007-037273), AstraZeneca, the British Heart Foundation, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular Program of Karolinska Institutet and Stockholm County Council, the Foundation for Strategic Research and the Stockholm County Council (560283). M.F and H.W acknowledge the support of the Wellcome Trust core award (090532/Z/09/Z) and the BHF Centre of Research Excellence. A.G and H.W acknowledge European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no. HEALTH-F2-2013-601456 (CVGenes@Target) & and A.G, the Wellcome Trust Institutional strategic support fund. This work was supported by the TriPartite Immunometabolism Consortium [TrIC]- Novo Nordisk Foundation's Grant number NNF15CC0018486.

RHS (Ragama Health Study): The RHS was supported by the Grant of National Center for Global Health and Medicine (NCGM).

SHEEP (Stockholm Heart Epidemiology Project): This study was supported by grants from the Swedish Research Council for Health, Working Life and Welfare (http://www.forte.se/en/), the Stockholm County Council (http://www.sll.se/om-landstinget/Information-in-English1/), the Swedish Research Council (http://www.vr.se/inenglish.4.12fff4451215cbd83e4800015152.html), the Swedish Heart and Lung Foundation (https://www.hjart-lungfonden.se/HLF/Om-Hjart-lungfonden/About-HLF/), and the Cardiovascular Programme at Karolinska Institutet (http://ki.se/en/mmk/cardiovascular-research-networks).

SHIP (Study of Health in Pomerania): SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG.

SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study): We thank all the individuals who took part in these studies and all the researchers who have enabled this work to be carried out. The Shanghai Women's Health Study and the Shanghai Men's Health Study are supported by research grants UM1CA182910 and UM1CA173640 from the U.S. National Cancer Institute, respectively.

TAICHI_G: This study was supported by the National Health Research Institutes, Taiwan (PH-100-SP-01, BS-094-PP-01, PH-100-PP-03), the National Science Council, Taiwan (Grant Nos NSC 98-2314-B-075A-002-MY3, NSC 96-2314-B-002-151, NSC 96-2314-B-002-152, NSC 98-2314-B-002-122-MY2, NSC 100-2314-B-002-115, NSC 101-2325-002-078, 101-2314-B-075A-006-MY3), the National Taiwan University Hospital, Taiwan (NTUH 98-N1266, NTUH 100-N1775, NTUH 101-N2010, NTUH 101-N, VN101-04, NTUH 101-S1784), and the staff in the 6th core laboratory for technical support.

THRV (Taiwan study of Hypertensives Rare Variants): The THRV study is supported by National Heart, Lung, and Blood Institute grant R01HL111249.

TRAILS (Tracking Adolescents' Individual Lives Survey): TRAILS (TRacking Adolescents' Individual Lives Survey) is a collaborative project involving various departments of the University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grant 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013); the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities, and Accare Center for Child and Adolescent Psychiatry. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation.

We are grateful to all adolescents who participated in this research and to everyone who worked on this project and made it possible.

TUDR (Taiwan-US Diabetic Retinopathy): This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R. and Y.-D.I.C.) and ARRA Supplement (EY014684-03S1, -04S1), the Eye Birth Defects Foundation Inc., the National Science Council, Taiwan (NSC 98-2314-B-075A-002-MY3 to W.H.S.) and the Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1003001C to W.H.S .). DNA handling and genotyping were supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

TWINGENE (TwinGene of the Swedish Twin Registry): The Swedish Twin Registry is financially supported by Karolinska Institutet. TwinGene project received funding from the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF) and the Heart and Lung foundation no. 20070481

UKHLS (Understanding Society / The UK Household Longitudinal Study): These data are from Understanding Society: The UK Household Longitudinal Study, which is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The data were collected by NatCen and the genome wide scan data were analysed by the Wellcome Trust Sanger Institute. The Understanding Society DAC have an application system for genetics data and all use of the data should be approved by them. The application form is at:

https://www.understandingsociety.ac.uk/about/health/data.

YFS (The Cardiovascular Risk in Young Finns Study): The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

The expert technical assistance in the statistical analyses by Leo-Pekka Lyytikäinen and Irina Lisinen is gratefully acknowledged.

1. The Wellcome Trust Sanger Institute, Cambridge, United Kingdom.
2. University of Cambridge Metabolic Research Laboratories, Cambridge, UK.
3. Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, UK.
4. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
5. Oxford NIHR Biomedical Research Centre, Oxford, UK.
6. Catalan Institute of Oncology (ICO), Barcelona, Spain.
7. CESP, University Paris-South, Faculty of Medicine, University Versailles-St Quentin, Inserm U1018, University ParisSaclay, Villejuif, France.
8. Navarre Public Health Institute (ISPN), Pamplona, Spain.
9. Unidad de Investigación Médica en Epidemiologia Clinica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 06720, Mexico.
10. Navarra Institute for Health Research (IdiSNA) Pamplona, Spain.
11. CIBER de Epidemiología y Salud Pública (CIBERESP), Spain.
12. German Institute of Human Nutrition Potsdam-Rehbruecke, Germany.
13. Public Health Division of Gipuzkoa, San Sebastian, Spain.
14. Lund University, Malmö, Sweden.
15. Umeå University, Umeå, Sweden.
16. MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom.
17. Epidemiology and Prevention Unit, Milan, Italy.
18. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
19. University of Oxford, United Kingdom, (20) Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
20. Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Spain.
21. Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark.
22. Aalborg University Hospital, Aalborg, Denmark.
23. Cancer Risk Factors and Life-Style Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence, Italy.
24. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy.
25. Public Health Directorate, Asturias, Spain.
26. Unit of Cancer Epidemiology, Citta' della Salute e della Scienza HospitalUniversity of Turin and Center for Cancer Prevention (CPO), Torino, Italy.
27. Human Genetics Foundation (HuGeF), Torino, Italy.
28. Andalusian School of Public Health, Granada, Spain.
29. Instituto de Investigación Biosanitaria de Granada (Granada.ibs), Granada (Spain).
30. International Agency for Research on Cancer, Lyon, France.
31. National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
32. Danish Cancer Society Research Center, Copenhagen, Denmark.
33. AIRE - ONLUS, Ragusa, Italy.

34. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.
35. School of Public Health, Imperial College London, UK.
36. Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
37. Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
38. Department of Pediatrics, Division of Genetic and Genomic Medicine, University of California, Irvine, Orange, CA 92868, USA.
39. Departments of Neurology and Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL 33136, USA.
40. Evelyn F McKnight Brain Institute, Miller School of Medicine, University of Miami, Miami, FL 33136, USA.
41. Jackson Memorial Hospital, University of Miami, Miami, FL 33136-1096, USA.
42. Institute of Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany.
43. Department of Research and Education, Kaiser Permanente Southern California, Pasadena, CA 91101, USA.
44. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94309, USA.
45. Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94309, USA.
46. Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305, USA. .
47. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 751 85, Sweden. .
48. Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA. .
49. Instituto BIO-Donostia, Basque Government, San Sebastian, Spain.
50. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Suita 565-0871, Japan.
51. Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan.
52. Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada.
53. Division of Medicine, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK. .