ORIGINAL ARTICLE

Loss-of-Function Mutations in *APOC3*, Triglycerides, and Coronary Disease

The TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute*

ABSTRACT

BACKGROUND

Plasma triglyceride levels are heritable and are correlated with the risk of coronary heart disease. Sequencing of the protein-coding regions of the human genome (the exome) has the potential to identify rare mutations that have a large effect on phenotype.

METHODS

We sequenced the protein-coding regions of 18,666 genes in each of 3734 participants of European or African ancestry in the Exome Sequencing Project. We conducted tests to determine whether rare mutations in coding sequence, individually or in aggregate within a gene, were associated with plasma triglyceride levels. For mutations associated with triglyceride levels, we subsequently evaluated their association with the risk of coronary heart disease in 110,970 persons.

RESULTS

An aggregate of rare mutations in the gene encoding apolipoprotein C3 (*APOC3*) was associated with lower plasma triglyceride levels. Among the four mutations that drove this result, three were loss-of-function mutations: a nonsense mutation (R19X) and two splice-site mutations (IVS2+1G→A and IVS3+1G→T). The fourth was a missense mutation (A43T). Approximately 1 in 150 persons in the study was a heterozygous carrier of at least one of these four mutations. Triglyceride levels in the carriers were 39% lower than levels in noncarriers ($P<1\times10^{-20}$), and circulating levels of APOC3 in carriers were 46% lower than levels in noncarriers ($P=8\times10^{-10}$). The risk of coronary heart disease among 498 carriers of any rare *APOC3* mutation was 40% lower than the risk among 110,472 noncarriers (odds ratio, 0.60; 95% confidence interval, 0.47 to 0.75; $P=4\times10^{-6}$).

CONCLUSIONS

Rare mutations that disrupt *APOC3* function were associated with lower levels of plasma triglycerides and APOC3. Carriers of these mutations were found to have a reduced risk of coronary heart disease. (Funded by the National Heart, Lung, and Blood Institute and others.)

Address reprint requests to Dr. Sekar Kathiresan at the Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge St., CPZN 5.252, Boston, MA 02114, or at skathiresan@partners.org.

*The members of the Triglycerides and High-Density Lipoprotein (TG and HDL) Working Group and their affiliations are listed in the Appendix. Ms. Jacy Crosby and Drs. Gina Peloso, Paul L. Auer, Alex P. Reiner, Eric Boerwinkle, and Sekar Kathiresan contributed equally to this article and assume responsibility for its content and integrity.

This article was published on June 18, 2014, at NEJM.org.

DOI: 10.1056/NEJMoa1307095

In observational studies, plasma triglyceride levels are associated with the risk of coronary heart disease.[1,2] Heritability accounts for more than 50% of the individual variation in triglyceride levels.[3] Genomewide association studies have identified common DNA sequence variants at more than 150 genetic loci that are related to plasma lipids[4,5] and have suggested that plasma triglyceride-rich lipoproteins directly influence the risk of coronary heart disease.[6] These findings lead to two unanswered questions: first, to what extent do rare DNA sequence variants, particularly those in protein-coding sequences, contribute to individual variation in plasma triglyceride levels and the risk of coronary heart disease at the population level, and second, are there specific genetic variants that might lower triglyceride levels and reduce the risk of coronary heart disease?

Recent advances in DNA sequencing technology allow comprehensive detection of rare DNA sequence variants. When sequencing is performed in large populations, a sufficient number of mutation carriers can be identified to evaluate the correlation of genotype with phenotype. In particular, it is advantageous to focus sequencing on exons, the elements of the genome that code for proteins (collectively called the exome),[7,8] since mutations in protein-coding sequences (e.g., missense, nonsense, or splice-site mutations) are most readily interpreted.

To address the two questions posed above, we sequenced the exomes of 3734 persons in the United States, identified mutations, and tested the mutations for association with plasma triglyceride levels. We subsequently investigated whether the same mutations were related to the risk of clinical coronary heart disease.

Methods

Study Design and Oversight

The study was conducted as part of the Exome Sequencing Project of the National Heart, Lung, and Blood Institute (NHLBI) (https://esp.gs.washington.edu/drupal).[9,10] The study was designed by the Triglycerides and High-Density Lipoprotein (TG and HDL) Working Group of the Exome Sequencing Project (see the Appendix). The institutional review boards at the Broad Institute and all participating sites approved the NHLBI Exome Sequencing Project study protocols. The last two authors vouch for the accuracy and completeness of the data and all analyses.

Discovery Cohorts

The Exome Sequencing Project study population includes enrollees in seven population-based cohorts (Atherosclerosis Risk in Communities,[11] Coronary Artery Risk Development in Young Adults,[12] the Cardiovascular Health Study,[13] the Framingham Heart Study,[14] the Jackson Heart Study,[15] the Multiethnic Study of Atherosclerosis,[16] and the Women's Health Initiative[17]) and participants in a study of early-onset myocardial infarction (Myocardial Infarction Genetics Consortium[18]) (see Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org).

All participants provided consent for participation in genetic studies. Total cholesterol, HDL cholesterol, and triglyceride levels were measured in fasting venous blood samples, according to standard procedures. Low-density lipoprotein (LDL) cholesterol values were calculated according to the Friedewald formula, with the assignment of a missing value to persons with triglyceride levels higher than 400 mg per deciliter (4.5 mmol per liter).

Exome Sequencing

Exome sequencing was performed at either the University of Washington or the Broad Institute. DNA sequence variants identified at either institution were then submitted to the University of Michigan, where they were subjected to quality control and filtering to exclude variants that were likely to be false positive findings. Protocols for exome sequencing and variant analysis are described in the Supplementary Appendix.

Primary Analyses

Sequence Variants and Triglycerides

We used linear regression to test the association of plasma triglyceride levels with identified DNA sequence variants in race-specific models and in models including multiple races. The outcome variable was the natural logarithm of the triglyceride level. The primary independent variable was the number of copies of the variant allele (coded as 0, 1, or 2). Covariates included age, sex, two principal components of ancestry, an indicator variable for race (in the model in which races were combined), and indicator variables for the sequencing ascertainment scheme (see page 45 of the Supplementary Appendix for further in-

formation on the covariate of age). The single-variant association analysis was limited to variants with a frequency above 0.1%, corresponding to variants with approximately seven copies of the minor allele among the 3734 participants.

Since many of the identified variants occurred too infrequently to be analyzed individually, we aggregated rare variants within each gene; here, the gene (rather than an individual variant) became the unit of analysis. Within each gene, we included all missense, nonsense, and splice-site variants below a frequency threshold of 1% and considered whether plasma triglyceride levels differed between carriers of these mutations and noncarriers.[19] Owing to considerations regarding statistical power, we limited the analysis to genes for which five or more persons carried minor alleles.

Replication of Association between APOC3 *Genotypes and Plasma Lipids*

The gene most strongly associated with plasma triglyceride levels in the discovery sample was the gene encoding apolipoprotein C3 (*APOC3*). To replicate this finding, we genotyped four *APOC3* mutations (R19X, IVS2+1G→A, A43T, and IVS3+1G→T), using the HumanExome BeadChip (Illumina) in 41,671 participants of African or European ancestry in seven replication studies (Table S2 in the Supplementary Appendix). This cohort was separate from the discovery cohort. We tested the association of these four mutations with plasma lipid levels. The methods used for genotyping and association testing are described in the Supplementary Appendix.

APOC3 *Genotypes,* APOC3 *Levels, and Coronary Heart Disease*

We next tested the association of *APOC3* mutations with coronary heart disease in participants of European, African, and Hispanic ancestry in 15 studies (for details, see Table S3 in the Supplementary Appendix). To investigate the association between APOC3 level and the risk of coronary heart disease, we used blood samples drawn from 3238 fasting participants in the Framingham Heart Study offspring cohort (examination cycle 5, 1991 to 1995).[20] Plasma APOC3 levels were measured with the use of a commercially available immunochemical assay (Wako Diagnostics). All participants underwent continual surveillance for incident coronary heart disease events until December 31, 2010. As described in the Supplementary Appendix, we used proportional-hazards regression models to examine the relationship of plasma APOC3 levels (logarithmically transformed on a natural log scale) to the risk of incident coronary heart disease.

SECONDARY ANALYSES

In patients with angiographic coronary heart disease, we evaluated the association of plasma APOC3 levels with total and cardiovascular mortality. We also tested the association of *APOC3* mutations with the presence of hepatic fat on computed tomography (CT). Details of the methods are presented in the Supplementary Appendix.

STATISTICAL ANALYSIS

For the single-variant analysis, we considered a P value of less than 5×10^{-7} to be significant because this threshold accounts for the 95,342 variants tested. For the gene-level analysis, we considered a P value of less than 3×10^{-6} to be significant to account for the testing of 18,666 genes. In the analyses of associations with coronary heart disease, we tested four variants and thus considered a P value of less than 0.0125 to be significant. Genetic association analyses were performed with use of PLINK (http://pngu.mgh.harvard.edu/~purcell/plink),[21] SCORE-Seq (http://dlin.web.unc.edu/software/SCORE-Seq),[22] SAS software, version 9.1, or R software, version R=2.14 (www.r-project.org).

RESULTS

BASELINE CHARACTERISTICS OF THE STUDY PARTICIPANTS

Among the 6823 persons for whom sequencing was performed through the NHLBI Exome Sequencing Project, 2861 were excluded because data on fasting triglyceride levels were not available, 149 were excluded because of relatedness, and 79 were excluded because of sample rejection during the quality-control process. These exclusions left 3734 persons eligible for the present investigation (the baseline characteristics of these participants are shown in Table S1 in the Supplementary Appendix). There were 2082 participants of European ancestry and 1652 of African ancestry.

EXOME SEQUENCING

We obtained sequence data for 256,143 exons from 18,666 genes. On average, targeted bases were read 89 times and 94% of targeted bases were covered

with at least 20 sequencing reads. Each participant had an average of 46 mutations causing the gain or loss of a stop codon, 5712 had missense mutations, and 19 had mutations that altered canonical splice sites.

ASSOCIATION OF SINGLE VARIANTS WITH TRIGLYCERIDES

We tested the association of plasma triglyceride levels with 95,342 nonsense, missense, or splice-site mutations with an allele frequency greater than 0.1% (Fig. S1 and Tables S4, S5, and S6 in the Supplementary Appendix). In analyses of the full sample, the variant most strongly associated with triglyceride levels was the previously described S19W variant in *APOA5*,[23] which had an allele frequency of 6% among participants of European ancestry ($P=3\times10^{-7}$).

GENE-LEVEL ASSOCIATIONS WITH TRIGLYCERIDES AND REPLICATION

We next assessed whether rare mutations (i.e., missense, nonsense, or splice-site variants with a frequency of <1%) tested collectively within each gene were associated with plasma triglyceride levels. *APOC3* was the gene most strongly associated with plasma triglyceride levels when analyses included only participants of European ancestry ($P=7\times10^{-6}$) and when analyses included participants of European ancestry and participants of African ancestry ($P=1\times10^{-5}$) (Table S7 in the Supplementary Appendix). An aggregate of rare alleles at *APOC3* was associated with lower plasma triglyceride levels (Table 1 and Fig. 1). As compared with noncarriers, carriers of any rare *APOC3* mutation had plasma triglyceride levels that were 39% lower ($P=6\times10^{-9}$), HDL cholesterol levels that were 22% higher ($P=3\times10^{-6}$), and LDL cholesterol levels that were 16% lower ($P=0.05$) (Table 1). These rare mutations collectively explained 0.9% of the variance in triglyceride levels.

The association between plasma lipid levels and *APOC3* was primarily attributable to four different sites within the gene. Three of the four mutations are predicted to severely disrupt the function of APOC3 — that is, lead to loss of function.[24] Loss-of-function variants included

Table 1. Rare Mutations in *APOC3* and Plasma Levels of Triglycerides.*

Chromosome Position	Mutation	Mutation Type	No. of Carriers	Mean Triglyceride Level	Mean HDL Cholesterol Level	Mean LDL Cholesterol Level
					mg/dl	
11:116701284	IVS1–2G→A	Splice site	1	66.0	65.0	140.0
11:116701326	A10T	Missense	2	71.5	64.0	138.5
11:116701353	R19X	Nonsense	2	45.5	56.8	152.5
11:116701354	IVS2+1G→A	Splice site	10	71.0	59.6	111.7
11:116701560	A43T	Missense	12	79.0	64.5	110.4
11:116701613	IVS3+1G→T	Splice site	5	82.8	67.0	154.2
11:116703493	D65 N	Missense	1	415.0	32.0	NA
Total *APOC3* carriers			33	84.5±86.0	61.9±16.6	122.8±54.8
Total *APOC3* noncarriers			3701	137.5±90.1	50.7±15.6	146.2±61.8
Carriers vs. noncarriers (%)				−38.5	22.1	−16.0
P value†				6×10^{-9}‡	4×10^{-6}	0.05

* Plus–minus values are means ±SD. The *APOC3* mutations were identified by means of exome sequencing. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for triglycerides to millimoles per liter, multiply by 0.01129. HDL denotes high-density lipoprotein, LDL low-density lipoprotein, and NA not available.

† P values were derived from a linear regression model, with adjustments for age, sex, ancestry, and principal components of ancestry.

‡ The P value for the triglyceride phenotype is based on triglyceride levels that were logarithmically transformed on a natural log scale.

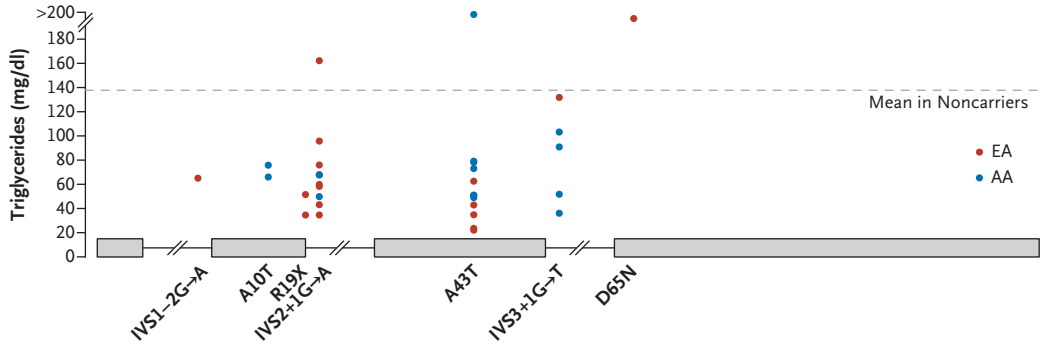


Figure 1. Rare Mutations in *APOC3* and Reductions in Plasma Triglyceride Levels.

A schematic diagram of the gene encoding apolipoprotein C3 (*APOC3*) shows four exons and three introns, or intervening sequences (IVS). In the graph, each dot represents a single study participant who carried a specific mutation. Red dots denote persons of European ancestry (EA), and blue dots persons of African ancestry (AA). The mean level of triglycerides among all persons who did not carry any mutation in the coding sequence of *APOC3* was 137 mg per deciliter (dashed line). To convert the values for triglycerides to millimoles per liter, multiply by 0.01129.

a nonsense substitution (R19X) and two DNA sequence variants that disrupt a splice site (IVS2+1G→A and IVS3+1G→T). The missense A43T mutation has been shown to decrease APOC3 function in vitro.[25,26] When the participants of European ancestry were combined with the participants of African ancestry, approximately 1 in 150 persons carried any one of these four loss-of-function mutations (Table S8 in the Supplementary Appendix).[27]

To replicate the results derived from exome sequencing, we genotyped one or more of these four mutations in 34,432 participants of European ancestry and 7239 participants of African ancestry, independently of the discovery sequencing study. When compared with participants who did not carry any of the four mutations, those who were heterozygous for any of the four mutations had lower plasma triglyceride levels among participants of European ancestry (42% lower, $P<1\times10^{-20}$), among participants of African ancestry (32% lower, $P=1\times10^{-9}$), and among participants in the two groups combined (39% lower, $P<1\times10^{-20}$) (Table 2, and Table S9 in the Supplementary Appendix). *APOC3* loss-of-function carrier status was associated with reduced levels of plasma triglycerides whether or not the S19W variant of *APOA5* was present (Table S10 in the Supplementary Appendix). *APOC3* loss-of-function mutation carriers also had significantly higher HDL cholesterol levels, with a trend toward lower LDL cholesterol levels (Table 2, and Table S9 in the Supplementary Appendix).

APOC3 LOSS-OF-FUNCTION MUTATIONS, CORONARY HEART DISEASE, AND HEPATIC FAT

Among 110,970 study participants (34,002 patients with coronary heart disease and 76,968 controls) from 14 studies, 498 were heterozygous for at least one of the four *APOC3* loss-of-function mutations. The risk of coronary heart disease for carriers of an *APOC3* loss-of-function mutation was 40% lower than the risk for noncarriers (odds ratio, 0.60; 95% confidence interval [CI], 0.47 to 0.75; $P=4\times10^{-6}$) (Fig. 2, and Tables S3 and S11 in the Supplementary Appendix). There was no evidence of heterogeneity across the 15 studies (Breslow–Day test for heterogeneity, $P=0.16$).

We did not observe homozygosity or compound heterozygosity for *APOC3* loss-of-function variants in any participant. However, with a combined allele frequency of 1:300 and genotyping of 110,970 participants, the expected mean (±SD) number of persons with homozygosity or compound heterozygosity for *APOC3* loss-of-function variants was only 1±1 (Table S12 in the Supplementary Appendix).

A total of 3051 participants had both an *APOC3* loss-of-function allele and a phenotype of hepatic steatosis on CT. The presence of hepatic fat on CT did not differ significantly between carriers of

Table 2. Association of Four *APOC3* Mutations with Plasma Lipid Levels in Replication Studies.*

Variable Affected by Mutation	Carriers of Any of Four *APOC3* Mutations		
	European Ancestry (N=34,432)	African Ancestry (N=7239)	European and African Ancestries Combined (N=41,671)
Triglycerides (% difference)	−42	−32	−39
P value	$<1\times10^{-20}$	1×10^{-9}	$<1\times10^{-20}$
LDL cholesterol (mg/dl)	−9.3±3.4	10.7±5.4	−3.8±2.9
P value	0.006	0.05	0.19
HDL cholesterol (mg/dl)	11.5±1.3	9.1±2.0	10.8±1.1
P value	$<1\times10^{-20}$	8×10^{-6}	$<1\times10^{-20}$

* Plus–minus values indicate effect size ±SE. Effect size represents the difference in mean lipid levels between carriers and noncarriers after adjustment for covariates. The four *APOC3* mutations tested were R19X, splice-site mutation IVS2+1G→A, A43T, and splice-site mutation IVS3+1G→T. P values are for the comparison with noncarriers. P values were derived from a linear regression model, with adjustments for age, sex, ancestry, and principal components of ancestry. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for triglycerides to millimoles per liter, multiply by 0.01129.

an *APOC3* loss-of-function allele (27 participants) and noncarriers (3034 participants) (P=0.82) (Table S13 in the Supplementary Appendix).

APOC3 MUTATIONS AND PLASMA APOC3

In 3237 participants from the Framingham Heart Study, the mean plasma level of APOC3 was 16.6 mg per deciliter (Fig. S2 in the Supplementary Appendix). The APOC3 plasma level was associated with a number of cardiovascular risk factors, plasma lipid levels, lipoprotein subfractions, levels of other apolipoproteins, and levels of blood biomarkers (Table S14 in the Supplementary Appendix).

Genotypes had been obtained for 2707 of these Framingham Heart Study participants. A total of 13 persons carried one mutant allele (with heterozygosity for either R19X or splice-site IVS2+1G→A). As compared with noncarriers, these 13 persons had a plasma level of APOC3 that was 46% lower ($P=8\times10^{-10}$) (Fig. S3 in the Supplementary Appendix).

PLASMA APOC3 LEVELS AND INCIDENT CORONARY HEART DISEASE

During a mean follow-up of 14.4 years, 303 participants in the Framingham Heart Study had a first coronary heart disease event. In models adjusted for age and sex, each decrease of 1 mg per deciliter in plasma levels of APOC3 was associated with a 4% decrease in the risk of incident coronary heart disease (hazard ratio, 0.96; 95% CI, 0.94 to 0.98; P<0.001) (Table S15 in the Supplementary Appendix). Participants with values in the lowest third of the distribution of plasma APOC3 levels had a reduced risk of incident coronary heart disease, as compared with participants with values in the highest third (odds ratio, 0.65; 95% CI, 0.48 to 0.87; P=0.003) (Fig. 3, and Table S16 in the Supplementary Appendix). However, this association was not significant after adjustment for additional cardiovascular risk factors (Tables S15 and S16 in the Supplementary Appendix). In the Verona Heart Study cohort,[18] which consisted of 794 patients with angiographic evidence of coronary artery disease, the lowest third of the distribution of APOC3 levels was associated with a reduced risk of death from cardiovascular disease, as compared with the highest third, in models that were minimally adjusted and fully adjusted for risk factors (Tables S17 through S20 in the Supplementary Appendix).

DISCUSSION

By sequencing the protein-coding regions of the genome in 3734 persons, we identified several rare coding-sequence variants of *APOC3* that were associated with a large effect on plasma triglyceride levels. Approximately 1 in 150 persons carried any one of four protein-altering or splice-site variants. As compared with noncarriers, carriers of any of these four *APOC3* mutations had plasma triglyceride levels that were 39% lower, HDL cholesterol levels that were 22% higher, LDL cholesterol levels that were 16% lower, and circulating APOC3 levels that were 46% lower, and they had a risk of coronary heart disease that was reduced by 40%.

These results show that loss of APOC3 function confers protection against clinical coronary heart disease. In a previous study, Pollin and colleagues[28] reported that approximately 5% of Amish persons in Lancaster County, Pennsylvania, carried the *APOC3* R19X null allele and that these carriers had a favorable lipid profile and a reduced amount of coronary artery calcium, a surrogate marker for atherosclerosis. The A43T missense mutation in *APOC3* (referred to as A23T in the

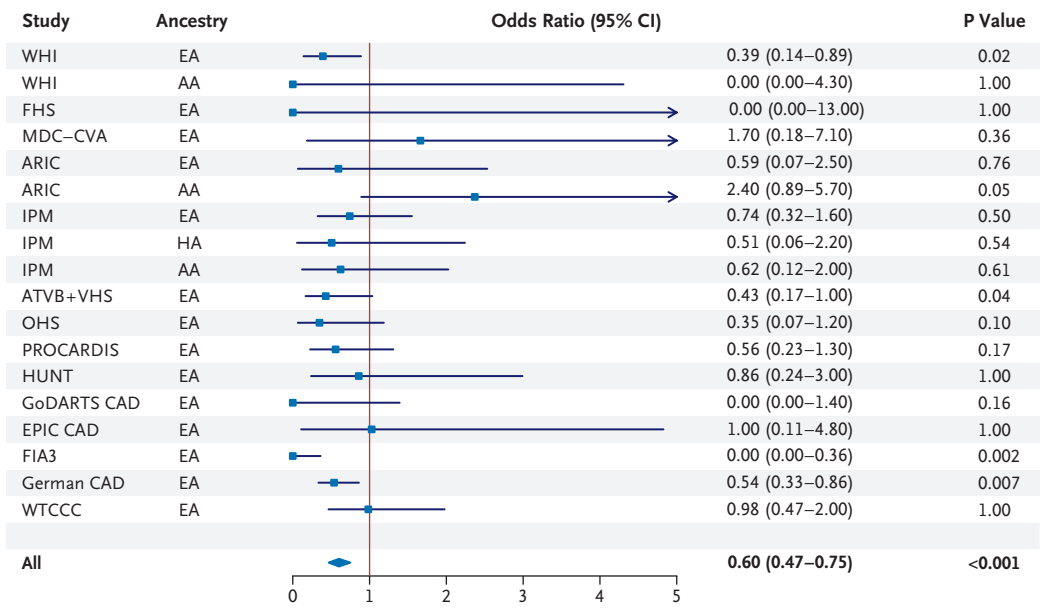


Figure 2. Association of *APOC3* Loss-of-Function Mutations with Risk of Coronary Heart Disease among 110,970 Participants in 15 Studies.

In each study, we tested the association of loss-of-function carrier status (heterozygous for any of four mutations: *APOC3* R19X, IVS2+1G→A, IVS3+1G→T, or A43T) with the risk of coronary heart disease. We calculated P values for the association tests and confidence intervals for the odds ratios with the use of exact methods. We performed a meta-analysis with the use of the Cochran–Mantel–Haenszel statistics for stratified 2-by-2 tables. The Cochran–Mantel–Haenszel method combines score statistics rather than Wald statistics and is particularly useful when some observed odds ratios are zero. For each study, squares indicate the estimated odds ratios and the corresponding lines indicate the 95% confidence intervals. The diamond indicates the combined estimate of the odds ratio and the corresponding 95% confidence interval. HA denotes Hispanic ancestry. The full study names are as follows: ARIC Atherosclerosis Risk in Communities. ATVB Italian Atherosclerosis, Thrombosis, and Vascular Biology Study, EPIC European Prospective Study into Cancer and Nutrition, FHS Framingham Heart Study, FIA3 First Myocardial Infarction, in AC County 3, GoDARTS Genetics of Diabetes Audit and Research Tayside Study, HUNT Nord–Trøndelag Health Study, IPM Mt. Sinai Institute for Personalized Medicine Biobank, MDC–CVA Malmö Diet and Cancer Study Cardiovascular Cohort, OHS Ottawa Heart Study, PROCARDIS, Precocious Coronary Artery Disease Study, VHS Verona Heart Study, WHI Women's Health Initiative, and WTCCC Wellcome Trust Case Control Consortium. All 15 studies are described in Table S3 in the Supplementary Appendix.

earlier nomenclature) was initially identified in two Mayan Indians with low plasma levels of triglycerides and APOC3.[25] Functional characterization in vitro showed deficient lipid binding[25] and attenuated secretion of very-low-density lipoprotein from hepatocytes.[26] Recently, in three families, 12 carriers of the A43T mutation or the two *APOC3* splice-site mutations were identified, and as compared with family members who were not carriers, the carriers had lower levels of plasma triglycerides and higher levels of HDL cholesterol.[29] In our study, which included 110,970 participants, carriers of *APOC3* loss-of-function mutations had a reduced risk of clinical coronary heart disease.

Several mechanisms may link APOC3 to atherogenesis.[30] Resident on the surface of triglyceride-rich lipoproteins, APOC3 inhibits the lipolytic activity of lipoprotein lipase, thereby increasing the plasma level of atherogenic triglyceride-rich lipoproteins, including very-low-density lipoproteins and chylomicrons. APOC3 also delays the clearance of atherogenic, cholesterol-rich remnants of the postprandial metabolism of triglyceride-rich lipoproteins.[31,32]

The question of whether triglyceride-rich lipoproteins in plasma have a causal effect on coronary heart disease remains unanswered. Several treatments that lower plasma levels of triglycerides, including fibrates and fish oils,

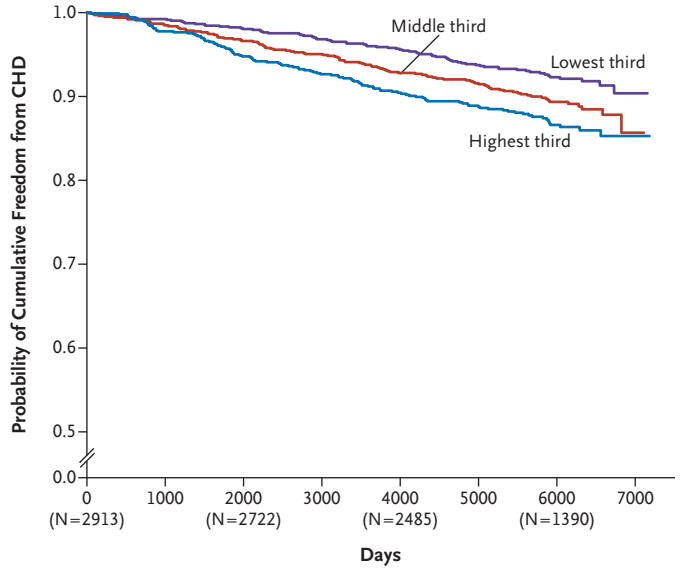


Figure 3. Cumulative Probability of Freedom from Coronary Heart Disease (CHD) According to Plasma Level of APOC3 at Baseline in the Framingham Heart Study.

Plasma APOC3 levels were 14.2 mg per deciliter or less in the lowest third of the population, 14.3 to 17.9 mg per deciliter in the middle third, and 18.0 mg per deciliter or more in the highest third. Median follow-up was 14.4 years. The numbers in parentheses are the numbers of study participants who were undergoing follow-up at the specified time points.

have failed to consistently reduce the risk of coronary heart disease in randomized, controlled trials.[33,34] Our findings with respect to the human *APOC3* loss-of-function alleles suggest that a lifelong decrease in APOC3 function is one means of reducing plasma levels of triglyceride-rich lipoproteins that may be atheroprotective. Recently, an antisense oligonucleotide therapeutic agent that decreased the production of APOC3 was shown to lower plasma levels of triglycerides and APOC3 in mice, nonhuman primates, and healthy human volunteers.[35] Our data provide confidence that such a therapeutic strategy might be expected to reduce the risk of clinical coronary heart disease.

In more general terms, the results of our study highlight the potential usefulness of naturally occurring loss-of-function mutations in guiding the selection of therapeutic targets.[24] For example, loss-of-function mutations in the gene encoding proprotein convertase subtilisin/kexin type 9 (PCSK9) were shown to reduce LDL cholesterol levels and lower the risk of coronary heart disease, a finding that prompted the development of monoclonal antibodies directed against PCSK9.[36-38]

Several limitations of our study deserve mention. First, we are unable to pinpoint the specific mechanism by which loss of APOC3 function might be atheroprotective. Carriers of *APOC3* loss-of-function mutations have a range of associated biomarkers, including lower levels of triglycerides, higher levels of HDL cholesterol, lower levels of LDL cholesterol, and lower levels of APOC3. Additional work will be required to understand the primary mechanism linking APOC3 to coronary heart disease. Second, earlier reports are conflicting as to whether common DNA sequence variants at *APOC3* affect hepatic steatosis.[39,40] We found no significant difference in CT findings with respect to hepatic fat between study participants who carried *APOC3* loss-of-function mutations and those who did not. However, larger samples may be required to confidently answer this question.

In conclusion, we identified rare DNA sequence variants in *APOC3* that were associated with lifelong reductions in plasma levels of triglycerides and APOC3. These variants conferred protection against coronary heart disease.

The views expressed in this article are solely those of the authors and do not necessarily represent the official views of the National Heart, Lung, and Blood Institute (NHLBI) or the National Institutes of Health (NIH).

Supported by the Grand Opportunity (GO) Exome Sequencing Project (ESP), through which grants were provided by the NHLBI (RC2 HL-103010 [HeartGO], RC2 HL-102923 [LungGO], RC2 HL-102924 [Women's Health Initiative ESP], RC2 HL-102925 [BroadGO], and RC2 HL-102926 [SeattleGO]), the Russell and Diana Hawkins Family Foundation Discovery Fellowship, an award from the NHLBI (T32HL007208, to Dr. Peloso), an award from the British Heart Foundation (BHF) (to Dr. Samani, who holds a chair supported by the BHF), a Banting Fellowship from the Canadian Institutes of Health Research (to Dr. Do), a career-development award from the NIH and NHLBI (K08-HL114642, to Dr. Stitziel), a Research Scholar Award from the Massachusetts General Hospital (MGH) (to Dr. Kathiresan), the Howard M. Goodman Fellowship from MGH, the Donovan Family Foundation, a grant from the NIH (R01HL107816, to Dr. Kathiresan), an investigator-initiated research grant from Merck (to Dr. Kathiresan), a grant from Fondation Leducq (to Dr. Kathiresan), a grant from the Swedish Heart–Lung Foundation (2020389, to the Swedish First Myocardial Infarction in AC County 3 [FIA3] Study), and a grant from Regione Emilia-Romagna (RFPS-2007-3-644382, to the Italian Atherosclerosis, Thrombosis, and Vascular Biology Study).

We thank the site investigators, field staff, and study participants for their contributions to this study. A complete list of acknowledgements is provided in the Supplementary Appendix.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

APPENDIX

The authors are as follows: Jacy Crosby, M.S., Gina M. Peloso, Ph.D., Paul L. Auer, Ph.D., David R. Crosslin, Ph.D., Nathan O. Stitziel, M.D., Ph.D., Leslie A. Lange, Ph.D., Yingchang Lu, M.D., Ph.D., Zheng-zheng Tang, M.S., He Zhang, Ph.D., George Hindy, M.D., Nicholas Masca, Ph.D., Kathleen Stirrups, Ph.D., Stavroula Kanoni, Ph.D., Ron Do, Ph.D., Goo Jun, Ph.D., Youna Hu, Ph.D., Hyun Min Kang, Ph.D., Chenyi Xue, M.S., Anuj Goel, M.Sc., Martin Farrall, F.R.C.Path., Stefano Duga, Ph.D., Pier Angelica Merlini, M.D., Rosanna Asselta, Ph.D., Domenico Girelli, M.D., Oliviero Olivieri, M.D., Nicola Martinelli, M.D., Ph.D., Wu Yin, Ph.D., Dermot Reilly, Ph.D., Elizabeth Speliotes, M.D., Ph.D., Caroline S. Fox, M.D., Kristian Hveem, M.D., Ph.D., Oddgeir L. Holmen, M.D., Majid Nikpay, Ph.D., Deborah N. Farlow, Ph.D., Themistocles L. Assimes, M.D., Ph.D., Nora Franceschini, M.D., M.P.H., Jennifer Robinson, M.D., Kari E. North, Ph.D., Lisa W. Martin, M.D., Mark DePristo, Ph.D., Namrata Gupta, Ph.D., Stefan A. Escher, Ph.D., Jan-Håkan Jansson, M.D., Natalie Van Zuydam, Ph.D., Colin N.A. Palmer, Ph.D., Nicholas Wareham, M.B., B.S., Ph.D., Werner Koch, M.D., Thomas Meitinger, M.D., Annette Peters, M.D., Wolfgang Lieb, M.D., Raimund Erbel, M.D., Inke R. Konig, Ph.D., Jochen Kruppa, M.Sc., Franziska Degenhardt, M.D., Omri Gottesman, M.D., Erwin P. Bottinger, M.D., Christopher J. O'Donnell, M.D., Bruce M. Psaty, M.D., Ph.D., Christie M. Ballantyne, M.D., Goncalo Abecasis, Ph.D., Jose M. Ordovas, Ph.D., Olle Melander, M.D., Ph.D., Hugh Watkins, F.R.C.P., Marju Orho-Melander, Ph.D., Diego Ardissino, M.D., Ruth J.F. Loos, Ph.D., Ruth McPherson, M.D., Ph.D., Cristen J. Willer, Ph.D., Jeanette Erdmann, Ph.D., Alistair S. Hall, F.R.C.P., Nilesh J. Samani, F.R.C.P., Panos Deloukas, Ph.D., Heribert Schunkert, M.D., James G. Wilson, M.D., Charles Kooperberg, Ph.D., Stephen S. Rich, Ph.D., Russell P. Tracy, Ph.D., Dan-Yu Lin, Ph.D., David Altshuler, M.D., Ph.D., Stacey Gabriel, Ph.D., Deborah A. Nickerson, Ph.D., Gail P. Jarvik, M.D., Ph.D., L. Adrienne Cupples, Ph.D., Alex P. Reiner, M.D., Eric Boerwinkle, Ph.D., and Sekar Kathiresan, M.D.

The authors' affiliations are as follows: the Department of Biostatistics, Bioinformatics, and Systems Biology, University of Texas Graduate School of Biomedical Sciences at Houston (J.C.), the Human Genetics Center, University of Texas Health Science Center at Houston (J.C., E.B.), Baylor College of Medicine (C.M.B., E.B.), and Methodist DeBakey Heart and Vascular Center (C.M.B.) — all in Houston; the Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital (G.M.P., R.D., D. Altshuler, S. Kathiresan), the Department of Medicine, Harvard Medical School (G.M.P., R.D., D. Altshuler, S. Kathiresan), Merck Sharp & Dohme (W.Y., D.R.), Nutrition and Genomics Laboratory, Jean Mayer–USDA Human Nutrition Research Center on Aging at Tufts University (J.M.O.), and Department of Biostatistics, Boston University School of Public Health (L.A.C.) — all in Boston; the Program in Medical and Population Genetics, Broad Institute, Cambridge, MA (G.M.P., R.D., D.N.F., M.D., N.G., D. Altshuler, S.G., S. Kathiresan); the School of Public Health, University of Wisconsin–Milwaukee, Milwaukee (P.L.A.); the Departments of Genome Sciences (D.R.C., D.A.N., G.P.J.), Medicine (Medical Genetics) (D.R.C., G.P.J.), and Epidemiology (A.P.R.) and the Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services (B.M.P.), University of Washington, the Group Health Research Institute, Group Health Cooperative (B.M.P.), and the Public Health Sciences Division, Fred Hutchinson Cancer Research Center (C.K., A.P.R.) — all in Seattle; the Cardiovascular Division, Department of Medicine, and the Division of Statistical Genomics, Washington University School of Medicine, St. Louis (N.O.S.); the Departments of Genetics (L.A.L.), Biostatistics (Z.T., D.-Y.L.), and Epidemiology (N.F., K.E.N.) and the Carolina Center for Genome Sciences (K.E.N.), University of North Carolina, Chapel Hill; the Charles Bronfman Institute for Personalized Medicine (Y.L., O.G., E.P.B., R.J.F.L.) and the Mindich Child Health and Development Institute (R.J.F.L.), Icahn School of Medicine at Mount Sinai, New York; the Departments of Internal Medicine (H.Z., E.S., C.J.W.), Computational Medicine and Bioinformatics (H.Z., C.X., E.S., C.J.W.), Human Genetics (H.Z., C.J.W.), and Biostatistics (G.J., Y.H., H.M.K., G.A.), the Division of Gastroenterology (E.S.), and the Center for Statistical Genetics (G.A.), University of Michigan, Ann Arbor; the Department of Clinical Sciences, Clinical Research Center (G.H., M.O.-M.), Department of Clinical Science, Genetic and Molecular Epidemiology Unit (S.A.E.), and Department of Clinical Sciences, Diabetes and Endocrinology (O.M.), Lund University and University Hospital Malmö, Malmö, the Department of Medicine, Skellefteå Hospital, Skellefteå (J.-H.J.), and the Department of Public Health and Clinical Medicine, Umeå University, Umeå, (J.-H.J.) — all in Sweden; the Department of Cardiovascular Sciences, University of Leicester, and National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Leicester (N. Masca, N.J.S.), the William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London (K.S., S. Kanoni, P.D.), the Division of Cardiovascular Medicine, Radcliffe Department of Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford (A.G., M.F., H.W.), the Medical Research Institute, Ninewells Hospital and Medical School, Dundee (N.V.Z., C.N.A.P.), the Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital (N.W., L.A.C.), and Wellcome Trust Sanger Institute (P.D.), Cambridge, and the Division of Epidemiology, School of Medicine, University of Leeds, Leeds (A.S.H.) — all in the United Kingdom; the Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano (S.D., R.A.) and the Division of Cardiology, Ospedale Niguarda (P.A.M.), Milan, the University of Verona School of Medicine, Department of Medicine, Verona (D.G., O.O., N. Martinelli), and the Division of Cardiology, Azienda Ospedaliero-Universitaria di Parma, Parma (D. Ardissino) — all in Italy; the National Heart, Lung and Blood Institute Framingham Heart Study (C.S.F., C.J.O.) and the Center for Population Studies (C.S.F.), Framingham, MA; the Nord-Trøndelag Health Study (HUNT) Research Center, Department of Public Health and General Practice, Norwegian University of Science and Technology (K.H., O.L.H.), and Levanger Hospital (K.H.), Levanger, and St. Olav Hospital, Trondheim University Hospital, Trondheim (O.L.H.) — both in Norway; the Division of Cardiology, University of Ottawa Heart Institute, Ottawa (M. Nikpay, R.M.); the Department of Medicine, Stanford University School of Medicine, Stanford, CA (T.L.A.); the Departments of Epidemiology and Medicine, College of Public Health, University of Iowa, Ames (J.R.); the Division of Cardiology, George Washington University School of Medicine and Health Sciences, Washington, DC (L.W.M.); Deutsches Herzzentrum München (W.K., H.S.), Medizinische Klinik, Klinikum rechts der Isar (W.K.), and Institute of Human Genetics (T.M.), Technische Universität München, and the German Center for Cardiovascular Research (W.K., A.P.), Munich, Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology II, Neuherberg (A.P.), Institute of Epidemiology and Biobank popgen, Christian-Albrechts University Kiel, Kiel (W.L.), Department of Cardiology, West German Heart Center, Essen (R.E.), Institut für Medizinische Biometrie und Statistik (I.R.K., J.K.) and Institut für Integrative und Experimentelle Genomik (J.E.), Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein (I.R.K., J.K.), and DZHK (German Research Center for Cardiovascular Research) (J.E.), Lübeck, and the Institute of Human Genetics, University of Bonn, Bonn (F.D.) — all in Germany; the Department of Cardiovascular Epidemiology and Population Genetics, National Center for Cardiovascular Investigation, and Instituto Madrileño de Estudios Avanzados–Alimentacion, Madrid (J.M.O.); King Abdulaziz University, Jeddah, Saudi Arabia (P.D.); the Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson (J.G.W.); the Center for Public Health Genomics, University of Virginia, Charlottesville (S.S.R.); and the Departments of Pathology and Biochemistry, University of Vermont College of Medicine, Burlington (R.P.T.).

REFERENCES

1. Miller M, Stone NJ, Ballantyne C, et al. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2011;123:2292-333.
2. Chapman MJ, Ginsberg HN, Amarenco P, et al. Triglyceride-rich lipoproteins and high-density lipoprotein cholesterol in patients at high risk of cardiovascular disease: evidence and guidance for management. Eur Heart J 2011;32:1345-61.
3. Namboodiri KK, Kaplan EB, Heuch I, et al. The Collaborative Lipid Research Clinics Family Study: biological and cultural determinants of familial resemblance for plasma lipids and lipoproteins. Genet Epidemiol 1985;2:227-54.
4. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707-13.
5. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45:1274-83.
6. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat Genet 2013;45:1345-52.
7. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461:272-6.
8. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 2009;106:19096-101.
9. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012;337:64-9.
10. Fu W, O'Connor TD, Jun G, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013;493:216-20.
11. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 1989;129:687-702.
12. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 1988;41:1105-16.
13. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991;1:263-76.
14. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951;41:279-81.
15. 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:Suppl 6:S4-S17.
16. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871-81.
17. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 1998;19:61-109.
18. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334-41.
19. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008;83:311-21.
20. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. Am J Epidemiol 1979;110:281-90.
21. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
22. Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. Am J Hum Genet 2011;89:354-67.
23. Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. Hum Mol Genet 2002;11:3031-8.
24. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 2012;335:823-8. [Erratum, Science 2012;336:296.]
25. Liu H, Labeur C, Xu CF, et al. Characterization of the lipid-binding properties and lipoprotein lipase inhibition of a novel apolipoprotein C-III variant Ala23Thr. J Lipid Res 2000;41:1760-71.
26. Sundaram M, Zhong S, Bou Khalil M, et al. Functional analysis of the missense APOC3 mutation Ala23Thr associated with human hypotriglyceridemia. J Lipid Res 2010;51:1524-34.
27. NHLBI Exome Sequencing Project. Exome Variant Server (http://evs.gs.washington.edu/EVS).
28. Pollin TI, Damcott CM, Shen H, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. Science 2008;322:1702-5.
29. Bochem AE, van Capelleveen JC, Dallinga-Thie GM, et al. Two novel mutations in apolipoprotein C3 underlie atheroprotective lipid profiles in families. Clin Genet 2014;85:433-40.
30. Huff MW, Hegele RA. Apolipoprotein C-III: going back to the future for a lipid drug target. Circ Res 2013;112:1405-8.
31. Windler E, Havel RJ. Inhibitory effects of C apolipoproteins from rats and humans on the uptake of triglyceride-rich lipoproteins and their remnants by the perfused rat liver. J Lipid Res 1985;26:556-65.
32. Zheng C, Khoo C, Furtado J, Sacks FM. Apolipoprotein C-III and the metabolic basis for hypertriglyceridemia and the dense low-density lipoprotein phenotype. Circulation 2010;121:1722-34.
33. The ACCORD Study Group. Effects of combination lipid therapy in type 2 diabetes mellitus. N Engl J Med 2010;362:1563-74. [Erratum, N Engl J Med 2010;362:1748.]
34. The ORIGIN Trial Investigators. n–3 Fatty acids and cardiovascular outcomes in patients with dysglycemia. N Engl J Med 2012;367:309-18.
35. Graham MJ, Lee RG, Bell TA III, et al. Antisense oligonucleotide inhibition of apolipoprotein C-III reduces plasma triglycerides in rodents, nonhuman primates, and humans. Circ Res 2013;112:1479-90.
36. Abifadel M, Varret M, Rabès JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 2003;34:154-6.
37. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264-72.
38. Stein EA, Mellis S, Yancopoulos GD, et al. Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. N Engl J Med 2012;366:1108-18.
39. Petersen KF, Dufour S, Hariri A, et al. Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. N Engl J Med 2010;362:1082-9.
40. Kozlitina J, Boerwinkle E, Cohen JC, Hobbs HH. Dissociation between APOC3 variants, hepatic triglyceride content and insulin resistance. Hepatology 2011;53:467-74.

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: The TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. N Engl J Med. DOI: 10.1056/NEJMoa1307095

ONLINE SUPPLEMENTARY APPENDIX

Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease

Jacy Crosby, M.S.*[1], Gina M. Peloso, Ph.D.*[2,3,4,5], Paul L. Auer, Ph.D.*[6], David R. Crosslin, Ph.D.[7,8], Nathan O. Stitziel, M.D., Ph.D.[9,10], Leslie A. Lange, Ph.D.[11], Yingchang Lu, M.D. Ph.D.[12], Zheng-zheng Tang, M.S.[13], He Zhang, Ph.D.[14,15,16], George Hindy, M.D.[17], Nicholas Masca, Ph.D.[18], Kathleen Stirrups, Ph.D.[19], Stavroula Kanoni, Ph.D.[19], Ron Do, Ph.D.[2-5], Goo Jun, Ph.D.[20], Youna Hu, Ph.D.[20], Hyun Min Kang, Ph.D.[20], Chenyi Xue, Ph.D.[15], Anuj Goel, M.Sc.[21], Martin Farrall, F.R.C.Path.[21], Stefano Duga, Ph.D.[22], Pier Angelica Merlini, M.D.[23], Rosanna Asselta[22], Domenico Girelli, M.D.[24], Oliviero Olivieri, M.D.[24], Nicola Martinelli, M.D., Ph.D.[24], Wu Yin, Ph.D.[25], Dermot Reilly, Ph.D.[25], Elizabeth Speliotes, M.D., Ph.D.[26], Caroline S. Fox, M.D.[27], Kristian Hveem, M.D., Ph.D.[28,29], Oddgeir L. Holmen, M.D.[28,30], Majid Nikpay, Ph.D.[31], NHLBI Exome Sequencing Project, Deborah N. Farlow, Ph.D.[5], Themistocles L. Assimes, M.D., Ph.D.[32], Nora Franceschini, M.D., M.P.H.[33], Jennifer Robinson, M.D.[34], Kari E. North, Ph.D.[33,35], Lisa W. Martin, M.D.[36], Mark DePristo, Ph.D.[5], Namrata Gupta, Ph.D.[5], Stefan A. Escher, Ph.D.[37], Jan-Håkan Jansson, M.D.[38,39], Natalie Van Zuydam, Ph.D.[40], Colin N.A. Palmer, Ph.D.[40], Nicholas Wareham, M.B.B.S., Ph.D.[41], Werner Koch, M.D.[42], Thomas Meitinger, M.D.[43], Annette Peters, M.D.[44], Wolfgang Lieb, M.D.[45], Raimund Erbel, M.D.[46], Inke R. Konig, Ph.D.[47], Jochen Kruppa, M.Sc.[47], Franziska Degenhardt[48], Omri Gottesman, M.D.[12], Erwin P. Bottinger, M.D.[12], Christopher J. O'Donnell, M.D.[49], Bruce M. Psaty, M.D., Ph.D.[50,51], Christie M. Ballantyne, M.D.[52], Goncalo Abecasis, Ph.D.[53], Jose M. Ordovas, Ph.D.[54,55,56], Olle Melander, M.D., Ph.D.[57], Hugh Watkins, F.R.C.P[21], Marju Orho-Melander, Ph.D.[17], Diego Ardissino, M.D.[58], Ruth J.F. Loos, Ph.D.[12,59], Ruth McPherson, M.D., Ph.D.[31], Cristen J. Willer, Ph.D.[14,15,16], Jeanette Erdmann, Ph.D.[60], Alistair S. Hall, F.R.C.P.[61], Nilesh J. Samani, F.R.C.P.[18], Panos Deloukas, Ph.D.[62], Heribert Schunkert, M.D.[63], James G. Wilson, M.D.[64], Charles Kooperberg, Ph.D.[65], Stephen S. Rich, Ph.D.[66], Russell P. Tracy, Ph.D.[67], Dan-Yu Lin, Ph.D.[13], David Altshuler, M.D., Ph.D.[2-4], Stacey Gabriel, Ph.D.[5], Deborah A. Nickerson, Ph.D.[7], Gail P. Jarvik, M.D., Ph.D.[7,8], L. Adrienne Cupples, Ph.D.[41,68], Alex P. Reiner, M.D.[69]*, Eric Boerwinkle, Ph.D.[70]*, Sekar Kathiresan, M.D.[2,3,4,5]*

*Contributed equally

Address for correspondence:
Sekar Kathiresan, MD
Cardiovascular Research Center and Center for Human Genetic Research
Massachusetts General Hospital
185 Cambridge Street, CPZN 5.251
Boston, MA 02114

skathiresan@partners.org

1. Department of Biostatistics, Bioinformatics, and Systems Biology, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, TX, USA and Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX, USA
2. Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA
3. Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA
4. Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
5. Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, MA, USA
6. School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
7. Department of Genome Sciences, University of Washington, Seattle, Washington, USA
8. Department of Medicine (Medical Genetics), University of Washington, Seattle, Washington, USA
9. Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA
10. Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA
11. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
12. The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, USA
13. Department of Biostatistics, University of North Carolina, 3101 McGavran-Greenberg Hall, Chapel Hill, NC, USA
14. Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA
15. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
16. Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA
17. Department of Clinical Sciences in Malmö, Lund University, Clinical Research Center, Malmö, Sweden
18. Department of Cardiovascular Sciences, University of Leicester and NIHR Leicester Cardiovascular Biomedical Research Unit, Leicester. LE3 9QP, UK
19. UK William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK
20. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
21. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
22. Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milan, Italy
23. Division of Cardiology, Ospedale Niguarda, Milano, Italy
24. University of Verona School of Medicine, Department of Medicine, Verona, Italy.
25. Merck Sharp and Dohme Corp., Boston, MA, USA
26. Department of Internal Medicine, Division of Gastroenterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
27. National Heart, Lung and Blood Institute's Framingham Heart Study and the Center for Population Studies, Framingham, MA, USA.
28. HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway
29. Levanger Hospital, Levanger, Norway
30. St. Olav Hospital, Trondheim University Hospital, Trondheim, Norway
31. Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada
32. Department of Medicine, Stanford University School of Medicine, Stanford, California, USA
33. Department of Epidemiology, University of North Carolina, 137 E. Franklin St., Suite 306, Chapel Hill, North Carolina, USA
34. Departments of Epidemiology & Medicine, College of Public Health, University of Iowa, Ames, IA, USA
35. Carolina Center for Genome Sciences, University of North Carolina, 5009 Genetic Medicine Building, Chapel Hill, North Carolina 27599, USA
36. Division of Cardiology, George Washington University School of Medicine and Health Sciences, Washington, DC, USA
37. Department of Clinical Science, Genetic & Molecular Epidemiology Unit, Lund University, Malmö, Sweden
38. Department of Medicine, Skellefteå Hospital, Skellefteå, Sweden
39. Department of Public Health & Clinical Medicine, Umeå University, Umeå, Sweden
40. Medical Research Institute, Ninewells Hospital and Medical School, Dundee, UK
41. MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Box 285, Cambridge, UK

42. Deutsches Herzzentrum München and Medizinische Klinik, Klinikum rechts der Isar, Technische Universität München, 80636 München, Germany and German Center for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, 80636 Munich, Germany
43. Institute of Human Genetics, Technische Universität München, 81675 München, Germany
44. Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Institute of Epidemiology II, 85764 Neuherberg, Germany and DZHK (German Centre for Cardiovascular Research), Partner Site Munich, Germany
45. Institute of Epidemiology and Biobank popgen, Christian-Albrechts-University Kiel, 24105 Kiel, Germany
46. Department of Cardiology, West German Heart Center, Essen, Germany
47. Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, 23538 Lübeck, Germany
48. Institute of Human Genetics, University of Bonn, Bonn, Germany
49. The National Heart, Lung, Blood Institute's Framingham Heart Study, Framingham,MA, USA
50. Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA
51. Group Health Research Institute, Group Health Cooperative, Seattle, WA
52. Baylor College of Medicine, and Methodist DeBakey Heart and Vascular Center, Houston, TX, USA
53. Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA
54. Department of Cardiovascular Epidemiology and Population Genetics, National Center for Cardiovascular Investigation, Madrid, Spain
55. IMDEA-Alimentacion, Madrid, Spain
56. Nutrition and Genomics Laboratory, Jean Mayer-USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA
57. Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden
58. Division of Cardiology, Azienda Ospedaliero-Universitaria di Parma, Parma, Italy
59. The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
60. Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, 23562 Lübeck, Germany and DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, 23562 Lübeck, Germany.
61. Division of Epidemiology, School of Medicine, University of Leeds, Leeds, UK
62. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA UK; William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK; and King Abdulaziz University, Jeddah 21589, Saudi Arabia
63. Technische Universität und Deutsches Herzzentrum München, Lazarettstrasse 36, 80636, Munich, and Deutsches Zentrum für Herz- und Kreislaufforschung (DZHK), Munich Heart Alliance, Germany
64. Department of Physiology and Biophysics, the University of Mississippi Medical Center, Jackson, MS, USA
65. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
66. Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
67. Departments of Pathology and Biochemistry, University of Vermont College of Medicine, Burlington, VT, USA
68. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
69. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA and Department of Epidemiology, University of Washington, Seattle, Washington, USA
70. Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX, USA and Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

Table of Contents

Participating Investigators: NHLBI GO Exome Sequencing Project 5

Acknowledgments 11

Supplementary Methods

- I. Exome Sequencing at the University of Washington 13
- II. Exome Sequencing at the Broad Institute 15
- III. Joint Calling of Variants for Entire ESP Project at the University of Michigan 18
- IV. Exome Array Genotyping 29
- V. Study Participants 29
- VI. APOC3 Genotypes: Replication For Plasma Lipids 30
- VII. APOC3 Genotypes: Association With CHD 31
- VIII. Plasma apoC-III Protein Concentration And Risk For Incident CHD 32
- IX. APOC3 Genotypes and Association With Hepatic Steatosis 33

Supplementary Figures 34

Supplementary Tables 38

List of Investigators

NHLBI GO Exome Sequencing Project

BroadGO

Stacey B. Gabriel (Broad Institute)[4, 5, 11, 16, 17], David M. Altshuler (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[1, 5, 7, 17], Gonçalo R. Abecasis (University of Michigan)[3, 5, 9, 13, 15, 17], Hooman Allayee (University of Southern California)[5], Sharon Cresci (Washington University School of Medicine)[5], Mark J. Daly (Broad Institute, Massachusetts General Hospital), Paul I. W. de Bakker (Broad Institute, Harvard Medical School, University Medical Center Utrecht)[3, 15], Mark A. DePristo (Broad Institute)[4, 13, 15, 16], Ron Do (Broad Institute)[5, 9, 13, 15], Peter Donnelly (University of Oxford)[5], Deborah N. Farlow (Broad Institute)[3, 4, 5, 12, 14, 16, 17], Tim Fennell (Broad Institute), Kiran Garimella (University of Oxford)[4, 16], Stanley L. Hazen (Cleveland Clinic)[5], Youna Hu (University of Michigan)[3, 9, 15], Daniel M. Jordan (Harvard Medical School, Harvard University)[13], Goo Jun (University of Michigan)[13], Sekar Kathiresan (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[5, 8, 9, 12, 14, 15, 17, 20], Hyun Min Kang (University of Michigan)[9, 13, 16], Adam Kiezun (Broad Institute)[5, 13, 15], Guillaume Lettre (Broad Institute, Montreal Heart Institute, Université de Montréal)[1, 2, 13, 15], Bingshan Li (University of Michigan)[3], Mingyao Li (University of Pennsylvania)[5], Christopher H. Newton-Cheh (Broad Institute, Massachusetts General Hospital, Harvard Medical School)[3, 8, 15], Sandosh Padmanabhan (University of Glasgow School of Medicine)[3, 12, 15], Gina Peloso (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[5], Sara Pulit (Broad Institute)[3, 15], Daniel J. Rader (University of Pennsylvania)[5], David Reich (Broad Institute, Harvard Medical School)[15], Muredach P. Reilly (University of Pennsylvania)[5], Manuel A. Rivas (Broad Institute, Massachusetts General Hospital)[5], Steve Schwartz (Fred Hutchinson Cancer Research Center)[5, 12], Laura Scott (University of Michigan)[1], David S. Siscovick (University of Washington)[5, 1, 25], John A. Spertus (University of Missouri Kansas City)[5], Nathan O. Stitziel (Brigham and Women's Hospital)[5, 15], Nina Stoletzki (Brigham and Women's Hospital, Broad Institute, Harvard Medical School)[13], Shamil R. Sunyaev (Brigham and Women's Hospital, Broad Institute, Harvard Medical School)[1, 3, 5, 13, 15], Benjamin F. Voight (Broad Institute, Massachusetts General Hospital), Cristen J. Willer (University of Michigan)[1, 9, 13, 15]

HeartGO

Stephen S. Rich (University of Virginia)[2, 4, 7, 8, 9, 11,14, 15, 17, 18, 31], Ermeg Akylbekova (Jackson State University, University of Mississippi Medical Center)[29], Larry D. Atwood* (Boston University)[1, 11, 28], Christie M. Ballantyne (Baylor College of Medicine, Methodist DeBakey Heart Center)[9, 22], Maja Barbalic (University of Texas Health Science Center Houston)[9, 14, 15, 17, 22], R. Graham Barr (Columbia University Medical Center)[10, 31], Emelia J. Benjamin (Boston University)[14, 20, 28], Joshua Bis (University of Washington)[15, 23], Eric Boerwinkle (University of Texas Health Science Center Houston)[3, 5, 9, 13, 15, 17, 22], Donald W. Bowden (Wake Forest University)[1, 31], Jennifer Brody (University of Washington)[3, 5, 15, 23], Matthew Budoff (Harbor-UCLA Medical Center)[31], Greg Burke (Wake Forest University)[5, 31], Sarah Buxbaum (Jackson State University)[3, 13, 15, 29], Jeff Carr (Wake Forest University)[25, 29, 31], Donna T. Chen (University of Virginia)[6, 11], Ida Y. Chen (Cedars-Sinai Medical Center)[1, 31], Wei-Min Chen (University of Virginia)[13, 15, 18], Pat Concannon (University of Virginia)[11], Jacy Crosby (University of Texas Health Science Center Houston)[22], L. Adrienne Cupples (Boston University)[1, 3, 5, 9, 13, 15, 18, 28], Ralph D'Agostino (Boston University)[28], Anita L. DeStefano (Boston University)[13, 18, 28], Albert Dreisbach (University of Mississippi Medical Center)[3, 29], Josée Dupuis (Boston University)[1, 28], J. Peter Durda (University of Vermont)[15, 23], Jaclyn Ellis (University of North Carolina Chapel Hill)[1], Aaron R. Folsom (University of Minnesota)[5, 22], Myriam Fornage (University of Texas Health Science Center Houston)[3, 18, 25], Caroline S. Fox (National Heart, Lung, and Blood Institute)[1, 28], Ervin Fox (University of Mississippi Medical Center)[3, 9, 29], Vincent Funari (Cedars-Sinai Medical Center)[1, 11, 31], Santhi K. Ganesh (University of

Michigan)[2, 22], Julius Gardin (Hackensack University Medical Center)[25], David Goff (Wake Forest University)[25], Ora Gordon (Cedars-Sinai Medical Center)[11, 31], Wayne Grody (University of California Los Angeles)[11, 31], Myron Gross (University of Minnesota)[1, 5, 14, 25], Xiuqing Guo (Cedars-Sinai Medical Center)[3, 15, 31], Ira M. Hall (University of Virginia), Nancy L. Heard-Costa (Boston University)[1, 11, 28], Susan R. Heckbert (University of Washington)[10, 14, 20, 23], Nicholas Heintz (University of Vermont), David M. Herrington (Wake Forest University)[5, 31], DeMarc Hickson (Jackson State University, University of Mississippi Medical Center)[29], Jie Huang (National Heart, Lung, and Blood Institute)[5, 28], Shih-Jen Hwang (Boston University, National Heart, Lung, and Blood Institute)[3, 28], David R. Jacobs (University of Minnesota)[25], Nancy S. Jenny (University of Vermont)[1, 2, 23], Andrew D. Johnson (National Heart, Lung, and Blood Institute)[2, 5, 11, 28], Craig W. Johnson (University of Washington)[15, 31], Steven Kawut (University of Pennsylvania)[10,31], Richard Kronmal (University of Washington)[31], Raluca Kurz (Cedars-Sinai Medical Center)[11, 31], Ethan M. Lange (University of North Carolina Chapel Hill)[3, 5, 9, 13, 34], Leslie A. Lange (University of North Carolina Chapel Hill)[1, 2, 3, 5, 9, 12, 13, 15, 17, 18, 20, 25, 34], Martin G. Larson (Boston University)[3, 15, 28], Mark Lawson (University of Virginia), Cora E. Lewis (University of Alabama at Birmingham)[25,34], Daniel Levy (National Heart, Lung, and Blood Institute)[3, 15, 17, 28], Dalin Li (Cedars-Sinai Medical Center)[11, 15, 31], Honghuang Lin (Boston University)[20, 28], Chunyu Liu (National Heart, Lung, and Blood Institute)[3, 28], Jiankang Liu (University of Mississippi Medical Center)[1, 29], Kiang Liu (Northwestern University)[25], Xiaoming Liu (University of Texas Health Science Center Houston)[15, 22], Yongmei Liu (Wake Forest University)[2, 5, 31], William T. Longstreth (University of Washington)[18, 23], Cay Loria (National Heart, Lung, and Blood Institute)[25], Thomas Lumley (University of Auckland)[9, 23], Kathryn Lunetta (Boston University)[28], Aaron J. Mackey (University of Virginia)[16, 18], Rachel Mackey (University of Pittsburgh)[1, 23, 31], Ani Manichaikul (University of Virginia)[8, 15, 18, 31], Taylor Maxwell (University of Texas Health Science Center Houston)[22], Barbara McKnight (University of Washington)[15, 23], James B. Meigs (Brigham and Women's Hospital, Harvard Medical School, Massachusetts General Hospital)[1, 28], Alanna C. Morrison (University of Texas Health Science Center Houston)[3, 15, 17], Solomon K. Musani (University of Mississippi Medical Center)[3, 29], Josyf C. Mychaleckyj (University of Virginia)[13, 15, 31], Jennifer A. Nettleton (University of Texas Health Science Center Houston)[9, 22], Kari North (University of North Carolina Chapel Hill)[1, 3, 9, 10, 13, 15, 17, 34], Christopher J. O'Donnell (Massachusetts General Hospital, National Heart, Lung, and Blood Institute)[2, 5, 9, 11, 12, 14, 15, 17, 20, 28], Daniel O'Leary (Tufts University School of Medicine)[25, 31], Frank S. Ong (Cedars-Sinai Medical Center)[3, 11, 31], Walter Palmas (Columbia University)[3, 15, 31], James S. Pankow (University of Minnesota)[1, 22], Nathan D. Pankratz (Indiana University School of Medicine)[15, 25], Shom Paul (University of Virginia), Marco Perez (Stanford University School of Medicine), Sharina D. Person (University of Alabama at Birmingham, University of Alabama at Tuscaloosa)[25], Joseph Polak (Tufts University School of Medicine)[31], Wendy S. Post (Johns Hopkins University)[3, 9, 11, 14, 20, 31], Bruce M. Psaty (Group Health Research Institute, University of Washington)[3, 5, 9, 11, 14, 15, 23], Aaron R. Quinlan (University of Virginia)[18, 19], Leslie J. Raffel (Cedars-Sinai Medical Center)[6, 11, 31], Vasan S. Ramachandran (Boston University)[3, 28], Alexander P. Reiner (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 9, 11, 12, 13, 14, 15, 20, 25, 34], Kenneth Rice (University of Washington)[15, 23], Jerome I. Rotter (Cedars-Sinai Medical Center)[1, 3, 6, 8, 11, 15, 31], Jill P. Sanders (University of Vermont)[23], Pamela Schreiner (University of Minnesota)[25], Sudha Seshadri (Boston University)[18, 28], Steve Shea (Brigham and Women's Hospital, Harvard University)[28], Stephen Sidney (Kaiser Permanente Division of Research, Oakland, CA)[25], Kevin Silverstein (University of Minnesota)[25], David S. Siscovick (University of Washington)[5, 1, 25], Nicholas L. Smith (University of Washington)[2, 15, 20, 23], Nona Sotoodehnia (University of Washington)[3, 15, 23], Asoke Srinivasan (Tougaloo College)[29], Herman A. Taylor (Jackson State University, Tougaloo College, University of Mississippi Medical Center)[5, 29], Kent Taylor (Cedars-Sinai Medical Center)[31], Fridtjof Thomas (University of Texas Health Science Center Houston)[3, 22], Russell P. Tracy (University of Vermont)[5, 9, 11, 12, 14, 15, 17, 20, 23], Michael Y. Tsai (University of Minnesota)[9, 31], Kelly A. Volcik (University of Texas Health Science Center Houston)[22], Chrstina L Wassel (University of California San Diego)[9, 15, 31], Karol Watson (University of California Los Angeles)[31], Gina Wei (National Heart, Lung, and Blood Institute)[25], Wendy White (Tougaloo College)[29], Kerri L. Wiggins (University of Vermont)[23], Jemma B. Wilk (Boston

University)[10, 28], O. Dale Williams (Florida International University)[25], Gregory Wilson (Jackson State University)[29], James G. Wilson (University of Mississippi Medical Center)[1, 2, 5, 8, 9, 11, 12, 14, 17, 20, 29], Phillip Wolf (Boston University)[28], Neil A. Zakai (University of Vermont)[2, 23]

ISGS and SWISS

John Hardy (Reta Lila Weston Research Laboratories, Institute of Neurology, University College London)[18], James F. Meschia (Mayo Clinic)[18], Michael Nalls (National Institute on Aging)[2, 18], Stephen S. Rich (University of Virginia)[2, 4, 7, 8, 9, 11, 14, 15, 17, 18, 31], Andrew Singleton (National Institute on Aging)[18], Brad Worrall (University of Virginia)[18]

LungGO

Michael J. Bamshad (Seattle Children's Hospital, University of Washington)[4, 6, 7, 8, 10, 11, 13, 15, 17, 27], Kathleen C. Barnes (Johns Hopkins University)[2, 10, 12, 14, 15, 17, 20, 24, 30, 32], Ibrahim Abdulhamid (Children's Hospital of Michigan)[27], Frank Accurso (University of Colorado)[27], Ran Anbar (Upstate Medical University)[27], Terri Beaty (Johns Hopkins University)[24, 30], Abigail Bigham (University of Washington)[13, 15, 27], Phillip Black (Children's Mercy Hospital)[27], Eugene Bleecker (Wake Forest University)[33], Kati Buckingham (University of Washington)[27], Anne Marie Cairns (Maine Medical Center)[27], Wei-Min Chen (University of Virginia)[13, 15, 18], Daniel Caplan (Emory University)[27], Barbara Chatfield (University of Utah)[27], Aaron Chidekel (A.I. Dupont Institute Medical Center)[27], Michael Cho (Brigham and Women's Hospital, Harvard Medical School)[13, 15, 24], David C. Christiani (Massachusetts General Hospital)[21], James D. Crapo (National Jewish Health)[24, 30], Julia Crouch (Seattle Children's Hospital)6, Denise Daley (University of British Columbia)[30], Anthony Dang (University of North Carolina Chapel Hill)[26], Hong Dang (University of North Carolina Chapel Hill)[26], Alicia De Paula (Ochsner Health System)[27], Joan DeCelie-Germana (Schneider Children's Hospital)[27], Allen Dozor (New York Medical College, Westchester Medical Center)[27,] Mitch Drumm (University of North Carolina Chapel Hill)[26], Maynard Dyson (Cook Children's Med. Center)[27], Julia Emerson (Seattle Children's Hospital, University of Washington)[27], Mary J. Emond (University of Washington)[10, 13, 15, 17, 27], Thomas Ferkol (St. Louis Children's Hospital, Washington University School of Medicine)[27], Robert Fink (Children's Medical Center of Dayton)[27], Cassandra Foster (Johns Hopkins University)[30], Deborah Froh (University of Virginia)[27], Li Gao (Johns Hopkins University)[24, 30, 32], William Gershan (Children's Hospital of Wisconsin)[27], Ronald L. Gibson (Seattle Children's Hospital, University of Washington)[10, 27], Elizabeth Godwin (University of North Carolina Chapel Hill)[26], Magdalen Gondor (All Children's Hospital Cystic Fibrosis Center)[27], Hector Gutierrez (University of Alabama at Birmingham)[27], Nadia N. Hansel (Johns Hopkins University, Johns Hopkins University School of Public Health)[10, 15, 30], Paul M. Hassoun (Johns Hopkins University)[10, 14, 32], Peter Hiatt (Texas Children's Hospital)[27], John E. Hokanson (University of Colorado)[24], Michelle Howenstine (Indiana University, Riley Hospital for Children)[27], Laura K. Hummer (Johns Hopkins University)[32], Seema M. Jamal (University of Washington)[11], Jamshed Kanga (University of Kentucky)[27], Yoonhee Kim (National Human Genome Research Institute)[24, 32], Michael R. Knowles (University of North Carolina Chapel Hill)[10, 26], Michael Konstan (Rainbow Babies & Children's Hospital)[27], Thomas Lahiri (Vermont Children's Hospital at Fletcher Allen Health Care)[27], Nan Laird (Harvard School of Public Health)[24], Christoph Lange (Harvard School of Public Health)[24], Lin Lin (Harvard Medical School)[21], Xihong Lin (Harvard School of Public Health)[21], Tin L. Louie (University of Washington)[13, 15, 27], David Lynch (National Jewish Health)[24], Barry Make (National Jewish Health)[24], Thomas R. Martin (University of Washington, VA Puget Sound Medical Center)[10, 21], Steve C. Mathai (Johns Hopkins University)[32], Rasika A. Mathias (Johns Hopkins University)[10, 13, 15, 30, 32], John McNamara (Children's Hospitals and Clinics of Minnesota)[27], Sharon McNamara (Seattle Children's Hospital)[27], Deborah Meyers (Wake Forest University)[33], Susan Millard (DeVos Children's Butterworth Hospital, Spectrum Health Systems)[27], Peter Mogayzel (Johns Hopkins University)[27], Richard Moss (Stanford University)[27], Tanda Murray (Johns Hopkins University)[30], Dennis Nielson (University of California at San Francisco)[27], Blakeslee Noyes (Cardinal Glennon Children's Hospital)[27], Wanda O'Neal (University of North Carolina Chapel Hill)[26], David Orenstein (Children's Hospital of Pittsburgh)[27], Brian O'Sullivan

(University of Massachusetts Memorial Health Care)[27], Rhonda Pace (University of North Carolina Chapel Hill)[26], Peter Pare (St. Paul's Hospital)[30], H. Worth Parker (Dartmouth-Hitchcock Medical Center, New Hampshire Cystic Fibrosis Center)[27], Mary Ann Passero (Rhode Island Hospital)[27], Elizabeth Perkett (Vanderbilt University)[27,] Adrienne Prestridge (Children's Memorial Hospital)[27], Nicholas M. Rafaels (Johns Hopkins University)[30], Bonnie Ramsey (Seattle Children's Hospital, University of Washington)[27], Elizabeth Regan (National Jewish Health)[24], Clement Ren (University of Rochester)[27], George Retsch-Bogart (University of North Carolina Chapel Hill)[27], Michael Rock (University of Wisconsin Hospital and Clinics)[27], Antony Rosen (Johns Hopkins University)[32], Margaret Rosenfeld (Seattle Children's Hospital, University of Washington)[27], Ingo Ruczinski (Johns Hopkins University School of Public Health)[13, 15, 30], Andrew Sanford (University of British Columbia)[30], David Schaeffer (Nemours Children's Clinic)[27], Cindy Sell (University of North Carolina Chapel Hill)[26], Daniel Sheehan (Children's Hospital of Buffalo)[27], Edwin K. Silverman (Brigham and Women's Hospital, Harvard Medical School)[24, 30], Don Sin (Children's Medical Center of Dayton)[30], Terry Spencer (Elliot Health System)[27], Jackie Stonebraker (University of North Carolina Chapel Hill)[26], Holly K. Tabor (Seattle Children's Hospital, University of Washington)[6, 10, 11, 17, 27], Laurie Varlotta (St. Christopher's Hospital for Children)[27], Candelaria I. Vergara (Johns Hopkins University)[30], Robert Weiss [30], Fred Wigley (Johns Hopkins University)[32], Robert A. Wise (Johns Hopkins University)[30], Fred A. Wright (University of North Carolina Chapel Hill)[26], Mark M. Wurfel (University of Washington)[10, 14, 21], Robert Zanni (Monmouth Medical Center)[27], Fei Zou (University of North Carolina Chapel Hill)[26]

SeattleGO

Deborah A. Nickerson (University of Washington)[3, 4, 5, 7, 8, 9, 11, 15, 17, 18, 19], Mark J. Rieder (University of Washington)[4, 11, 13, 15, 16, 17, 19], Phil Green (University of Washington), Jay Shendure (University of Washington)[1, 8, 14, 16, 17], Joshua M. Akey (University of Washington)[13, 14, 15], Michael J. Bamshad (Seattle Children's Hospital, University of Washington)[4, 6, 7, 8, 10, 11, 13, 15, 17, 27], Carlos D. Bustamante (Stanford University School of Medicine)[3, 13, 15], David R. Crosslin (University of Washington)[2, 9], Evan E. Eichler (University of Washington)[19], P. Keolu Fox[2], Wenqing Fu (University of Washington)[13], Adam Gordon (University of Washington)[11], Simon Gravel (Stanford University School of Medicine)[13, 15], Gail P. Jarvik (University of Washington)[9, 15], Jill M. Johnsen (Puget Sound Blood Center, University of Washington)[2], Mengyuan Kan (Baylor College of Medicine)[13], Eimear E. Kenny (Stanford University School of Medicine)[3, 13, 15], Jeffrey M. Kidd (Stanford University School of Medicine)[13, 15], Fremiet Lara-Garduno (Baylor College of Medicine)[15], Suzanne M. Leal (Baylor College of Medicine)[1, 13, 15, 16, 17, 19, 20], Dajiang J. Liu (Baylor College of Medicine)[13, 15], Sean McGee (University of Washington)[13, 15, 19], Timothy D. O'Connor (University of Washington)[13], Bryan Paeper (University of Washington)[16], Peggy D. Robertson (University of Washington)[4], Joshua D. Smith (University of Washington)[4, 16, 19], Jeffrey C. Staples (University of Washington), Jacob A. Tennessen (University of Washington)[13], Emily H. Turner (University of Washington)[4, 16], Gao Wang (Baylor College of Medicine)[1,13,20], Qian Yi (University of Washington)[4]

WHISP

Rebecca Jackson (Ohio State University)[1, 2, 4, 5, 8, 12, 14, 15, 17, 18, 20, 34], Kari North (University of North Carolina Chapel Hill)[1, 3, 9, 10, 13, 15, 17, 34,] Ulrike Peters (Fred Hutchinson Cancer Research Center)[1, 3, 11, 12, 13, 15, 17, 18, 34], Christopher S. Carlson (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 12, 13, 14, 15, 16, 17, 18, 19, 34], Garnet Anderson (Fred Hutchinson Cancer Research Center)[34], Hoda Anton-Culver (University of California at Irvine)[34], Themistocles L. Assimes (Stanford University School of Medicine)[5, 9, 11, 34], Paul L. Auer (Fred Hutchinson Cancer Research Center)[1, 2, 3, 5, 11, 12, 13, 15, 16, 18, 34], Shirley Beresford (Fred Hutchinson Cancer Research Center)[34], Chris Bizon (University of North Carolina Chapel Hill)[3, 9, 13, 15, 34], Henry Black (Rush Medical Center)[34], Robert Brunner (University of Nevada)[34], Robert Brzyski (University of Texas Health Science Center San Antonio)[34], Dale Burwen (National Heart, Lung, and Blood Institute WHI Project Office)[34], Bette Caan (Kaiser Permanente

Division of Research, Oakland, CA)[34], Cara L. Carty (Fred Hutchinson Cancer Research Center)[18, 34], Rowan Chlebowski (Los Angeles Biomedical Research Institute)[34], Steven Cummings (University of California at San Francisco)[34], J. David Curb* (University of Hawaii)[9, 18, 34], Charles B. Eaton (Brown University, Memorial Hospital of Rhode Island)[12, 34], Leslie Ford (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[34], Nora Franceschini (University of North Carolina Chapel Hill)[2, 3, 9, 10, 15, 34], Stephanie M. Fullerton (University of Washington)[6, 11, 34], Margery Gass (University of Cincinnati)[34], Nancy Geller (National Heart, Lung, and Blood Institute WHI Project Office)[34], Gerardo Heiss (University of North Carolina Chapel Hill)[5, 34], Barbara V. Howard (Howard University, MedStar Research Institute)34, Li Hsu (Fred Hutchinson Cancer Research Center)[1, 13, 15, 18, 34], Carolyn M. Hutter (Fred Hutchinson Cancer Research Center)[13, 15, 18, 34], John Ioannidis (Stanford University School of Medicine)[11, 34], Shuo Jiao (Fred Hutchinson Cancer Research Center)[34], Karen C. Johnson (University of Tennessee Health Science Center)[3, 34], Charles Kooperberg (Fred Hutchinson Cancer Research Center)[1, 5, 9, 13, 14, 15, 17, 18, 34], Lewis Kuller (University of Pittsburgh)[34], Andrea LaCroix (Fred Hutchinson Cancer Research Center)[34], Kamakshi Lakshminarayan (University of Minnesota)[18, 34], Dorothy Lane (State University of New York at Stony Brook)[34], Ethan M. Lange (University of North Carolina Chapel Hill)[3, 5, 9, 13, 34], Leslie A. Lange (University of North Carolina Chapel Hill)[1, 2, 3, 5, 9, 12, 13, 15, 17, 18, 20, 25, 34], Norman Lasser (University of Medicine and Dentistry of New Jersey)[34], Erin LeBlanc (Kaiser Permanente Center for Health Research, Portland, OR)[34], Cora E. Lewis (University of Alabama at Birmingham)[25,34], Kuo-Ping Li (University of North Carolina Chapel Hill)[9, 34], Marian Limacher (University of Florida)[34], Dan-Yu Lin (University of North Carolina Chapel Hill)[1, 3, 9, 13, 15, 34], Benjamin A. Logsdon (Fred Hutchinson Cancer Research Center)[2, 34], Shari Ludlam (National Heart, Lung, and Blood Institute WHI Project Office)[34], JoAnn E. Manson (Brigham and Women's Hospital, Harvard School of Public Health)[34], Karen Margolis (University of Minnesota)[34], Lisa Martin (George Washington University Medical Center)[9, 34], Joan McGowan (National Heart, Lung, and Blood Institute WHI Project Office)[34], Keri L. Monda (Amgen, Inc.)[1, 15, 34], Jane Morley Kotchen (Medical College of Wisconsin)[34], Lauren Nathan (University of California Los Angeles)[34], Judith Ockene (Fallon Clinic, University of Massachusetts)[34], Mary Jo O'Sullivan (University of Miami)[34], Lawrence S. Phillips (Emory University)[34], Ross L. Prentice (Fred Hutchinson Cancer Research Center)[34], Alexander P. Reiner (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 9, 11, 12, 13, 14, 15, 20, 25, 34], John Robbins (University of California at Davis)[34], Jennifer G. Robinson (University of Iowa)[9, 11, 18, 34], Jacques E. Rossouw (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[5, 14, 17, 20, 34], Haleh Sangi-Haghpeykar (Baylor College of Medicine)[34], Gloria E. Sarto (University of Wisconsin)[34], Sally Shumaker (Wake Forest University)[34], Michael S. Simon (Wayne State University)[34], Marcia L. Stefanick (Stanford University School of Medicine)[34], Evan Stein (Medical Research Labs)[34], Hua Tang (Stanford University)[2, 34], Kira C. Taylor (University of Louisville)[1, 3, 13, 15, 20, 34], Cynthia A. Thomson (University of Arizona)[34], Timothy A. Thornton (University of Washington)[13, 15, 18, 34], Linda Van Horn (Northwestern University)[34], Mara Vitolins (Wake Forest University)[34], Jean Wactawski-Wende (University of Buffalo)[34], Robert Wallace (University of Iowa)[2, 34], Sylvia Wassertheil-Smoller (Boston University)[18, 34], Donglin Zeng (University of North Carolina Chapel Hill)[9, 34]

*deceased

NHLBI GO ESP Project Team

Deborah Applebaum-Bowden (National Heart, Lung, and Blood Institute)[4, 7, 12, 17], Michael Feolo (National Center for Biotechnology Information)[12], Weiniu Gan (National Heart, Lung, and Blood Institute)[7, 8, 16, 17], Dina N. Paltoo (National Heart, Lung, and Blood Institute)[4, 6, 11, 17], Jacques E. Rossouw (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[5, 14, 17, 20, 34], Phyliss Sholinsky (National Heart, Lung, and Blood Institute)[4, 12, 17], Anne Sturcke (National Center for Biotechnology Information)[12]

ESP Groups

[1]Anthropometry Project Team, [2]Blood Count/Hematology Project Team, [3]Blood Pressure Project Team, [4]Data Flow Working Group, [5]Early MI Project Team, [6]ELSI Working Group, [7]Executive Committee, [8]Family Study Project Team, [9]Lipids Project Team, [10]Lung Project Team, [11]Personal Genomics Project Team, [12]Phenotype and Harmonization Working Group, [13]Population Genetics and Statistical Analysis Working Group, [14]Publications and Presentations Working Group, [15]Quantitative Analysis Ad Hoc Task Group, [16]Sequencing and Genotyping Working Group, [17]Steering Committee, [18]Stroke Project Team, [19]Structural Variation Working Group, [20]Subclinical/Quantitative Project Team

ESP Cohorts

[21]Acute Lung Injury (ALI), [22]Atherosclerosis Risk in Communities (ARIC), [23]Cardiovascular Health Study (CHS), [24]Chronic Obstructive Pulmonary Disease (COPDGene), [25]Coronary Artery Risk Development in Young Adults (CARDIA), [26]Cystic Fibrosis (CF), [27]Early Pseudomonas Infection Control (EPIC), [28]Framingham Heart Study (FHS), [29]Jackson Heart Study (JHS), [30]Lung Health Study (LHS), [31]Multi-Ethnic Study of Atherosclerosis (MESA), [32]Pulmonary Arterial Hypertension (PAH), [33]Severe Asthma Research Program (SARP), [34]Women's Health Initiative (WHI)

Acknowledgments

HeartGO:

Atherosclerosis Risk in Communities (ARIC): NHLBI (N01 HC-55015, N01 HC-55016, N01HC-55017, N01 HC-55018, N01 HC-55019, N01 HC-55020, N01 HC-55021); **Cardiovascular Health Study (CHS)**: NHLBI (HHSN268201200036C, N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, and grant HL080295), with additional support from NINDS and from NIA (AG-023629, AG-15928, AG-20098, and AG-027058); **Coronary Artery Risk Development in Young Adults (CARDIA)**: NHLBI (N01-HC95095 & N01-HC48047, N01-HC48048, N01-HC48049, and N01-HC48050); **Framingham Heart Study (FHS)**: NHLBI (N01-HC-25195 and grant R01 NS17950) with additional support from NIA (AG08122 and AG033193); **Jackson Heart Study (JHS)**: NHLBI and the National Institute on Minority Health and Health Disparities (N01 HC-95170, N01 HC-95171 and N01 HC-95172); **Multi-Ethnic Study of Atherosclerosis (MESA):** NHLBI (N01-HC-95159 through N01-HC-95169 and RR-024156).

Lung GO:

Cystic Fibrosis (CF): Cystic Fibrosis Foundation (GIBSON07K0, KNOWLE00A0, OBSERV04K0, RDP R026), the NHLBI (R01 HL-068890, R02 HL-095396), NIH National Center for Research Resources (UL1 RR-025014), and the National Human Genome Research Institute (NHGRI) (5R00 HG-004316). **Chronic Obstructive Pulmonary Disease (COPDGene):** NHLBI (U01 HL-089897, U01 HL-089856), and the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, and Sunovian. The COPDGene clinical centers and investigators are available at www.copdgene.org. **Acute Lung Injury (ALI)**: NHLBI (RC2 HL-101779). **Lung Health Study (LHS)**: NHLBI (RC2 HL-066583), the NHGRI (HG-004738), and the NHLBI Division of Lung Diseases (HR-46002). **Pulmonary Arterial Hypertension (PAH):** NIH (P50 HL-084946, K23 AR-52742), and the NHLBI (F32 HL-083714). **Asthma:** NHLBI (RC2 HL-101651), and the NIH (HL-077916, HL-69197, HL-76285, M01 RR-07122).

SWISS and ISGS:

Siblings with Ischemic Stroke Study (SWISS): National Institute of Neurological Disorders and Stroke (NINDS) (R01 NS039987); Ischemic Stroke Genetics Study (ISGS): NINDS (R01 NS042733)

WHISP:

Women's Health Initiative (WHI): The WHI Sequencing Project is funded by the NHLBI (HL-102924) as well as the National Institutes of Health (NIH), U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221, and HHSN268201100046C. 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: https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Cardiovascular Health Study:
This research was supported by contracts HHSN268201200036C, N01HC85239, N01 HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grant HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/PI.htm.

PROCARDIS:
PROCARDIS was funded by European Commission (LSHM-CT- 2007- 037273), the Swedish Heart-Lung Foundation, the Swedish Research Council (8691), the Knut and Alice Wallenberg Foundation, the Foundation for Strategic Research, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular Programme of Karolinska Institutet and the Stockholm County Council (ALF project program), and the Stockholm County Council (560183) and the Strategic Cardiovascular and Diabetes Programmes of Karolinska Institutet and Stockholm County Council. www.procardis.org.

WTCCC:
Recruitment of the BHF-FHS was funded by the British Heart Foundation (BHF). Genotyping of cases was partly by the Wellcome Trust and partly by the BHF and the UK National Institute for Health Research (NIHR). Genotyping of the controls was funded by the Wellcome Trust. NJS holds a Chair supported by the BHF and NM is funded by the NIHR Leicester Cardiovascular Biomedical Research Unit.

I. Exome Sequencing at the University of Washington

Quality control of sample DNA

Quality control (QC) of the DNA samples included quantification with PicoGreen, confirmation of high-molecular weight DNA, tests for PCR amplification (four amplicons), and sex determination using a Taq-man assay[1]. Prior to preparation for exome sequencing, all samples were genotyped (Illumina BeadXpress) with 96 high frequency (30-50% minor allele frequency) exome specific SNPs, derived from the content found on genome wide arrays for both the Illumina and Affymetrix platforms. Genotype data at these variant sites were used to ensure sample tracking integrity through sample preparation and the sequencing pipeline. Samples failed QC if: (1) the total mass, concentration or integrity of DNA was low; (2) genotype call rates were low (<90%); or (3) sex-typing was inconsistent with the sample manifest. Following QC, all remaining genomic DNA (~ 3.5 ug) was reformatted into 96 well plates for library preparation and for exome capture.

Library production and exome capture

All protocols for library construction and exome capture were automated on a Perkin-Elmer Janus II liquid handling robot, and performed in 96-well plate format. Samples were prepared by subjecting genomic DNA (~3.5 ug) to a series of shotgun library construction steps, including fragmentation through acoustic fragmentation (Covaris), end-polishing and A tailing, ligation of sequencing adaptors, and PCR amplification. Sample shotgun libraries were captured for exome enrichment using one of three in-solution capture products: CCDS 2008 (~26Mb), the SeqCap EZ Human Exome Library v1.0 (~32 Mb), or the SeqCap EZ Human Exome Libraray v2.0 (~34Mb). Briefly, 1 ug of shotgun library was hybridized to biotinylated capture probes for 72 hours and recovered via streptavidin beads. Unbound DNA was washed away, and the captured

DNA PCR amplified. Following capture, washing, and PCR, libraries were assessed again on the Agilent Bioanalyzer for concentration, molecular weight distribution, and the presence of PCR artifacts. The fragment size distributions of the libraries were highly consistent (typically 125 ± 15 bp).

Clustering and sequencing

Library concentration and flow-cell loading cluster densities were determined using a standardized qPCR protocol (Kapa Biosystems). Using the automated Illumina cBot cluster station, non-multiplexed samples were processed in batches of eight (one for each lane of the flow-cell), diluted and denatured to their final effective loading concentrations. Hybridization was followed by cluster generation via bridge PCR as per standard protocols (Illumina). Enriched libraries were sequenced on an Illumina GAIIx paired-end 76 bp reads.

Read mapping and variant analysis for QC purposes

Samples were processed from real-time base-calls (RTA 1.7 software [Bustard], converted to qseq.txt files, and aligned to a human reference (hg19) using BWA (Burrows-Wheeler Aligner)[2]. Read-pairs not mapping within two standard deviations of the average library size (~125 ± 15 bp) were removed. Data were processed using the Genome Analysis ToolKit[3] (GATK refv1.2905). All aligned read data were subjected to "duplicate removal", i.e. the removal of reads with duplicate start positions, indel realignment (GATK IndelRealigner) and base qualities recalibration (GATK TableRecalibration). Variant detection and genotyping were performed using the UnifiedGenotyper (UG) tool from GATK and on the targeted exome regions. Variant data for each sample were formatted (variant call format [VCF]) as "raw" calls for all samples, and sites flagged using the filtration walker (GATK) to mark sites that are of lower quality/false positives (i.e. low quality scores (<50), allelic imbalance (0.75), long homopolymer runs (>3),

and/or low quality by depth (QD<5)). Samples were considered complete when exome targeted read coverage was >8x over >90% of the exome target. Typically, the mean target coverage was 60-80x.

Data analysis QC at University of Washington

Individual exome sequence data were evaluated against the following QC metrics which included an assessment of: (1) total reads, or a minimum of 30M PE reads; (2) library complexity: the ratio of unique reads to total reads mapped to target; (3) capture efficiency: the ratio of reads mapped to target versus the reads mapped to human; (4) coverage distribution: 90% at >8x required for completion; (5) capture uniformity; (6) raw error rates; (7) Ts/Tv ratio (3.2 for known sites and 2.9 for novel sites); (8) distribution of known and novel variants relative to dbSNP; (9) fingerprint concordance with 96 QC SNPs >99%; (10) homozygosity; (11) heterozygosity. All QC metrics for both single-lane and merged data were reviewed to identify data deviations from known or historical norms. Lanes/samples that failed QC were re-queued for library prep for further sequencing.

II. Exome Sequencing at the Broad Institute

Receipt/QC of Sample DNA

Samples were shipped to the Biological Samples Platform laboratory at the Broad Institute of MIT and Harvard. DNA concentration was determined by the Picogreen assay (Invitrogen, Carlsbad, California) before storage in 2D-barcoded 0.75 mL Matrix tubes at −20°C in the SmaRTStore™ (RTS, Manchester, UK) automated sample handling system. We performed initial QC on all samples involving sample quantification (PicoGreen), confirmation of high-molecular

weight DNA and fingerprint genotyping and gender determination (Illumina iSelect). Samples were failed if the total mass, concentration, integrity of DNA or quality of preliminary genotyping data was too low.

Library construction and in-solution hybrid selection

Starting with 3µg of genomic DNA, library construction and in-solution hybrid selection were performed as described by Fisher et al[4]. A subset of samples, however, was prepared using the Fisher et al. protocol with some slight modifications. Initial genomic DNA input into shearing was reduced from 3µg to 100ng in 50µL of solution. In addition, for adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters with unique 8 base index sequences embedded within the adapter.

Preparation of libraries for cluster amplification and sequencing

After in-solution hybrid selection, libraries were quantified using quantitative PCR (kit purchased from KAPA biosystems) with probes specific to the ends of the adapters. This assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 2nM and then denatured using 0.1 N NaOH using Perkin-Elmer's MultiProbe liquid handling platform. A subset of the samples prepared using forked, indexed adapters was quantified using qPCR, normalized to 2nM using Perkin-Elmer's Mini-Janus liquid handling platform, and pooled by equal volume using the Agilent Bravo. Pools were then denatured using 0.1 N NaOH. Denatured samples were diluted into strip tubes using the Perkin-Elmer MultiProbe.

Cluster amplification and sequencing

Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) using either Genome Analyzer v3, Genome Analyzer v4, or HiSeq 2000 v2 cluster chemistry and flowcells. After cluster amplification, SYBR Green dye was added to all flowcell lanes, and a portion of each lane visualized using a light microscope, in order to confirm target cluster density. Flowcells were sequenced either on Genome Analyzer II using v3 and v4 Sequencing-by-Synthesis Kits, then analyzed using RTA v1.7.48, or on HiSeq 2000 using HiSeq 2000 v2 Sequencing-by-Synthesis Kits, then analyzed using RTA v1.10.15. All samples were run on 76 cycle, paired end runs. For samples prepared using forked, indexed adapters, Illumina's Multiplexing Sequencing Primer Kit was also used.

Read mapping and variant analysis

Samples were processed from real-time base-calls (RTA 1.7 software [Bustard], converted to qseq.txt files, and aligned to a human reference (hg19) using BWA (Burrows-Wheeler Aligner) [2]. Aligned reads duplicating the start position of another read were flagged as duplicates and not analyzed ("duplicate removal"). Data were processed using the Genome Analysis ToolKit[3] (GATK **v1.1.3**). Reads were locally realigned (GATK IndelRealigner) and their base qualities were recalibrated (GATK TableRecalibration). Variant detection and genotyping were performed on both exomes and flanking 50bp of intronic sequence using the UnifiedGenotyper (UG) tool from the GATK. Variant data for each sample was formatted (variant call format [VCF]) as "raw" calls for all samples. SNP and Indel sites were flagged using the Variant Filtration walker (GATK) to mark sites of low quality that are likely false positives. SNPs were marked as potential errors if they exhibited strong strand bias (SB >= 0.10), low average quality (QD <5.0), or fell in a homopolymer run (HRun > 4). Indels were marked as potential errors for low quality

(QUAL < 30.0), low average quality (QD < 2.0), or if the site exhibited strong strand bias (SB > -1.0). Samples were considered complete when exome targeted read coverage was ≥20x over ≥80% of the exome target.

Data Analysis QC

Processed sequence data were required to match known fingerprint genotypes for their respective samples, and to achieve a sequence coverage of >20x for >70% of targeted bases. Variant calls were evaluated on both bulk and per-sample properties: novel and known variant counts, Ts/Tv ratio, Het/Hom ratio, and Deletion/Insertion ratio. Both bulk and sample metrics were compared to historical values for exome sequencing projects at the Broad. No significant deviation of the ESP calls or ESP samples from historical values were noted.

III. Joint Calling of Variants for Entire ESP Project at the University of Michigan

SNVs were called using the UMAKE pipeline at University of Michigan, which allowed all samples to be analyzed simultaneously, both for variant calling and filtering. Briefly, we used BAM files summarizing BWA alignments generated at the University of Washington and the Broad Institute as input. These BAM files summarized alignments generated by BWA, refined by duplicate removal, recalibration, and indel re-alignment. We excluded all reads that were not confidently mapped (Phred-scaled mapping quality < 20) from further analysis. To avoid PCR artifacts, we clipped overlapping ends in paired reads. We then computed genotype likelihoods for exome targeted regions and 50 flanking bases, accounting for per base alignment quality (BAQ) using samtools[5]. Variable sites and their allele frequencies were identified using a

maximum-likelihood model, implemented in glfMultiples[6]. These analyses assumed a uniform prior probability of polymorphism at each site.

Variant and Sample Level Quality Control

SVM Filter: We used a support vector machine (SVM) classifier to separate likely true positive and false-positive variant sites using a battery of SNP quality metrics. These include allelic balance (the proportional representation of each allele in likely heterozygotes), base quality distribution for sites supporting the reference and alternate alleles, and the distribution of supporting evidence between strands and sequencing cycle, amongst others. We used as the positive training set variants identified by dbSNP or 1000 Genomes and we used variants that failed multiple filters as the negative training set. We found this method to be effective at removing sequencing artifacts while preserving good-quality data, as indicated by the Ts/Tv ratio for previously known and newly identified variant sites, the proportion of high frequency variants overlapping with dbSNP, and the ratio of synonymous to non-synonymous variants, as well as attempts at validation of a subset of sites. A total of 1,908,614 SNVs passed the SVM filter.

Filter based on Depth10: There were 52 pairs of duplicate samples in the final set of exomes from the Exome Sequencing Project - ESP6800 dataset. For each of these 52 pairs, we calculated the non-reference genotype concordance rates. The non-reference concordance (NRC) rate is a measure of concordance that only considers genotypes where at least one sample was called a heterozygote or a non-reference homozygote. Missing genotypes do not contribute to this calculation. Standard concordance rates for rare-variants tend to be dominated by an abundance of reference homozygous calls, thus we chose non-reference concordance rates as a measure of

genotyping specificity. To investigate whether to use a genotype filter based on read depth, we calculated NRC rates across a variety of read depth cutoffs (depth = 1, 5, 10, 15, 20, 25, 30). For each cutoff *c*, we replaced any genotype with an associated read depth less than *c*, with a missing value. As a measure of sensitivity, we calculated the total number of genotypes retained after enforcing the read depth cutoff. For the 52 pairs of duplicates, **Figure 1** shows the NRC rate by the percent of total genotypes retained for a variety of read depth cutoffs. From this plot we concluded that a filter based on a read depth of 10 markedly improved concordance rates while maintaining over 90% of the total genotypes. Thus we replaced genotypes with a corresponding read depth less than 10 with a missing value in the gene-based analysis. This was not applied to the single-variant analysis as this analysis involved common variants.

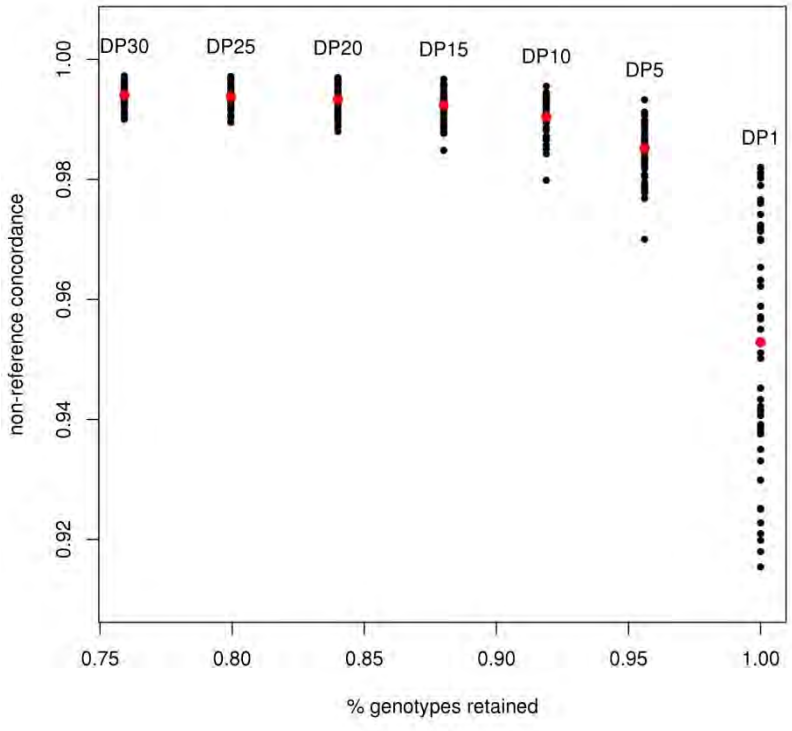


Figure 1: Non-reference concordance rates are plotted on the y-axis versus % genotypes retained on the x-axis for each of the 52 duplicate samples in the ESP6800. All data-points are shown in black, the mean non-reference concordance for each read depth cutoff is shown in read.

Filter based on mean Depth 500: We investigated further a variant level filter based on average per-SNP read depth. We used transition-transversion (ti-tv) ratios as a means of considering the overall quality of set of SNPs. In the exome, ti-tv ratios near a value of three are thought to be indicative of true positive SNPs. We noticed a general trend of increasing ti-tv ratios as the average per-variant depth increased, with a decrease in ti-tv ratios at very high average depths (**Figure 2**). This is most likely do to pseudo-SNPs. This happens when regions of the genome with close sequence homology (e.g., only 1 base-pair differentiates the two sequences) are subjected to short-read shotgun sequencing and the alignment software preferentially maps the reads from both regions to only one location. This results in a pile-up of reads at the preferential location, that appear to be polymorphic and an incorrect heterozygous call is made. To guard against these pseudo-SNPs we filtered out all variants with an average depth greater than 500. The low ti-tv ratios at very low depths were accounted for by enforcing the read depth 10 filter described above.

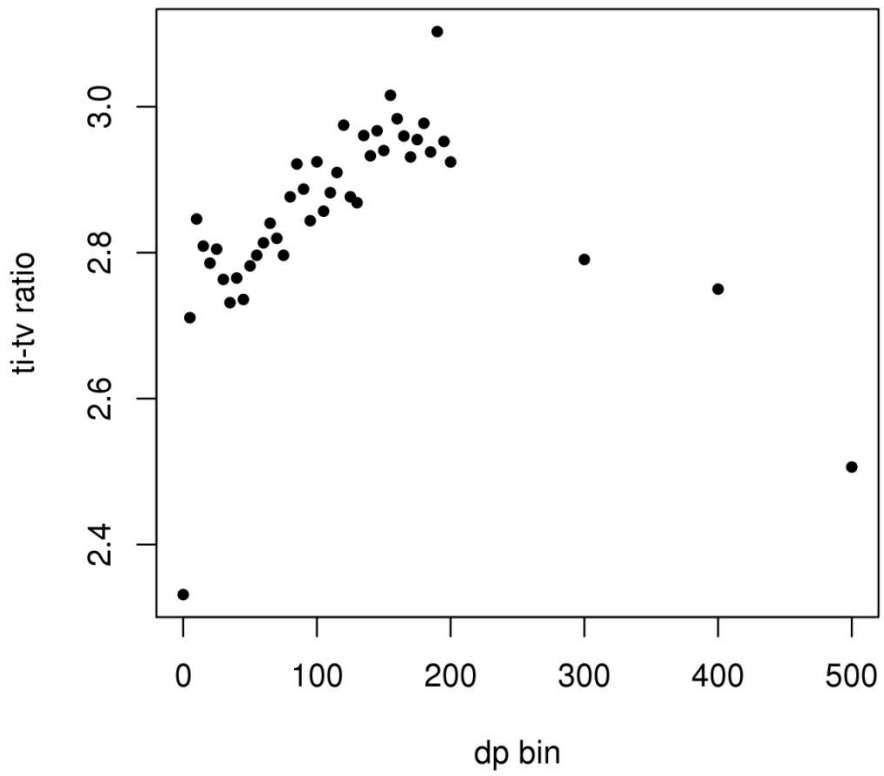


Figure 2: Transition to transversion ratios against average variant-level read depth bins.

Principal Component Analysis and Ancestry Designation: After performing the SVM and read depth 10 filter on the ESP6800 call-set, we ran a principal component analysis (PCA) to determine sample-level outliers and to cross-check our self-reported ancestry. To do so we only included SNPs with a minor allele frequency (MAF) greater than or equal to 0.1% and a call-rate of greater than 95%. Only autosomal SNPs were included in the PCA. We ran the PCA in PLINK[7] after pruning out SNPs in linkage disequilibrium (LD). This was done by looking in windows of 50 SNPs and shifting the windows 5 SNPs at each step. If a pair of SNPs had a genotype R^2 value greater than 0.5 one of the SNPs was removed. The resulting SNPs were used to determine a matrix of genome-wide Identity by State (IBS) pairwise distances which were

subsequently input to the PLINK multidimensional scaling (MDS) algorithm. Figure 2 shows the first two dimensions from the MDS (analogous to the first two principal components). The first two PCs clearly separate the African American (AA) samples from the European American (EA) samples. However, there is a clear group of admixed individuals between these two clusters where many self-reported Hispanic individuals were clustered. We removed from all subsequent analyses those individuals of indeterminate genetic ancestry located between the two vertical lines in **Figure 3**. For simplicity, we also removed from analysis any individual self-reporting race different from AA or EA. Of the remaining samples, all points to the left of the left-most vertical bar were designated as having AA genetic ancestry. All points to the right of the right-most vertical bar were designated as having EA genetic ancestry. Those samples with discrepant self-reported and designated ancestry were removed from all subsequent analyses.

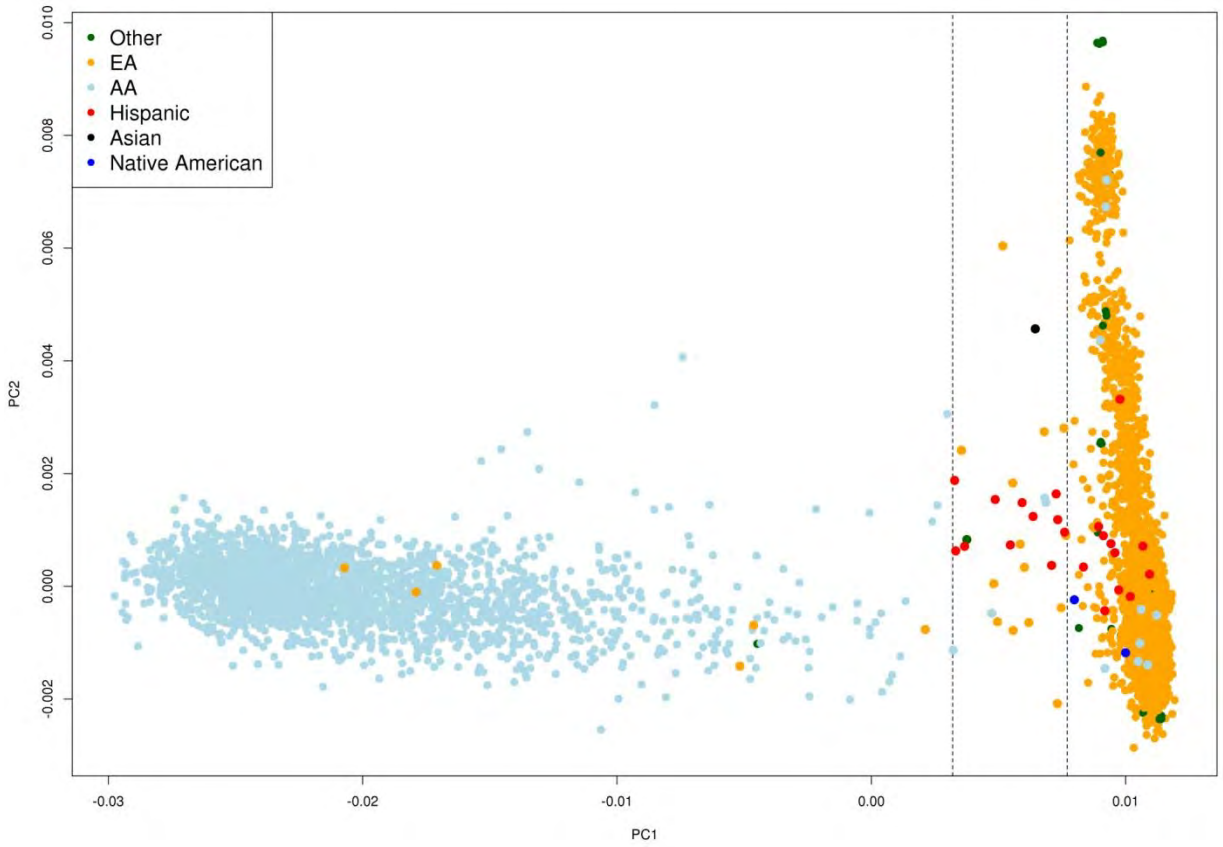


Figure 3: The first two principal components from the ESP6800 call-set. Self-reported EAs are shown in orange, AAs in light-blue, Hispanics in red, Asians in black, and Native Americans in dark-blue. Missing self-reported race is shown in green.

Analysis of Relatedness: After designating samples to AA and EA ancestry groups, we ran a race stratified kinship analysis to identify any cryptically related individuals in the ESP6800 call-set. To do so we only considered variants that passed the SVM filter, the Depth 500 filter, and after replacing genotypes with a corresponding read depth less than 10 with a missing value. Furthermore, we only considered variants that were in the intersection of the four capture targets that were used. The degree of relatedness was estimated using the KING software[8]. As with the MDS analysis, only LD-pruned autosomal variants with MAF > 0.001 were used as input. Pairs

of samples with kinship coefficient range of > 0.354, [0.177, 0.354], [0.0884, 0.177), [0.0442, 0.0884) were designated as duplicates, 1st-degree, 2nd-degree, and 3rd-degree relatives, respectively. **Figure 4** displays the estimated kinship coefficients plotted against the proportion of SNPS with zero identical by state.

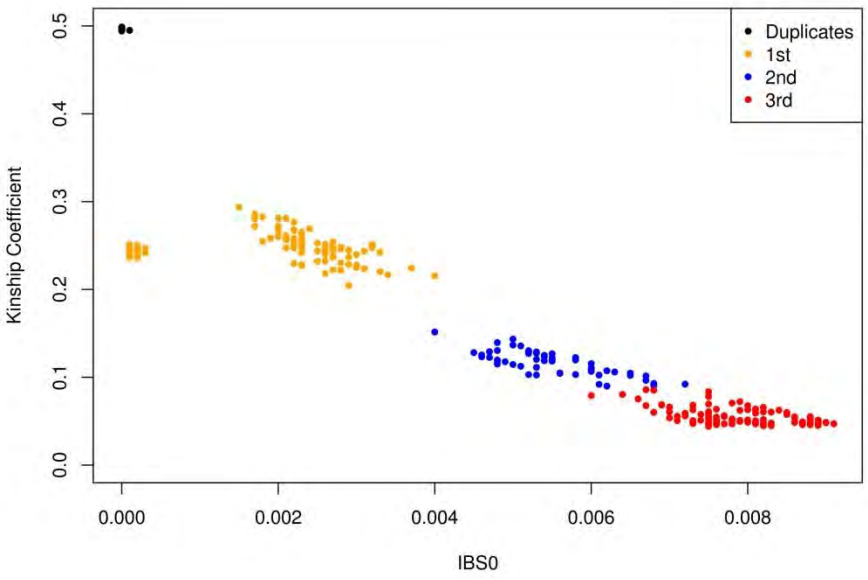


Figure 4: Estimated kinship coefficients plotted against the proportion of SNPs with zero identical by state. Duplicates are shown in black, 1st-degree relatives in yellow, 2nd-degree relatives in blue, and 3rd-degree relatives in red.

Hardy-Weinberg Variant Level Filter: After running the kinship analysis, we considered whether variants were in Hardy-Weinberg equilibrium (HWE). This analysis was stratified by race, and only 1 individual from each duplicate/relative pair was included (the sample with the higher call-rate). Variants with a p-value testing HWE < $5x10^{-20}$ based on an exact test for HWE[9], were excluded from further analyses.

Sample Level Missingness: After enforcing the read depth 10 cutoff, we calculated sample level missing rates (**Figure 5**). There is a clear difference between the four target capture arrays that were used. Within each of the four targets, only one sample (in black) was a clear outlier. This sample was excluded from further analyses.

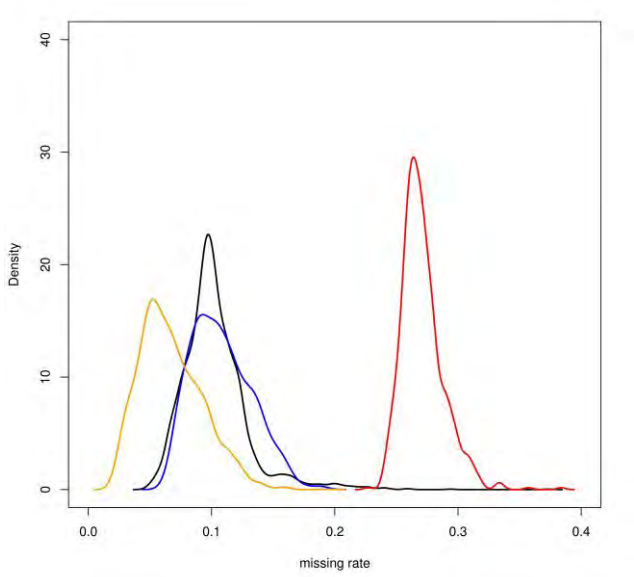


Figure 5: Sample level missing rates after enforcing a DP10 filter on the genotypes. Missing rates for the 4 capture targets are shown in black, red, blue and yellow

Sample Level Homozygosity: For each sample we calculated inbreeding coefficients in PLINK. We used the same set of variants that were included in the MDS analysis. One EA sample was found it have an exceedingly high inbreeding coefficient compared to the other samples (**Figure 6**). This sample was removed from subsequent analyses.

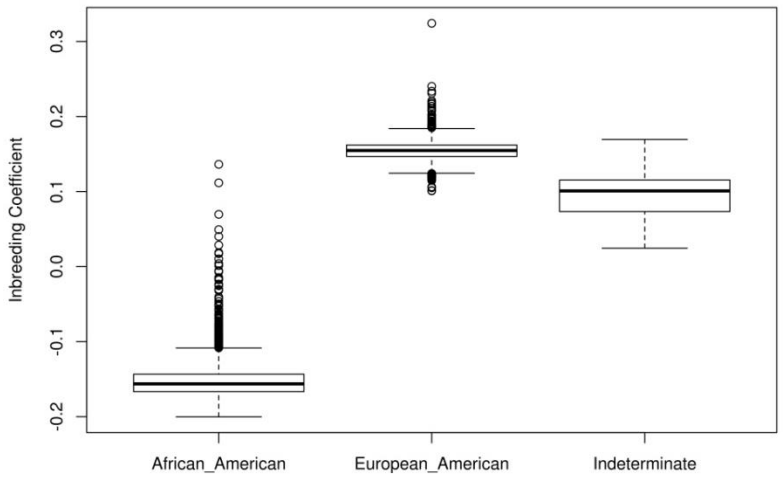


Figure 6: Sample level homozygosity estimates stratified by the three race groups (AA, EA, Indeterminate).

Sex Check: To guard against potential sample swaps, we cross-checked self-reported sex against a normalized measure of read depth on the X and Y chromosomes. Because of the way the samples were processed, we normalized the read depth for the first 2,484 samples differently from the last 4,339 samples. **Figure 7** shows the normalized coverage on the two sex chromosomes. There are very clearly two distinct clusters (males and females) in each plot. Samples where the self-reported sex was clearly different from the XY coverage cluster (highlighted in **Figure 7**) were considered sample swaps and excluded from further analysis.

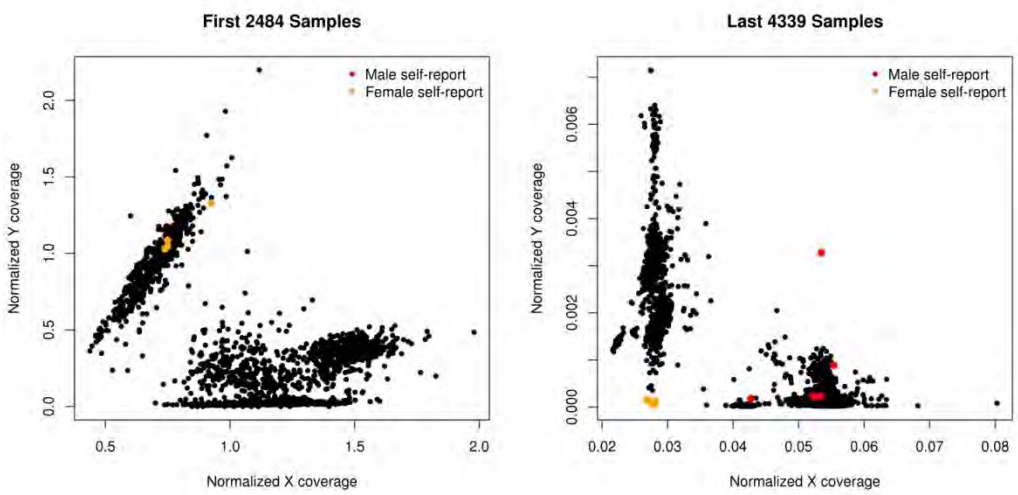


Figure 7: Plots of normalized chrY depth of coverage versus normalized chrX depth of coverage. The first 2,484 samples are shown in the left panel, the second 4,339 samples are shown in the right panel. Samples self-reported as male but falling in the female cluster are displayed in red, self-reported females that fall in the male cluster are displayed in yellow.

GWAS Concordance: When we had access to genome-wide SNP array data we ran concordance checks between the ESP variants that overlapped with the variants typed on the arrays. Samples identified as having very low concordance rates were subsequently dropped from further analysis due to the strong likelihood that they were sample swaps.

Variant Level Missingness: We did not enforce a call-rate filter for the per-variant analyses. For the gene-level analyses, for each gene we first removed samples with >10% missing rate for the variants in that gene. Once these samples were removed we filtered out variants with missing rate > 10%.

Variant Annotation: All variants in the ESP6800 were submitted to the SeattleSeq annotation server (http://snp.gs.washington.edu/SeattleSeqAnnotation134/) on May 29, 2012. We used annotation version 134, the hg19 build of the human reference genome, and the NCBI full genes

(NM, XM) gene model option. For variants mapping to multiple transcripts, we retained the most damaging classification (from most damaging to least: nonsense, splice, missense, synonymous, utr, other).

IV. Exome Array Genotyping

Study samples were processed on the HumanExome BeadChip v1.0 (Illumina, Inc., San Diego, CA) using standard protocols suggested by the manufacturer at local genotyping centers. Genotypes were assigned using GenomeStudio v2010.3 using the calling algorithm/genotyping module version 1.8.4 along with the custom cluster file StanCtrExChp_CEPH.egt. At most genotyping centers, these calls were supplemented by the application of the zCall rare variant calling algorithm.[10] Across ~66,000 samples from the CHARGE Consortium, the raw data files for the samples were assembled into a single project for joint calling. Genotype data for the four *APOC3* mutations (exm957809, exm957810, exm957815, and exm957817) were extracted prior to analysis.

V. Study Participants

Discovery study samples: The U.S. National Heart, Lung, and Blood Institute's Exome Sequencing Project (ESP) sought to use exome sequencing as a tool to discover novel genes and mechanisms contributing to heart, lung, and blood disorders (https://esp.gs.washington.edu/drupal/).[11,12] Participants for the present analysis were 3,734 individuals who had both exome sequence and plasma triglycerides available (**Table S1**). Participants were enrollees in seven population-based cohorts [Atherosclerosis Risk in Communities (ARIC),[13] Coronary Artery Risk Development in Young Adults (CARDIA),[14]

Cardiovascular Health Study (CHS),[15] Framingham Heart Study (FHS),[16] Jackson Heart Study (JHS),[17] Multi-Ethnic Study of Atherosclerosis (MESA),[18] and the Women's Health Initiative (WHI)[19]] and a study of early-onset myocardial infarction, Myocardial Infarction Genetics Consortium, MIGen).[20]

Replication Study Samples: We genotyped 41,671 African-Americans (AA) or participants of European ancestry (EA) from seven replication studies: ARIC (EA and AA), JHS (AA), WHI (EA and AA), Malmo Diet and Cancer Study Cardiovascular Cohort (MDC-CVA, EA),[21] Ottawa Heart Study (EA), Precocious Coronary Artery Disease (PROCARDIS) study (EA),[22] and Italian Atherosclerosis, Thrombosis, and Vascular Biology (ATVB) study (EA).[23] These participants were independent from those sequenced in the discovery study.

VI. APOC3 Genotypes: Replication For Plasma Lipids

To follow-up the strongest result for triglycerides observed in the discovery sample, i.e., *APOC3*, we performed genotyping of four mutations (R19X, IVS2+1 G>A, A43T, and IVS3+1 G>T) using the Illumina HumanExome Beadchip. Three of the four mutations are predicted to severely disrupt APOC3 function, i.e., lead to loss of function (LoF).[24] LoF variants included a nonsense substitution (i.e., R19X) and two DNA sequence variants disrupting a splice site (i.e., IVS2+1 G>A and IVS3+1 G>T). Each of these *APOC3* variants and a fourth, missense variant A43T, were associated with lower plasma triglycerides, suggesting that all four variants lead to loss of APOC3 function.

We genotyped 41,671 African-Americans (AA) or participants of European ancestry (EA) from seven replication studies (**Table S2**). These participants were independent from those sequenced in the discovery study. We performed race-specific linear regression with the outcome

variable of plasma triglycerides (or other lipid fractions), independent variable of variant allele carrier status (coded as 0,1,2), and covariates of age, gender, and at least two principal components of ancestry. We also considered a model where carriers of any of the four LoF mutations were collapsed into a single independent variable – *APOC3* LoF carrier. Statistical evidence across the studies was summarized through meta-analysis with inverse of the variance as weights.

VII. APOC3 Genotypes: Association With CHD

We next tested the association of *APOC3* LoF carrier status with CHD in EA, AA, and Hispanic ancestry (HA) participants from 15 studies. Participants were genotyped using the Illumina HumanExome Beadchip. Descriptions of the studies and the definitions for CHD outcomes are provided in **Table S3**. We calculated P values for the association tests and the confidence intervals for the odds ratios by using exact methods. We performed meta-analyses by using Cochran-Mantel-Haenszel statistics for stratified 2X2 tables. The Cochran-Mantel-Hanszel method combines score statistics rather than Wald statistics and is particularly attractive when the observed odds ratios are zero. All the results were obtained from the Freq procedure in SAS.

As an alternate approach, we performed logistic regression where the outcome variable was either incident CHD or prevalent CHD, the independent variable was *APOC3* LoF carrier status (coded as 0 or 1) and covariates of age, sex, and at least two principal components of ancestry; these analyses yielded similar results (data not shown).

VIII. Plasma apoC-III Protein Concentration And Risk For Incident CHD

Blood was drawn from fasting participants in the Framingham Heart Study Offspring cohort examination cycle 5 (1991 – 1995).[25] Plasma apoC-III protein concentration was assessed in 3,238 individuals using a commercially available immmunochemical assay from Wako Diagnostics (Richmond, USA). All participants underwent continuous surveillance for incident CHD events until December 31, 2010. CHD events included fatal MI, non-fatal MI, angina pectoris, and coronary insufficiency as described previously.[26] We prospectively studied 2,913 persons without prevalent CHD. Using proportional-hazards regression, we examined the relations of plasma apoC-III (natural logarithmically transformed) to risk of incident CHD. We tested two models: (1) age- and sex-adjusted; and (2) multivariable models adjusting for age, sex, smoking, diabetes mellitus, LDL cholesterol, HDL cholesterol, hypertension treatment, systolic and diastolic blood pressure, lipid-lowering treatment, and fasting serum glucose.

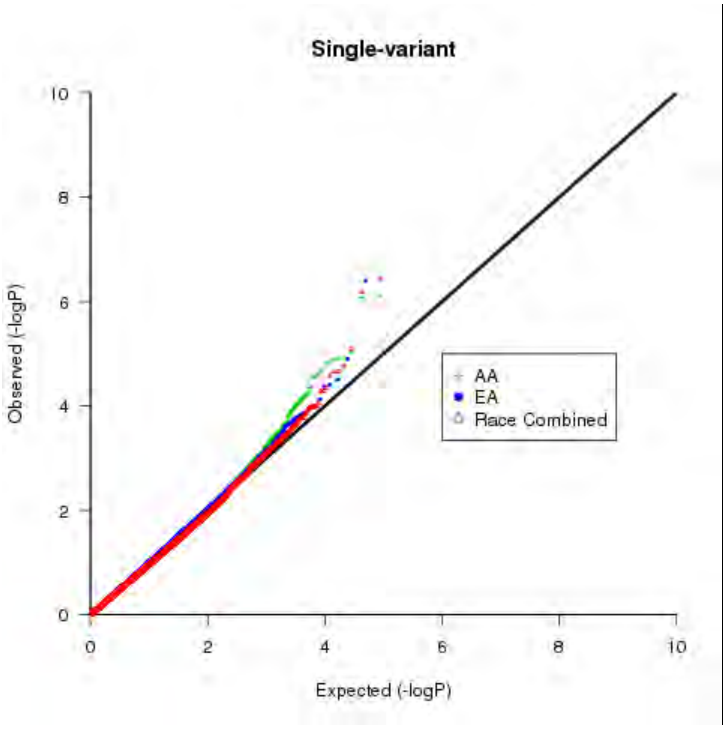
In order to evaluate plasma apoC-III protein in the secondary prevention setting, we studied the association of plasma apoC-III protein with incident total and cardiovascular mortality in the Verona Heart Study. We recruited 794 subjects with angiographic coronary artery disease and measured plasma apoC-III as previously described.[27] During a median follow-up of 59 months, there were 134 deaths, with 92 due to cardiovascular disease (coronary artery disease, heart failure, peripheral artery disease, or cerebrovascular disease). Using proportional-hazards regression, we examined the relations of plasma apoC-III (natural logarithmically transformed) to risk of incident total or cardiovascular mortality. We tested two models: (1) age- and sex-adjusted; and (2) multivariable models adjusting for age, sex, smoking, diabetes mellitus, LDL cholesterol, HDL cholesterol, hypertension, lipid-lowering treatment, and fasting serum glucose.

IX. APOC3 Genotypes and Association With Hepatic Steatosis

Between 2002 and 2005, 1,400 individuals from the Framingham Offspring Study and 2,011 individuals from third generation underwent multi-dectector computed tomograms on which we evaluated liver attenuation as previously described.[28] We tested the association of *APOC3* LoF genotypes with CT liver fat after inverse normal transformation. Covariates in the regression models included age, age^2, gender, and number of alcoholic drinks per week.

Figure S1. Quantile-quantile plot of results testing the association of plasma triglycerides with single coding sequence variants (A) and with variants aggregated at the gene level (B). AA denotes African Americans; EA, European ancestry

1A.



1B.

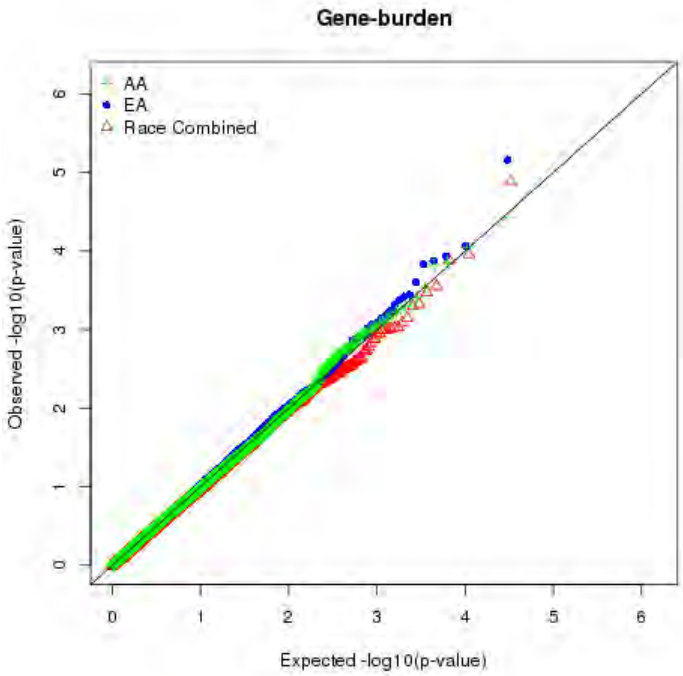

Gene-burden
AA
EA
Race Combined
Observed -log10(p-value)
Expected -log10(p-value)
0
1
2
3
4
5
6

Figure S2. Distribution of plasma apolipoprotein C-III concentrations in the Framingham Heart Study Offspring cohort (n=3,237) in mg/dl.

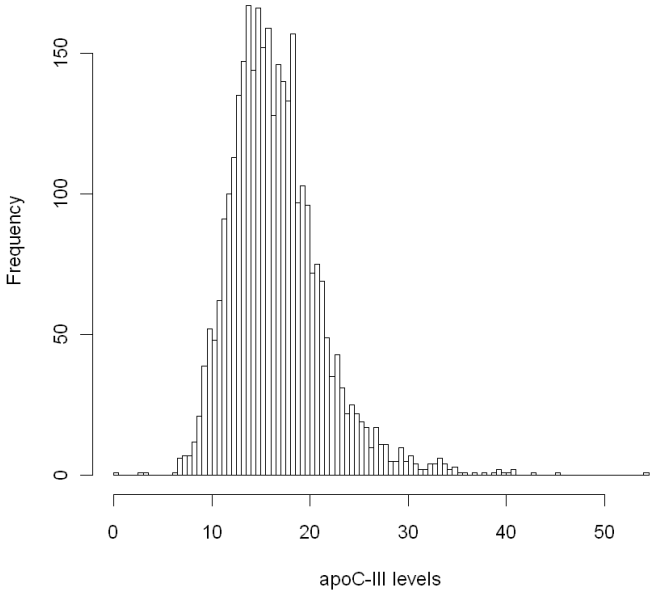


Figure S3. Mean plasma apolipoprotein C-III concentration in carriers of *APOC3* variants [R19X (n=2) or splice site IVS2+1 G>A (n=11)] compared non-carriers in the Framingham Heart Study Offspring cohort. Mean plasma apoC-III concentration in carriers (n=13) was 8.96 mg/dl (SD 1.69) whereas the mean concentration in non-carriers (n=2,694) was 16.63 mg/dl (SD 4.71) ($P=8 \times 10^{-10}$).

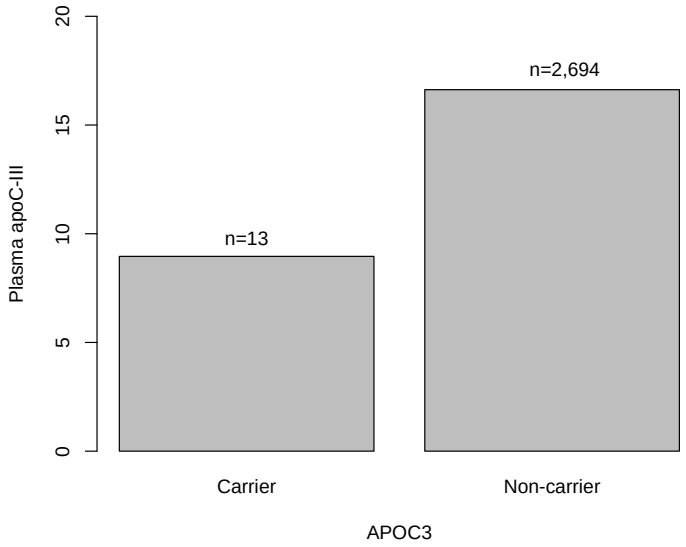


Table S1. Baseline characteristics of 3,734 participants sequenced across exome and with plasma triglycerides

Study	ARIC	CARDIA	CHS	MIGen	FHS	JHS	MESA	WHI
N	798	195	204	11	416	311	394	1405
Mean age ± SD, yrs	53.8 ± 5.8	26.5 ± 2.9	73.5 ± 5.8	52.4 ± 9.0	39.7 ± 9.9	54.0 ± 12.2	61.4 ± 9.7	63.3 ± 7.6
Gender, % female	53.1% (n=424)	43.1% (n=84)	25.5% (n=52)	81.8% (n=9)	34.1% (n=142)	55.3% (n=172)	37.8% (n=149)	100% (n=1405)
African-American Ancestry, %	37% (n=295)	44% (n=86)	31% (n=64)	100% (n=11)	0% (n=0)	100% (n=311)	38% (n=151)	52% (n=734)
Total cholesterol, mg/dl	226.0 ± 69.0	185.6 ± 49.1	207.0 ± 46.9	184.5 ± 52.2	197.3 ± 43.8	206.0 ± 53.4	193.5 ± 51.3	230.3 ± 46.9
Low-density lipoprotein cholesterol, mg/dl	148.0 ± 68.3	116.8 ± 45.6	128.7 ± 46.5	112.9 ± 47.6	125.1 ± 40.0	132.8 ± 51.4	118.4 ± 48.4	147.1 ± 45.0
High-density lipoprotein cholesterol, mg/dl	48.4 ± 17.3	52.0 ± 13.1	47.8 ± 14.5	43.8 ± 4.3	47.3 ± 13.7	49.3 ± 15.7	50.2 ± 15.0	54.1 ± 15.3
Triglycerides	150.5 ± 98.7	84.5 ± 71.3	156.4 ± 81.0	135.7 ± 62.7	129.8 ± 110.8	122.9 ± 83.1	125.8 ± 73.8	142.4 ± 84.2

For lipid traits, data shown are mean ± standard deviation; ARIC denotes Atherosclerosis Risk in Communities Study[11]; CARDIA, Coronary Artery Risk Development in Young Adults[12]; CHS, Cardiovascular Health Study[13]; MIGen, Myocardial Infarction Genetics Consortium[14]; FHS, Framingham Heart Study[15]; JHS, Jackson Heart Study[16]; MESA, Multi-Ethnic Study of Atherosclerosis[17]; WHI, Women's Health Initiative[18]

Table S2. Characteristics of participants in replication study of *APOC3* coding sequence variants with plasma lipid levels

Cohort	**ARIC**	**FHS**	**MDC-CVA**	**WHI**	**OHS Cases**	**OHS Controls**	**Procardis Cases**	**Procardis Controls**	**ATVB Cases**	**ATVB Controls**	**ARIC**	**JHS**	**WHI**
Ancestry	EA	EA	EA	EA	EA	EA	EA	EA	EA	EA	AA	AA	AA
N	10,349	7,033	4,924	4,157	800	2,111	1,070	1,776	1,252	960	2,933	2,154	2,152
Mean age ± SD, yrs	54.4 ± 5.7	37.7 ± 9.6	57.6 ± 5.9	66.9 ± 6.6	54.1 ± 9.3	74.7 ± 6.0	58.4 ± 7.6	67.0 ± 4.8	39.7 ± 4.9	39.3 ± 5.1	53.7 ± 5.8	52.9 ± 12.7	67.1 ± 5.2
Gender, % female	53%	53%	59%	100%	17%	49%	40%	49%	12%	14%	62%	63%	100%
Total cholesterol, mg/dl	214.5 ± 38.6	192.3 ± 37.8	239.5 ± 43.8	239.0 ± 44.1	238.6 ± 48.0	219.6 ± 40.2	234.2 ± 48.9	219.5 ± 38.8	221.3 ± 56.0	201.5 ± 37.2	213.5 ± 42.2	205.6 ± 42.7	233.6 ± 45.8
Low-density lipoprotein cholesterol, mg/dl	137.2 ± 35.5	118.9 ± 34.3	162.4 ± 40.2	155.4 ± 39.7	155.9 ± 39.6	137.9 ± 34.2	146.7 ± 44.4	132.5 ± 32.7	147.7 ± 52.3	125.7 ± 34.9	136.6 ± 39.5	133.8 ± 39.0	153.9 ± 42.7
High-density lipoprotein cholesterol, mg/dl	50.7 ± 16.7	53.0 ± 15.5	53.3 ± 14.4	51.6 ± 13.5	42.7 ± 13.9	56.7 ± 16.5	47.8 ± 13.3	55.5 ± 15.1	42.0 ± 13.0	49.1 ± 12.6	55.0 ± 17.4	51.7 ± 14.6	57.0 ± 14.5
Triglycerides	136.3 ± 90.6	102.7 ± 82.4	121.1 ± 69.9	161.1 ± 91.2	234.3 ± 203.1	126.7 ± 85.7	190.3 ± 120.5	143.0 ± 81.1	177.7 ± 132.4	121.7 ± 70.7	113.1 ± 84.3	103.2 ± 78.0	113.2 ± 69.6

For lipid traits, data shown are mean ± standard deviation; ARIC denotes Atherosclerosis Risk in Communities Study; FHS, Framingham Heart Study; MDC-CVA, Malmo Diet and Cancer Study Cardiovascular Arm; WHI, Women's Health Initiative; OHS, Ottawa Heart Study; PROCARDIS, Precocious Coronary Artery Disease Study; ATVB, Italian Atherosclerosis, Thrombosis, and Vascular Biology Study; JHS, Jackson Heart Study

Table S3. Definitions of coronary heart disease across fifteen studies

Study	Design	Definition of CHD	Ascertainment of controls	Refs
WHI	Prospective, cohort	WHI participants included in this study were 50-79 years of age at enrollment in 1993-1998. These women were followed for development of clinical CHD until 2012. A CHD event was defined as a definite or probable myocardial infarction, silent myocardial infarction, coronary revascularization, hospitalized angina, or death due to CHD.	Participants free of CHD on follow-up	18
FHS	Prospective, cohort	Incident nonfatal or fatal MI, angina pectoris, and coronary insufficiency	Participants free of CHD on follow-up	19
MDC-CVA	Prospective, cohort	Incident nonfatal or fatal MI	Participants free of CHD on follow-up	20
ARIC	Prospective, cohort	Incident definite or probable MI, silent MI (indicated by electrocardiogram) between 4 examinations in 1987–1998, definite CHD death, or coronary revascularization	Participants free of CHD on follow-up	11
IPM	Case-control	CAD cases were ascertained from Institute for Personalized Medicine Biobank; CAD was defined using the electronic health record. Cases were documented ICD9 codes 410.xx to 414.xx and (abnormal stress test or abnormal coronary angiography)	Controls were individuals in biobank who did not meet case criteria	NIH dbGaP Study Accession: phs000388.v1.p1
ATVB	Case-control	MI in men or women ≤ 45 yo	No history of thromboembolic disease	21
VHS	Case-control	Documented diagnosis of MI, coronary artery bypass grafting (CABG), CAD (by angiography) in males ≤ 50 yo for males and in females ≤ 60 yo	Coronary angiography normal	14
Ottawa	Case-control	Angiography (>1 coronary vessel with >50% stenosis); ≤ 50 yo for males and ≤ 60 yo for females; without type 2 diabetes	Asymptomatic, males >65, females >70	22
PROCARDIS	Case-control	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 (80%), coronary artery bypass graft (10%), acute coronary syndrome (6%), coronary angioplasty (1%) or stable angina (hospitalization for angina or documented obstructive coronary disease) (3%)	No personal or sibling history of CAD before age 66 years.	23
HUNT	Case-control	MI cases collected by the Norwegian Nord-Trøndelag health study (HUNT) Biobank	Free of MI on Norwegian ischemic heart disease national register	24
		The GoDARTS (Genetics of Diabetes Audit and Research in Tayside	Controls were free of coronary artery disease,	25

		Scotland) study is a joint initiative of the Department of Medicine and the Medicines Monitoring Unit (MEMO) at the University of Dundee, the diabetes units at three Tayside healthcare trusts (Ninewells Hospital and Medical School, Dundee; Perth Royal Infirmary; and Stracathro Hospital, Brechin), and a large group of Tayside general practitioners with an interest in diabetes care. Cases were a first-ever CAD event, defined as fatal and non-fatal myocardial infarction, unstable angina or coronary revascularisation	stroke and peripheral vascular disease	
GoDARTS CAD	Case-cohort			
EPIC CAD	Nested case-cohort	The EPIC (European Prospective Study into Cancer and Nutrition) study sub-cohorts from the UK were used, subjects were collected in collaboration with general practicioners, mainly in Cambridgeshire and Norfolk. Cases were individuals who developed a fatal or non-fatal CAD during an average follow-up of 11 years, until June 2006. Participants were identified if they had a hospital admission and/or died with CAD as the underlying cause. CAD was defined as cause of death codes ICD9 410-414 or ICD10 I20-I25, and hospital discharge codes ICD10 I20.0, I21, I22 or I23 according to the International Classification of Diseases, 9th and 10th revisions.	Controls were study participants who remained free of any cardiovascular disease during follow-up (defined as ICD9 401-448 and ICD10 I10-I79).	
FIA3	Nested case-control	Cases of MI occurring in participants from Västerbotten Intervention Program (VIP), WHO's Multinational Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA) study in northern Sweden and the Mammography Screening Project (MSP) in Västerbotten.	Individuals free of MI from VIP and MSP	[26,27]
German CAD	Case-control	**KORA-MI:** Hospitalized survivors of MI who are 26–74 years of age. The diagnosis of a MI (<60) was made with the use of the algorithm of the MONICA project. **PopGen CAD**: the PopGen CAD sample comprised unrelated German CAD patients with early onset of disease who were recruited in Schleswig–Holstein, Germany (www.PopGen.de). **Angio-Lüb:** the Lübeck angiographic study (Angio-Lüb) includes patients with angiographically proven CAD who underwent cardiac catheterization at the University Hospital Schleswig-Holstein, Campus Lübeck between 2005 and 2008. Patients were not selected for particular risk factors or phenotypes. **Munich-MI:** Participants of the Munich MI sample included in this study were consecutively recruited from 1993 to 2002 and examined with coronary angiography at Deutsches Herzzentrum München and 1. Medizinische Klinik rechts der Isar der Technischen Universität München. The diagnosis of MI was established in the presence of chest pain lasting >20 minutes combined with ST-segment elevation or pathological Q waves on a surface electrocardiogram. Patients with MI had to show either an angiographically occluded infarct-related artery or regional wall motion abnormalities corresponding to the electrocardiographic infarct localization, or both.	Controls were subjects from population-based studies from Germany (PopGen, Heinz-Nixdorf-Recall, KORA).	[28-32]

WTCCC	Case-control	CAD cases in the WTCCC Study were from those recruited in the British Heart Foundation Heart Family Heart Study (BHF-FHS) and supplemented by additional cases from WTCCC-CAD2	Controls were subjects from the UK 1958 Birth Cohort.	[33,34]

WHI, Women's Health Initiative; FHS, Framingham Heart Study; MDC-CVA, Malmo Diet and Cancer Study-Cardiovascular Arm; ARIC, Atherosclerosis Risk in Communities Study; IPM, Mt. Sinai Institute for Personalized Medicine Biobank; ATVB, Italian Atherosclerosis, Thrombosis, and Vascular Biology Study; Verona, Verona Heart Study; Ottawa, Ottawa Heart Study; PROCARDIS, Precocious Coronary Artery Disease Study; HUNT, Nord-Trøndelag health study; GoDARTS, Genetics of Diabetes Audit and Research Tayside; FIA3, First Myocardial Infarction in AC county 3; EPIC, European Prospective Study into Cancer and Nutrition; WTCCC, Wellcome Trust Case Control Consortium
MI denotes myocardial infarction; CAD, coronary artery disease

Table S4. Association of individual gene variants and plasma triglycerides in African Americans

Chromosome, position	Gene	N	Beta	Statistic	P	Minor allele frequency	Protein annotation
chr11_116662407	APOA5	1562	0.165	4.962	7.74E-07	0.07	S19W
chr6_153019197	MYCT1	1564	0.591	4.948	8.33E-07	0.005	T54A
chr16_5140548	FAM86A	1561	1.027	4.443	9.51E-06	0.001	T121A
chr12_106632875	CKAP4	1564	1.015	4.388	1.22E-05	0.001	G579D
chr2_113671410	IL37	1564	-0.078	-4.380	1.27E-05	0.33	T42A
chr1_183514098	SMG7	1564	0.824	4.373	1.31E-05	0.002	P632H
chr16_702524	WDR90	1476	0.146	4.357	1.41E-05	0.07	G371S
chr2_197298051	HECW2	1564	0.759	4.338	1.53E-05	0.002	A33T
chr22_37465121	TMPRSS6	1467	0.326	4.281	1.99E-05	0.02	R711L
chr1_240071937	CHRM3	1564	0.534	4.270	2.07E-05	0.005	L396M
chr19_53014422	ZNF578	1559	-0.130	-4.221	2.58E-05	0.08	I263T
chr19_54652192	CNOT3	1456	0.795	4.211	2.71E-05	0.002	G402S
chr15_89870432	POLG	1561	0.969	4.196	2.88E-05	0.001	A467T
chr6_90408618	MDN1	1564	0.945	4.094	4.46E-05	0.001	E3045G
chr22_29446079	ZNRF3	1353	1.93	4.088	4.61E-05	0.001	H637R

Covariates included age, age^2, sex, two principal components of ancestry, an indicator variable for race (in race-combined model only) and indicator variables for sequencing ascertainment scheme.

Table S5. Association of individual gene variants and plasma triglycerides in participants of European ancestry

Chromosome, position	Gene	N	Beta	Statistic	P	Minor allele frequency	Protein annotation
chr4_4304605	ZBTB49	2079	0.786	5.085	4.00E-07	0.003	A348T
chr6_39832264	DAAM2	2079	1.054	4.377	1.27E-05	0.001	R105H
chr3_62307648	C3orf14	2074	0.744	4.173	3.13E-05	0.002	L33M
chr11_116701560	APOC3	2075	-0.992	-4.120	3.94E-05	0.001	A43T
chr4_4322570	ZBTB49	2079	0.778	4.103	4.24E-05	0.002	E609K
chr8_121357700	COL14A1	2079	-0.949	-3.970	7.45E-05	0.001	P1659A
chr9_32633036	TAF1L	2079	-0.931	-3.889	0.0001	0.001	D848N
chr20_55941872	RAE1	2079	0.628	3.885	0.0001	0.003	P129S
chr3_49314251	C3orf62	2079	0.925	3.870	0.0001	0.001	R19G
chr4_84384688	FAM175A	2079	-0.725	-3.831	0.0001	0.002	R252Q
chr2_190608005	ANKAR	2079	-0.775	-3.823	0.0001	0.001	R1272H
chr12_10532326	KLRK1,KLRC4-KLRK1	2079	-0.077	-3.811	0.0001	0.21	T72A
chr12_57863433	GLI1	2079	-0.482	-3.797	0.0002	0.004	R382W
chr20_58476811	SYCP2	2066	-0.222	-3.780	0.0002	0.02	S363N
chr9_116132334	BSPRY	2079	0.0793	3.772	0.0002	0.19	T374I

Covariates included age, age^2, sex, two principal components of ancestry, an indicator variable for race (in race-combined model only) and indicator variables for sequencing ascertainment scheme.

Table S6. Association of individual gene variants and plasma triglycerides in participants of African American and European ancestry

Chromosome, position	Gene	N	Beta	Statistic	P	Minor allele frequency African Americans	Minor allele frequency European Americans	Protein annotation
chr11_116662407	APOA5	3728	0.124	5.092	3.71E-07	0.07	0.06	S19W
chr2_27730940	GCKR	3734	0.0686	4.984	6.50E-07	0.1	0.4	L446P
chr6_153019197	MYCT1	3731	0.586	4.470	8.05E-06	0.005	0	T54A
chr20_55941872	RAE1	3734	0.563	4.300	1.75E-05	0.001	0.003	P129S
chr3_62307648	C3orf14	3728	0.648	4.246	2.23E-05	0.0006	0.002	L33M
chr8_19819724	LPL	3734	-0.0891	-4.244	2.25E-05	0.07	0.1	S474X
chr17_38031648	ZPBP2	3734	-0.457	-4.202	2.71E-05	0.0003	0.005	K262E
chr12_57863433	GLI1	3734	-0.476	-4.081	4.57E-05	0.0003	0.004	R382W
chr5_102423628	GIN1	3416	-0.371	-4.051	5.22E-05	0.01	0.0002	N515D
chr4_84384688	FAM175A	3734	-0.724	-4.045	5.34E-05	0	0.002	R252Q
chr8_121292281	COL14A1	3734	0.702	3.918	9.10E-05	0.0006	0.001	A1197T
chr18_65181506	DSEL	3734	-0.698	-3.892	0.0001	0.0003	0.002	A124T
chr2_29259543	FAM179A	3734	-0.183	-3.888	0.0001	0.005	0.02	V852A
chr1_41978890	HIVEP3	3732	0.288	3.881	0.0001	0.01	0	R2001Q
chr11_19955322	NAV2	3724	-0.254	-3.873	0.0001	0.003	0.01	T447M

Covariates included age, age^2, sex, two principal components of ancestry, an indicator variable for race (in race-combined model only) and indicator variables for sequencing ascertainment scheme.

Table S7. Gene-based association results aggregating coding sequence variants with minor allele frequency < 1%

Ancestry	Gene	Full gene name	Location	P
EA	APOC3	apolipoprotein C-III	11q23.3	6.89E-06
EA	C12orf56	chromosome 12 open reading frame 56	12q14.2	8.63E-05
AA	GIN1	gypsy retrotransposon integrase 1	5q21.1	9.16E-05
AA	MARCH6	membrane-associated ring finger (C3HC4) 6, E3 ubiquitin protein ligase	5p15.2	3.80E-05
Combined	APOC3	apolipoprotein C-III	11q23.3	1.31E-05

Presented here are genes with $P < 0.0001$ for triglycerides in European ancestry, African American ancestry, and overall.

Table S8. Combined allele frequency of four rare *APOC3* loss-of-function mutations

Ancestry	*APOC3* R19X rs76353203 alternate allele count/total number of chromosomes	IVS2+1 G>A rs138326449 alternate allele count/total number of chromosomes	IVS3+1 G>T rs140621530 alternate allele count/total number of chromosomes	A43T rs147210663 alternate allele count/total number of chromosomes	Total alternate allele count/total number of chromosomes	Combined Allele frequency	Combined Carrier frequency
EA	3/8588	16/8586	1/8590	8/8592	28/8592	0.00326 (1:307)	1:154
AA	0/4402	3/4401	5/4400	7/4402	15/4402	0.00341 (1:293)	1:147

Table S9. Association of four *APOC3* coding sequence variants and plasma lipid levels

Mutation	R19X	R19X	IVS2+1 G>A	IVS2+1 G>A	A43T	A43T	IVS3+1 G>T	IVS3+1 G>T	Carriers of any of four *APOC3* mutations	Carriers of any of four *APOC3* mutations	Carriers of any of four *APOC3* mutations
Ancestry	EA	AA	EA	AA	EA	AA	EA	AA	EA	AA	Race-combined
N	33,068	5,066	14,623	2,152	24,840	7,282	10.618	7,279	34,432	7,239	41,671
TG*	-0.58	-0.38	-0.53	-0.76	-0.49	-0.13	-0.43	-0.63	-0.55	-0.38	-0.49
Beta (SE)	(0.09)	(0.21)	(0.07)	(0.14)	(0.13)	(0.08)	(0.37)	(0.16)	(0.05)	(0.06)	(0.04)
P TG	1.6e-11	0.07	7.2e-16	8.1e-08	2.0e-04	0.15	0.24	5.4e-05	<1.0e-20	1.4e-09	<1.0e-20
LDL-C*	-15.5	3.1	-8.6	-5.9	-4.1	13.4	13.1	2.4	-9.3	10.7	-3.8
Beta (SE)	(6.4)	(19.2)	(5.4)	(13.6)	(9.3)	(7.4)	(26.5)	(14.2)	(3.4)	(5.4)	(2.9)
P LDL-C	0.02	0.87	0.11	0.67	0.66	0.07	0.62	0.86	5.6e-03	0.05	0.19
HDL-C*	17.1	-1.2	9.0	7.4	4.5	6.3	13.5	25.5	11.5	9.1	10.8
Beta (SE)	(2.5)	(7.4)	(1.8)	(4.7)	(3.7)	(2.7)	(10.6)	(5.3)	(1.3)	(2.0)	(1.1)
P HDL-C	3.5e-12	0.87	6.7e-07	0.12	0.21	0.02	0.20	1.2e-06	<1.0e-20	7.5e-06	<1.0e-20

TG denotes triglycerides; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol

P values are for the comparison with noncarriers. P values were derived from a linear regression model, with adjustments for age, sex, ancestry, and principal components of ancestry. The P value for the triglyceride phenotype is based on triglyceride levels logarithmically transformed on a natural log scale.

*Units for TG are ln (triglycerides), for LDL-C is mg/dl, for HDL-C is mg/dl

Table S10: Association of *APOC3* LoF carrier status with plasma lipids before and after conditioning on *APOA5* S19W

	Whites		Blacks	
	Before accounting for *APOA5* S19W	**After accounting for *APOA5* S19W**	**Before accounting for *APOA5* S19W**	**After accounting for *APOA5* S19W**
Outcome variable	**Beta (SE) P**	**Beta (SE) P**	**Beta (SE) P**	**Beta (SE) P**
Triglycerides	-0.56 (0.12) 4×10^{-6}	-0.56 (0.12) 4×10^{-6}	-0.39 (0.10) 0.0001	-0.40 (0.10) 9×10^{-5}
HDL cholesterol	+12.0 (3.6) 0.001	+11.9 (3.6) 0.001	+12.5 (3.6) 0.0006	+12.6 (3.6) 0.0005
LDL cholesterol	-12.4 (8.5) 0.15	-12.4 (8.5) 0.15	+11.4 (8.4) 0.18	+11.2 (8.4) 0.18
n	10,349	10,349	2,932	2,932

Table S11. Association of rare *APOC3* mutations and risk for coronary heart disease

	Non-carriers	Carriers of *APOC3* R19X, IVS2+1 G>A, IVS3+1 G>T, or A43T	Proportion of cases who carry variant	Proportion of controls who carry variant
Study 1 - WHI				
EA Cases	2412	6	0.25%	0.63%
EA Controls	14009	89		
AA Cases	126	0	0%	0.62%
AA Controls	2249	14		
Study 2 - FHS				
EA Cases	126	0	0%	0.20%
EA Controls	3474	8		
Study 3 - MDC-CVA				
EA Cases	339	2	0.59%	0.35%
EA Controls	4507	16		
Study 4 - ARIC				
EA Cases	1792	2	0.11%	0.19%
EA Controls	8523	16		
AA Cases	556	8	1.4%	0.86%
AA Controls	3129	19		
Study 5 - IPM				
EA Cases	693	10	1.4%	1.91%
EA Controls	1696	33		
HA Cases	1053	2	0.19%	0.37%
HA Controls	3465	13		
AA Cases	553	3	0.54%	0.86%
AA Controls	3212	28		
Study 6 & 7 – ATVB + VHS				
EA Cases	1595	9	0.56%	1.3%
EA Controls	1217	16		
Study 8 - Ottawa				
EA Cases	1021	3	0.29%	0.84%
EA Controls	2248	19		
Study 9 - PROCARDIS				
EA Cases	2426	10	0.41%	0.73%
EA Controls	2163	16		
Study 10 - HUNT				
EA Cases	2891	6	0.21%	0.24%
EA Controls	2899	7		
Study 11 - GoDARTS CAD				
EA Cases	1694	0	0%	0.17%
EA Controls	2869	5		
Study 12 - EPIC CAD				
EA Cases	1394	2	0.14%	0.14%
EA Controls	7158	10		
Study 13 – FIA3				
EA Cases	2657	0	0%	0.38%
EA Controls	2112	8		
Study 14 - German				

CAD				
EA Cases	9681	37	0.38%	0.71%
EA Controls	5769	41		
Study 15 – WTCCC				
EA Cases	2880	13	0.45%	0.46%
EA Controls	5884	27		
Total Cases	33,889	113	0.33%	0.50%
Total Controls	76,583	385		

EA denotes European ancestry; AA, African American; HA, Hispanic ancestry; WHI, Women's Health Initiative; FHS, Framingham Heart Study; MDC-CVA, Malmo Diet and Cancer Study-Cardiovascular Arm; ARIC, Atherosclerosis Risk in Communities Study; IPM, Mt. Sinai Institute for Personalized Medicine Biobank; ATVB, Italian Atherosclerosis, Thrombosis, and Vascular Biology Study; Verona, Verona Heart Study; Ottawa, Ottawa Heart Study; PROCARDIS, Precocious Coronary Artery Disease Study; HUNT, Nord-Trøndelag health study; GoDARTS, Genetics of Diabetes Audit and Research Tayside; EPIC, European Prospective Study into Cancer and Nutrition; FIA3, FörstagångsInsjuknande i hjärtinfarkt i AC-län; WTCCC, Wellcome Trust Case Control Consortium

Table S12. Number of individuals expected to be homozygous or compound heterozygous for any of four *APOC3* loss-of-function mutations

N genotyped	Combined allele frequency = q	Expected number of homozygotes = q^2*n	Variance in number of homozygotes = $q^2(1-q^2)*n$	Standard deviation = square root of variance
110,970	1:300	1.23	1.23	1.11

Table S13. Association of *APOC3* LoF mutations with CT hepatic fat in 3,051 Framingham Heart Study participants

Outcome variable	Predictor variable	Covariates	Beta (SE)	P
CT hepatic fat	APOC3 R19X or IVS2+1 G>A (n=27)	age, age^2, gender	-0.04 (0.19)	0.82
CT hepatic fat	APOC3 R19X or IVS2+1 (n=27)	age, age^2, gender, # of alcoholic drinks per week	-0.04 (0.19)	0.84

Table S14. Correlation of plasma apolipoprotein C-III level with plasma lipids, apolipoproteins, and cardiovascular risk factors in the Framingham Heart Study Offspring Cohort

Variable	Correlation Cofficient	Pr(>|t|)
Total cholesterol	0.473	2.26E-180
High density lipoprotein cholesterol	-0.135	1.25E-14
Low-density lipoprotein cholesterol	0.233	5.01E-40
Triglycerides	0.752	<1E-222
Log (Triglycerides)	0.789	<1E-222
Body mass index	0.200	1.42E-30
Fasting glucose	0.271	2.31E-55
Intermediate-density lipoprotein determined by NMR, Exam 4	0.274	4.92E-48
VLDL size determined by NMR, Exam 4	0.390	6.15E-100
LDL size determined by NMR, Exam 4	-0.327	3.20E-69
HDL size determined by NMR, Exam 4	-0.178	6.77E-21
Large VLDL particles determined by NMR, Exam 4	0.445	1.34E-132
Medium VLDL particles determined by NMR, Exam 4	0.557	1.21E-222
Small VLDL particles determined by NMR, Exam 4	0.135	1.41E-12
Large LDL particles determined by NMR, Exam 4	-0.114	2.47E-09
Medium LDL particles determined by NMR, Exam 4	0.249	9.21E-40
Small LDL particles determined by NMR, Exam 4	0.329	1.02E-69
large HDL particles determined by NMR, Exam 4	-0.135	1.60E-12
medium HDL particles determined by NMR, Exam 4	0.209	3.14E-28
small HDL particles determined by NMR, Exam 4	0.094	9.33E-07
Apolipoprotein AI concentration by ELISA (mg/dl), Exam 4	0.132	6.79E-14
Apolipoprotein AII concentration by ELISA (mg/dl), Exam 4	0.294	4.02E-65
Apolipoprotein B concentration by ELISA (mg/dl), Exam 4	0.359	4.26E-98
Cholesterol in remnant like particles in mg/dl, Exam 4	0.421	2.12E-106
Triglycerides in remnant like particles in mg/dl, Exam 4	0.365	5.35E-76
Systolic blood pressure	0.249	7.63E-47
Diastolic blood pressure	0.172	7.17E-23
Log (C-reactive protein), exam 5	0.174	9.53E-30
Log (C-reactive protein), exam 6	0.193	1.90E-24
Sex	0.019	0.276
Age	0.200	2.05E-30

Correlations are unadjusted. All measurements are made in exam cycle 5 unless specified.
VLDL denotes very-low density lipoprotein; NMR, nuclear magnetic resonance; LDL, low-density lipoprotein; HDL, high-density lipoprotein

Table S15. Association of continuous plasma apolipoprotein C-III levels with incident CHD in the Framingham Heart Study

Model	Beta	SE	OR	95% CI lower	95% CI upper	P
1	0.044	0.011	1.045	1.023	1.068	4.90E-05
2	0.017	0.015	1.017	0.988	1.047	0.26

Model 1 covariates include age and sex
Model 2 covariates include age, sex, smoking, diabetes mellitus, LDL cholesterol, HDL cholesterol, hypertension treatment, alcohol consumption, systolic and diastolic blood pressure, lipid-lowering treatment, and fasting serum glucose

Table S16. Association of tertiles of plasma apolipoprotein C-III levels with incident events in the Framingham Heart Study Offspring cohort

Model	Comparison	Beta	SE	OR	95% CI lower	95% CI upper	P
1	Lowest third vs. Highest third	-0.435	0.148	0.648	0.484	0.865	0.003
	Middle third vs. Highest third	-0.174	0.133	0.841	0.647	1.091	0.19
2	Lowest third vs. Highest third	-0.117	0.159	0.890	0.651	1.214	0.46
	Middle third vs. Highest third	0.029	0.141	1.029	0.780	1.357	0.84

Model 1: age and sex
Model 2: age, sex, smoking, diabetes mellitus, LDL cholesterol, HDL cholesterol, hypertension treatment, alcohol consumption, systolic and diastolic blood pressure, lipid-lowering treatment, and fasting serum glucose

Table S17. Association of continuous plasma apolipoproteinC-III levels with total mortality in CAD patients in Verona Heart Study

Model	Beta	SE	OR	95% CI lower	95% CI upper	P
1	0.078	0.016	1.081	1.047	1.116	2E-06
2	0.107	0.025	1.113	1.059	1.168	2E-05

Model 1 covariates include age and sex
Model 2 covariates include age, sex, diabetes mellitus, hypertension, LDL cholesterol, HDL cholesterol, lipid-lowering treatment, and fasting serum glucose

Table S18. Association of tertiles of plasma apolipoproteinC-III levels with total mortality in CAD patients in Verona Heart Study

Model	Comparison	Beta	SE	OR	95% CI lower	95% CI upper	P
1	Lowest third vs. Highest third	-0.677	0.211	0.508	0.336	0.769	0.001
	Middle third vs. Highest third	-0.565	0.215	0.569	0.373	0.867	0.009
2	Lowest third vs. Highest third	-0.774	0.317	0.461	0.248	0.858	0.015
	Middle third vs. Highest third	-0.819	0.301	0.441	0.244	0.795	0.006

Model 1: age and sex

Model 2 covariates include age, sex, diabetes mellitus, hypertension, LDL cholesterol, HDL cholesterol, lipid-lowering treatment, and fasting serum glucose

Table S19. Association of continuous plasma apolipoproteinC-III levels with cardiovascular mortality in CAD patients in Verona Heart Study

Model	Beta	SE	OR	95% CI lower	95% CI upper	P
1	0.069	0.020	1.071	1.029	1.115	0.001
2	0.088	0.033	1.092	1.023	1.165	0.008

Model 1 covariates include age and sex
Model 2 covariates include age, sex, diabetes mellitus, hypertension, LDL cholesterol, HDL cholesterol, lipid-lowering treatment, and fasting serum glucose

Table S20. Association of tertiles of plasma apolipoproteinC-III levels with cardiovascular mortality in CAD patients in Verona Heart Study

Model	Comparison	Beta	SE	OR	95% CI lower	95% CI upper	P
1	Lowest third vs. Highest third	-0.673	0.268	0.510	0.302	0.862	0.012
	Middle third vs. Highest third	-0.246	0.248	0.782	0.481	1.271	0.321
2	Lowest third vs. Highest third	-0.850	0.399	0.427	0.195	0.934	0.033
	Middle third vs. Highest third	-0.523	0.341	0.593	0.304	1.156	0.125

Model 1: age and sex
Model 2 covariates include age, sex, diabetes mellitus, hypertension, LDL cholesterol, HDL cholesterol, lipid-lowering treatment, and fasting serum glucose

References

1. Yu L, Martinez FD, Klimecki WT. Automated high-throughput sex-typing assay. Biotechniques 2004;37:662-4.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
3. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.
4. Fisher S, Barry A, Abreu J, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 2011;12:R1.
5. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
6. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 2008;18:1851-8.
7. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007;81:3.
8. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867-73.
9. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:887-93.
10. Goldstein JI, Crenshaw A, Carey J, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics 2012;28:2543-5.
11. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687-702.
12. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 1988;41:1105-16.
13. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991;1:263-76.
14. Myocardial Infarction Genetics Consortium, Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334-41.
15. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951;41:279-81.
16. 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:S6-4-17.
17. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871-81.
18. 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.
19. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979;110:281-90.
20. Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med 2008;358:1240-9.

21. Atherosclerosis, Thrombosis, and Vascular Biology Working Group. No evidence of association between prothrombotic gene polymorphisms and the development of acute myocardial infarction at a young age. Circulation 2003;107:1117-22.
22. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007;316:1488-91.
23. Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med 2009;361:2518-28.
24. Krokstad S, Langhammer A, Hveem K, et al. Cohort Profile: The HUNT Study, Norway. Int J Epidemiol 2012;42:968-977.
25. Morris AD, Boyle DI, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. BMJ 1997;315:524-8.
26. Norberg M, Blomstedt Y, Lonnberg G, et al. Community participation and sustainability--evidence over 25 years in the Vasterbotten Intervention Programme. Global Health Action 2012;5:1-9.
27. Stegmayr B, Lundberg V, Asplund K. The events registration and survey procedures in the Northern Sweden MONICA Project. Scandinavian Journal of Public Health Supplement 2003;61:9-17.
28. Lowel H, Meisinger C, Heier M, Hormann A. The population-based acute myocardial infarction (AMI) registry of the MONICA/KORA study region of Augsburg. Gesundheitswesen 2005;67 Suppl 1:S31-7.
29. Linsel-Nitschke P, Jansen H, Aherrarhou Z, et al. Macrophage cholesterol efflux correlates with lipoprotein subclass distribution and risk of obstructive coronary artery disease in patients undergoing coronary angiography. Lipids in Health and Disease 2009;8:14.
30. Erdmann J, Willenborg C, Nahrstaedt J, et al. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. Eur Heart J 2011;32:158-68.
31. Rivera NV, Carreras-Torres R, Roncarati R, et al. Assessment of the 9p21.3 locus in severity of coronary artery disease in the presence and absence of type 2 diabetes. BMC Medical Genetics 2013;14:11.
32. Erbel R, Lehmann N, Churzidse S, et al. Gender-specific association of coronary artery calcium and lipoprotein parameters: The Heinz Nixdorf Recall Study. Atherosclerosis 2013;229:531-540.
33. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443-53.
34. CARDIoGRAMPlusC4D Consortium, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013;45:25-33.