ORIGINAL ARTICLE

Inactivating Mutations in *NPC1L1* and Protection from Coronary Heart Disease

The Myocardial Infarction Genetics Consortium Investigators

ABSTRACT

BACKGROUND

Ezetimibe lowers plasma levels of low-density lipoprotein (LDL) cholesterol by inhibiting the activity of the Niemann–Pick C1-like 1 (NPC1L1) protein. However, whether such inhibition reduces the risk of coronary heart disease is not known. Human mutations that inactivate a gene encoding a drug target can mimic the action of an inhibitory drug and thus can be used to infer potential effects of that drug.

METHODS

We sequenced the exons of *NPC1L1* in 7364 patients with coronary heart disease and in 14,728 controls without such disease who were of European, African, or South Asian ancestry. We identified carriers of inactivating mutations (nonsense, splice-site, or frameshift mutations). In addition, we genotyped a specific inactivating mutation (p.Arg406X) in 22,590 patients with coronary heart disease and in 68,412 controls. We tested the association between the presence of an inactivating mutation and both plasma lipid levels and the risk of coronary heart disease.

RESULTS

With sequencing, we identified 15 distinct *NPC1L1* inactivating mutations; approximately 1 in every 650 persons was a heterozygous carrier for 1 of these mutations. Heterozygous carriers of *NPC1L1* inactivating mutations had a mean LDL cholesterol level that was 12 mg per deciliter (0.31 mmol per liter) lower than that in noncarriers ($P=0.04$). Carrier status was associated with a relative reduction of 53% in the risk of coronary heart disease (odds ratio for carriers, 0.47; 95% confidence interval, 0.25 to 0.87; $P=0.008$). In total, only 11 of 29,954 patients with coronary heart disease had an inactivating mutation (carrier frequency, 0.04%) in contrast to 71 of 83,140 controls (carrier frequency, 0.09%).

CONCLUSIONS

Naturally occurring mutations that disrupt NPC1L1 function were found to be associated with reduced plasma LDL cholesterol levels and a reduced risk of coronary heart disease. (Funded by the National Institutes of Health and others.)

The authors are listed in the Appendix. 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.

This article was published on November 12, 2014, at NEJM.org.

DOI: 10.1056/NEJMoa1405386

Ezetimibe, a drug that is commonly prescribed to reduce plasma levels of low-density lipoprotein (LDL) cholesterol, inhibits the function of the protein encoded by the Niemann–Pick C1-like 1 gene (*NPC1L1*).[1] NPC1L1 protein, which is expressed in the small intestine and liver, functions as a transporter of dietary cholesterol from the gut lumen into intestinal enterocytes.[2,3] Because of its ability to block sterol absorption by about 50%,[4] ezetimibe lowers plasma LDL cholesterol levels by 15 to 20%.[5] However, it is uncertain whether inhibiting NPC1L1 — either through ezetimibe treatment or by other means — reduces the risk of clinical coronary heart disease.[6]

Naturally occurring DNA sequence variants in humans that affect the activity of one or more protein targets can be used to estimate the potential efficacy and toxicity of a drug targeting such proteins.[7,8] Genomewide association studies have identified common DNA sequence variants in *NPC1L1* associated with modest alterations in plasma LDL cholesterol levels.[9] However, it is difficult to discern precisely how variants that are discovered through genomewide association studies affect the activity of a gene.

In contrast, some DNA mutations that arise in the protein-coding sequence can completely inactivate a gene. Inactivating mutations can be single-base changes that introduce a stop codon and that lead to premature truncation of a protein (nonsense mutations), insertions or deletions (indels) of DNA that scramble the protein translation beyond the variant site (frameshift mutations), or point mutations at modification sites of the nascent pre–messenger RNA transcript that alter the splicing process[10] (splice-site mutations). Because such mutations — which are variously termed protein-disruptive, protein-inactivating, loss-of-function, or null — profoundly affect protein function, they are typically very rare in the population as a consequence of natural selection.

We tested the hypothesis that protein-inactivating mutations in *NPC1L1* reduce both the LDL cholesterol level and the risk of coronary heart disease. We sequenced the coding regions of *NPC1L1* in a large number of persons, identified carriers of mutations that inactivate this gene, and determined whether persons who carry a heterozygous inactivating mutation had a lower LDL cholesterol level and a lower risk of coronary heart disease than noncarriers of these mutations.

Methods

Study Design

We conducted this study using data and DNA samples from 16 case–control studies and cohort studies. All study participants provided written informed consent for genetic studies. The first and last authors designed the study. The institutional review boards at the Broad Institute and each participating site approved the study protocols. The first and last authors vouch for the accuracy and completeness of the data and all analyses.

Study Participants

During the first phase of the study, we sequenced the 20 protein-coding exons in *NPC1L1* in samples obtained from 22,092 participants from seven case–control studies and two prospective cohort studies (see Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). The case–control studies included the Exome Sequencing Project Early-Onset Myocardial Infarction (ESP-EOMI) study conducted by the National Heart, Lung, and Blood Institute,[11] the Italian Atherosclerosis Thrombosis and Vascular Biology (ATVB) study,[12] the Ottawa Heart Study (OHS),[13] the Precocious Coronary Artery Disease (PROCARDIS) study,[14] the Pakistan Risk of Myocardial Infarction Study (PROMIS),[15] the Registre Gironi del COR (Gerona Heart Registry or REGICOR) study,[16] and the Munich Myocardial Infarction (Munich-MI) study.[17] The prospective cohort studies included the Atherosclerosis Risk in Communities (ARIC) study[18] and the Jackson Heart Study (JHS).[19]

During the second phase of the study, we genotyped the most common inactivating mutation in *NPC1L1* on the basis of data obtained during the sequencing phase (p.Arg406X) in nine independent sample sets from a total of 91,002 participants (Table S2 in the Supplementary Appendix). These nine sample sets were from participants in the ARIC study (participants who did not undergo sequencing), the Vanderbilt University Medical Center Biorepository (BioVU),[20] the Genetics of Diabetes Audit and Research Tayside (GoDARTS) study,[21] the German North and German South Coronary Artery Disease studies,[22] the Mayo Vascular Diseases Biorepository (Mayo),[23] PROCARDIS (participants who did not undergo sequencing), the Women's Genome Health Study (WGHS),[24] and the Women's Health Initiative (WHI).[25]

CLINICAL DATA

Data obtained for all the participants from both the sequencing and genotyping phases of the study included a medical history and laboratory assessment for cardiovascular risk factors, as described previously for each study. The participants were of African ancestry (2836 participants from ARIC, 2251 from JHS, and 455 from ESP-EOMI), South Asian ancestry (1951 participants from PROMIS), or European ancestry (all the other participants).

For each study cohort, available clinical data were used to define coronary heart disease. The definitions, which therefore varied from cohort to cohort, are provided in Tables S1 and S2 in the Supplementary Appendix.

SEQUENCING AND GENOTYPING

Sequence data for *NPC1L1* were extracted from exome sequences generated at the Broad Institute, the Human Genome Sequencing Center at Baylor College of Medicine, or the University of Washington with the use of protocols that are described in the Supplementary Appendix. Briefly, sequence reads were aligned to the human reference genome (build HG19), and the basic alignment files for sequenced samples were combined for the purpose of identifying variant positions. Single-nucleotide variants (SNVs) and indels were identified, and quality control procedures were applied to remove outlier samples and outlier variants, as described in the Supplementary Appendix.

For the purposes of this study, we defined inactivating mutations as any one of the following: SNVs leading to a stop codon substitution (nonsense mutations), SNVs occurring within two base pairs of an exon–intron boundary (splice-site mutations), or DNA insertions or deletions leading to a change in the reading frame and the introduction of a premature stop codon (frameshift mutations). The positions of nonsense, splice-site, and frameshift mutations were based on the complementary DNA reference sequence for *NPC1L1* (NM_013389.2) with the ATG initiation codon, encoding methionine, numbered as residue 1 or p.Met1.

To obtain additional data for a particular nonsense mutation (p.Arg406X) observed from sequencing *NPC1L1*, we genotyped the variant site in additional samples using the HumanExome BeadChip Kit (Illumina), according to the manufacturer's recommended protocol. (See the Methods section in the Supplementary Appendix for details.)

TECHNICAL VALIDATION OF SEQUENCING AND GENOTYPING

To assess the accuracy of next-generation sequencing methods, we performed Sanger sequencing on samples obtained from all participants who carried inactivating mutations in the ATVB study. To assess the accuracy of the genotyping of *NPC1L1* p.Arg406X with the HumanExome BeadChip kit, we compared these genotypes with those derived from next-generation sequencing for a subset of samples.

STATISTICAL ANALYSIS

We first tested the association between *NPC1L1* protein-inactivating mutations and plasma lipid levels. For participants who were receiving lipid-lowering therapy, we accounted for an average reduction in total cholesterol and LDL cholesterol levels of 20% and 30%, respectively,[26] by adjusting the measured values accordingly. We did not adjust levels of high-density lipoprotein (HDL) cholesterol or triglycerides in these participants. Status with respect to the use of lipid-lowering medication was available for participants in ARIC, JHS, Munich-MI, PROCARDIS, REGICOR, and WGHS. When possible, we combined primary data for studies that included only one participant with an inactivating mutation with data for other studies involving participants of the same ancestry in order to create a larger data set. We performed regression analysis with a linear model that was adjusted for age and sex, along with an indicator variable for the study if applicable, to test for an association between the presence of inactivating mutations in *NPC1L1* and levels of total cholesterol, LDL cholesterol, HDL cholesterol, and log-transformed triglyceride levels in each sample set. We combined results first within ancestry groups and then across ancestry groups, using fixed-effects meta-analyses.

We next tested for an association between protein-inactivating mutations in *NPC1L1* and the risk of coronary heart disease. In each study, we estimated the odds ratio for disease among carriers of any *NPC1L1* inactivating mutation, as compared with noncarriers. We then calculated the summary odds ratios and 95% confidence intervals for coronary heart disease among carriers, using a Mantel–Haenszel fixed-effects meta-analysis without continuity correction, a method that is robust with low (and even zero) counts and resultant odds ratios. A P value of less than 0.05 was considered to indicate statis-

tical significance. The R software program (R Project for Statistical Computing) was used for all analyses.

RESULTS

RARE INACTIVATING MUTATIONS IN *NPC1L1*

After sequencing *NPC1L1* in 7364 patients with coronary heart disease and in 14,728 controls without such disease, we identified 15 mutations that were expected to inactivate NPC1L1 (Table 1). These mutations included 10 nonsense single-nucleotide substitutions, 3 single-nucleotide substitutions that were predicted to disrupt splicing, and 2 frameshift indels (Fig. 1). In aggregate, these 15 mutations were seen in a total of 34 participants with heterozygous mutations; no homozygotes or compound heterozygotes were identified. *NPC1L1* inactivating mutations were rare, with such variants found in approximately 1 in 650 participants.

The most frequently observed individual mutation was p.Arg406X, which had a minor allele frequency of 0.02% among participants of European ancestry (seven alleles observed in 29,198 chromosomes) and was not observed in participants of African or South Asian ancestry. We genotyped this single variant in an additional 22,590 participants with coronary heart disease and in 68,412 controls. Among these 91,002 participants, we identified 48 additional heterozygous carriers (Table 1). The baseline characteristics of participants carrying *NPC1L1* inactivating mutations and those without such mutations were similar across all 16 studies (Table S3 in the Supplementary Appendix).

As a quality-control measure to assess the accuracy of next-generation sequencing, we performed Sanger sequencing and independently confirmed the presence of inactivating mutations in all carriers who were identified in the ATVB study. (See the Supplementary Appendix for details.) In a similar effort to assess the quality of genotyping, we compared genotype calls for p.Arg406X across 4092 samples that had undergone both genotyping and sequencing. On the basis of these data, we observed 100% specificity and sensitivity in identifying p.Arg406X carriers with the use of genotyping.

NPC1L1 MUTATIONS AND PLASMA LIPID LEVELS

Plasma lipid measurements were available for 13,626 participants in the ARIC study, 2082 in the JHS, and 22,515 in the WGHS. In addition, plasma lipid levels were available for 5405 controls without coronary heart disease from the ATVB, ESP-EOMI, Munich-MI, OHS, PROCARDIS, and REGICOR studies. LDL cholesterol levels were available for 42,813 of these 43,628 study participants. To minimize the effect of ascertainment bias, we excluded patients with coronary heart disease from case–control studies in the lipids analysis. As compared with noncarriers, carriers of inactivating *NPC1L1* mutations had significantly lower levels of total cholesterol (mean adjusted difference, –13 mg per deciliter [0.34 mmol per liter]; P=0.03) (Table 2) and LDL cholesterol (mean adjusted difference, –12 mg per deciliter [0.31 mmol per liter]; P=0.04) (Table 2 and Fig. 2). Triglyceride levels were also reduced among carriers, although the difference was not significant (mean change, –12%; P=0.11). We did not observe any significant difference in HDL cholesterol levels between carriers and noncarriers, with an increase of 2 mg per deciliter (0.05 mmol per liter) among carriers (P=0.29). Participants of European and African ancestry had a similar magnitude of LDL reduction (–13 mg per deciliter and –10 mg per deciliter [0.26 mmol per liter], respectively) (Fig. 2).

ASSOCIATION BETWEEN *NPC1L1* MUTATIONS AND CORONARY RISK

Carriers of the 15 inactivating mutations that we identified in *NPC1L1* were underrepresented among patients with coronary heart disease, as compared with controls (Table 3). In total, only 11 participants among 29,954 patients with coronary heart disease had an inactivating mutation (carrier frequency, 0.04%) in contrast to 71 of 83,140 controls (carrier frequency, 0.09%). This represented a 53% reduction in the risk of coronary heart disease among carriers of inactivating *NPC1L1* mutations (odds ratio for disease among carriers, 0.47; 95% confidence interval [CI], 0.25 to 0.87; P=0.008) (Table 3, and Table S4 in the Supplementary Appendix).

We observed a reduced risk of coronary heart disease among both participants of African descent and those of European descent. In the African ancestry subgroup, only 2 of 887 patients with coronary heart disease carried an *NPC1L1* inactivating mutation (carrier frequency, 0.23%), as compared with 13 of 4655 controls (carrier frequency, 0.28%), representing a 17%

Table 1. Inactivating Mutations in *NPC1L1* in Participants from 16 Studies. *

Cohort and Mutation	Type of Mutation	Study (No. of Participants/ No. of Carriers)	Ancestry
Sequencing and genotyping cohorts		All studies (113,094/82)	
Sequencing cohort			
All participants†		All studies in sequencing cohort (22,092/34)	
p.L71RfsX50	Frameshift	ARIC (2836/1), JHS (2251/1)	African
p.Q167X	Nonsense	OHS (1953/1)	European
p.A296VfsX57	Frameshift	ATVB (3539/3)	European
p.R406X	Nonsense	ATVB (3539/4), PROCARDIS (1902/1), ARIC (5718/2)	European
p.Y483X	Nonsense	PROMIS (1951/1)	South Asian
c.1681+1G→A	Splice-site	ARIC (2836/2), JHS (2251/1)	African
p.W592X	Nonsense	ARIC (5718/1)	European
p.R601X	Nonsense	ARIC (2836/1)	African
p.Q604X	Nonsense	ESP-EOMI (455/1), ARIC (2836/2)	African
p.R738X	Nonsense	REGICOR (783/2)	European
p.E803X	Nonsense	ARIC (5718/1)	European
c.2637+2T→G	Splice-site	ARIC (5718/1), Munich-MI (704/1)	European
p.C967X	Nonsense	ARIC (2836/1)	African
p.A1201V	Splice-site	JHS (2251/2)	African
p.R1325X	Nonsense	ARIC (2836 of African ancestry/1), JHS (2251 of African ancestry/2), ARIC (5718 of European ancestry/1)	African and European
Genotyping cohort			
All participants‡		All studies in genotyping cohort (91,002/48)	
p.R406X	Nonsense	ARIC (5237/4), BioVU (21,143/12), German North (7350/1), German South (8176/3), GoDARTS (3765/4), Mayo (2669/2), PROCARDIS (2227/1), WGHS (22,617/11), WHI (17,818/10)	European

* ARIC denotes Atherosclerosis Risk in Communities, ATVB Atherosclerosis Thrombosis and Vascular Biology, BioVU Vanderbilt University Medical Center Biorepository, ESP-EOMI Exome Sequencing Project Early-Onset Myocardial Infarction, German North German North Coronary Artery Disease Study, German South German South Coronary Artery Disease Study, GoDARTS Genetics of Diabetes Audit and Research Tayside, JHS Jackson Heart Study, Mayo Mayo Vascular Diseases Biorepository, Munich-MI Munich Myocardial Infarction, OHS Ottawa Heart Study, PROCARDIS Precocious Coronary Artery Disease, PROMIS Pakistan Risk of Myocardial Infarction Study, REGICOR Registre Gironi del COR, WGHS Women's Genome Health Study, and WHI Women's Health Initiative.

† The number of participants in the sequencing cohort includes 7364 patients with coronary heart disease and 14,728 controls without such disease.

‡ The number of participants in the genotyping cohort includes 22,590 patients with coronary heart disease and 68,412 controls without such disease.

reduction in the risk of coronary heart disease among carriers (cohort-based meta-analysis odds ratio, 0.83). In participants of European ancestry, 9 of 28,223 patients with coronary heart disease carried the mutation (carrier frequency, 0.03%), as compared with 57 of 77,378 controls (carrier frequency, 0.07%), representing a 57% reduction in the risk of coronary heart disease among carriers (cohort-based meta-analysis odds ratio, 0.43).

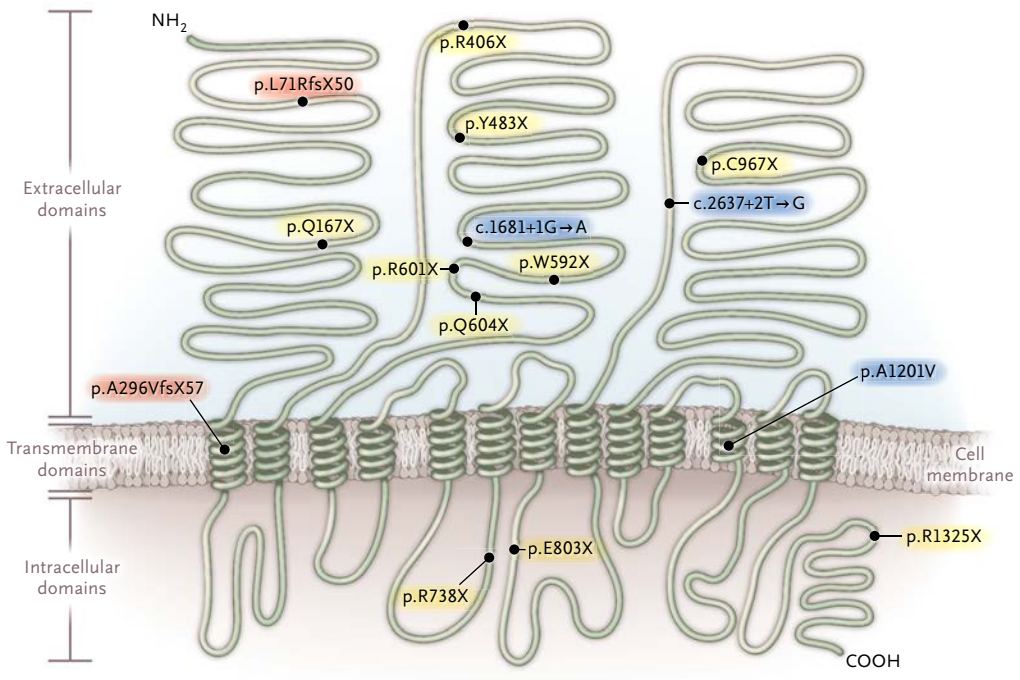


Figure 1. Inactivating Mutations in *NPC1L1* Identified in the Study.

Black circles indicate individual mutations along with the effect expected to lead to NPC1L1 inactivation. Mutations p.L71RfsX50 and p.A296VfsX57 (red shading) are indels that shift the open reading frame and induce a premature termination codon after an additional 50 and 57 amino acids, respectively. Mutations c.1681+1G→A, c.2637+2T→G, and p.A1201V (c.3602C→T) (blue shading) alter the splicing process at sites of modification of the nascent pre–messenger RNA transcript (splice-site mutations). All other mutations (yellow shading) are single-nucleotide variants that introduce a termination codon. The locations of the three main extracellular domains, 13 transmembrane domains, and intracellular domains are based on data from Betters and Yu.[27] NH_2 denotes the N-terminal at which protein translation is initiated, and COOH the C-terminal at which translation terminates.

DISCUSSION

We sequenced the protein-coding regions of *NPC1L1* in 22,092 participants and identified 15 rare mutations that were expected to disrupt the protein. We also genotyped the most frequently observed of these inactivating mutations (p.Arg406X) in an additional 91,002 participants. Carriers of any *NPC1L1* inactivating mutation had a mean LDL cholesterol level that was 12 mg per deciliter lower than the level in noncarriers, along with a 53% lower risk of coronary heart disease. These results show that lifelong inactivation of one copy of *NPC1L1* is protective against coronary heart disease.

The observation that genetic inhibition of NPC1L1 reduces the risk of coronary heart disease increases the prior probability that pharmacologic inhibition of NPC1L1 will also reduce the risk of disease. In 2002, ezetimibe was initially approved as a therapeutic agent in the United States on the basis of the capacity of the drug to lower LDL cholesterol levels. Although it has been assumed that any pharmacologic means of lowering LDL cholesterol levels will reduce the risk of coronary heart disease, the findings from the Ezetimibe and Simvastatin in Hypercholesterolemia Enhances Atherosclerosis Regression (ENHANCE) trial have led some observers to question this assumption.[28] In ENHANCE, the addition of ezetimibe to background statin therapy in patients with familial hypercholesterolemia did not reduce the progression of carotid intima–media thickness, a surrogate measure for atherosclerosis.[6] In the ongoing phase 3, randomized Im-

proved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT; ClinicalTrials.gov number, NCT00202878), investigators are evaluating whether the addition of ezetimibe to background simvastatin therapy will reduce the risk of recurrent cardiovascular events in patients with a recent acute coronary syndrome.[29]

Our findings do not predict with certainty that ezetimibe will be found to reduce cardiovascular risk in the IMPROVE-IT trial or other clinical studies, for several reasons. First, lifelong genetic inhibition, as tested in our study, has important differences from pharmacologic inhibition that is initiated in adulthood and lasts for several years. Second, our genetic study focuses on a first cardiovascular event, whereas IMPROVE-IT is evaluating recurrent events. Finally, the net clinical benefit of a pharmacologic therapy is a complex interplay among multiple factors, including many that are specific to the drug (e.g., toxic effects) and that would not be tested in a genetic model such as the one used in our study.

The reduction in the risk of coronary heart disease (53%) that we observed among carriers exceeds the reduction that would be expected for a decrease of 12 mg per deciliter in LDL cholesterol on the basis of results from statin trials.[26] Several factors may explain this difference. Modest reductions in plasma lipid levels over a lifetime, as achieved in carriers of an inactivating mutation, appear to lead to a larger modification of the risk

Table 2. Association between the Presence of Inactivating Mutations in *NPC1L1* and Plasma Lipid Levels.*

Variable	Mean Difference between Carriers and Noncarriers*	P Value
Cholesterol (mg/dl)		
Total	−13	0.03
Low-density lipoprotein	−12	0.04
High-density lipoprotein	2	0.29
Triglycerides (% change)	−12	0.11†

* The mean difference is the summary effect estimate for carriers of inactivating mutations in *NPC1L1*, as compared with noncarriers, after adjustment for age, sex, and study. Participants from population-based studies (ARIC, JHS, and WGHS) and controls without coronary heart disease from case–control studies were included in this analysis. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586.

† This P value was calculated with the use of natural log transformation of the values.

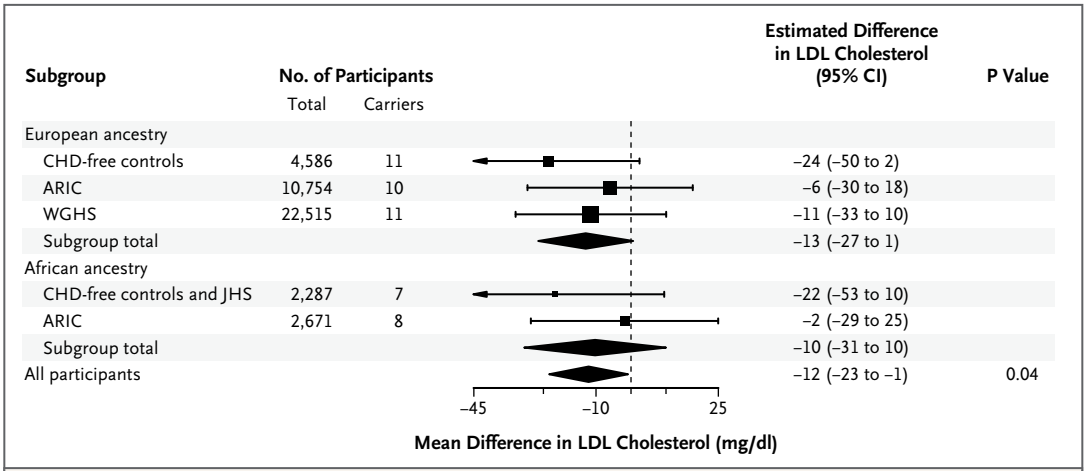


Figure 2. Association between the Presence of Inactivating Mutations in *NPC1L1* and LDL Cholesterol Levels, According to Genetic Ancestry.

In each group of participants, we tested the association between the presence of inactivating mutations in *NPC1L1* and plasma levels of low-density lipoprotein (LDL) cholesterol, after adjustment for age, sex, and study. The squares indicate the estimated adjusted difference in the LDL cholesterol level for carriers, as compared with noncarriers, in each ancestry group. The sizes of the squares are inversely proportional to the variance of the estimates. The diamonds indicate the combined results, based on a fixed-effects meta-analysis performed first within and then across ancestry groups. Participants from population-based studies — the Atherosclerosis Risk in Communities (ARIC) study, the Jackson Heart Study (JHS), and the Women's Genome Health Study (WGHS) — and controls without coronary heart disease (CHD) from case–control studies were included in this analysis. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586.

Table 3. Association between the Presence of Inactivating Mutations in *NPC1L1* and the Risk of Coronary Heart Disease (CHD).

Inactivating Mutation	Mutation Carriers		Total Participants		Carrier Frequency	
	With CHD	Without CHD	With CHD	Without CHD	Participants with CHD	Participants without CHD
	number				*percent*	
All mutations*	11	71	29,954	83,140	0.04	0.09
p.L71RfsX50	0	2	709	4,378	0	0.05
p.Q167X	0	1	966	987	0	0.10
p.A296VfsX57	0	3	1,794	1,745	0	0.17
p.R406X	6	49	26,507	75,654	0.02	0.06
p.Y483X	0	1	844	1,107	0	0.09
c.1681+1G→A†	0	3	709	4,378	0	0.07
p.W592X	1	0	1,157	4,561	0.09	0
p.R601X	1	0	474	2,362	0.21	0
p.Q604X	0	3	652	2,639	0	0.11
p.R738X	0	2	382	401	0	0.50
p.E803X	1	0	1,157	4,561	0.09	0
c.2637+2T→G†	1	1	1,525	4,897	0.07	0.02
p.C967X	0	1	474	2,362	0	0.04
p.A1201V†	0	2	235	2,016	0	0.10
p.R1325X	1	3	1,866	8,939	0.05	0.03

* The overall odds ratio for coronary heart disease in mutation carriers, as compared with noncarriers, was 0.47 (95% confidence interval, 0.25 to 0.87; P=0.008) on the basis of a meta-analysis of independent samples.
† This mutation was predicted to disrupt messenger RNA splicing.

of coronary heart disease than pharmacologic treatment that is initiated later in life. Such an effect has been observed in persons with genetic loss of function in several lipid genes.[30] In addition to affecting LDL cholesterol levels, genetic loss of NPC1L1 function is associated with reduced plant sterol absorption.[31,32] Levels of plant sterols are markedly elevated in patients with autosomal recessive sitosterolemia, a disease that is associated with accelerated atherosclerotic vascular disease even among patients without significantly elevated plasma LDL cholesterol levels.[33] These observations raise the possibility that genetic inhibition of NPC1L1 may also lower the risk of coronary heart disease by reducing the absorption of noncholesterol sterols. Also, the effect of NPC1L1 inhibition on cardiovascular risk in our study may be overestimated owing to the "winner's curse"[34] phenomenon, in which the effects of newly discovered associations are inflated as compared with the true effect sizes. Our results suggest a broad range of plausible risk estimates associated with these mutations.

Several limitations of the study deserve mention. The combined statistical evidence supporting a protective association with coronary heart disease (P=0.008) is significant for a test of a single hypothesis but falls short of the exomewide significance threshold that would be used to account for multiple hypothesis testing across all genes ($P=1.7\times10^{-6}$ on the basis of a Bonferroni correction for 21,000 protein-coding and 9000 long noncoding RNA genes).[35] This stringent threshold is used to limit false positive results of genetic association studies[35] involving many hypotheses, in which the prior probability of a true association is low. Here, however, we have evaluated a gene that is known to alter LDL cholesterol levels, a proven causal factor for coronary heart disease. Therefore, the prior probability that this

gene alters the risk of coronary heart disease is considerably higher than that for a random gene drawn from the genome.

In addition, we focused only on the classes of genetic variation — nonsense, splice-site, and frameshift — that are clearly expected to lead to a loss in NPC1L1 function and did not include missense variants. Although some missense mutations in *NPC1L1* clearly inhibit function, many others have no effect on the protein.[31,32,36,37] On average, the inclusion of neutral missense variants has been shown to dilute association signals and decrease statistical power.[38] As a result of our focus on rare inactivating mutations, the associations that we discovered are based on a relatively modest number of observations. Finally, we were unable to evaluate whether *NPC1L1* inactivating mutations lead to other phenotypic consequences.

In conclusion, on the basis of sequencing and genotyping in 113,094 study participants, we found that inactivating mutations in *NPC1L1* were associated with both reduced LDL cholesterol levels and a reduced risk of coronary heart disease. Whether pharmacologic therapies that are focused on inhibiting NPC1L1 function reduce the risk of coronary heart disease remains to be determined.

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

Supported by grants from the NHLBI (K08HL114642, to Dr. Stitziel; T32HL007208, to Dr. Peloso; R01HL107816, to Dr. Kathiresan; RC2HL102926, to Dr. Nickerson; and RC2HL102925, to Dr. Gabriel), a grant from the NHGRI (5U54HG003067-11, to Drs. Gabriel and Lander), a grant from the Foundation for Barnes–Jewish Hospital (to Dr. Stitziel), a Banting Fellowship from the Canadian Institutes of Health Research (to Dr. Do), and a grant from the Donovan Family Foundation, an investigator-initiated research grant from Merck, and a grant from Fondation Leducq (all to Dr. Kathiresan). Funders for the various studies that are discussed in this article are listed 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: Nathan O. Stitziel, M.D., Ph.D., Hong-Hee Won, Ph.D., Alanna C. Morrison, Ph.D., Gina M. Peloso, Ph.D., Ron Do, Ph.D., Leslie A. Lange, Ph.D., Pierre Fontanillas, Ph.D., Namrata Gupta, Ph.D., Stefano Duga, Ph.D., Anuj Goel, M.Sc., Martin Farrall, F.R.C.Path., Danish Saleheen, M.B., B.S., Ph.D., Paola Ferrario, Ph.D., Inke König, Ph.D., Rosanna Asselta, Ph.D., Piera A. Merlini, M.D., Nicola Marziliano, Ph.D., Maria Francesca Notarangelo, M.D., Ursula Schick, M.S., Paul Auer, Ph.D., Themistocles L. Assimes, M.D., Ph.D., Muredach Reilly, M.D., Robert Wilensky, M.D., Daniel J. Rader, M.D., G. Kees Hovingh, M.D., Ph.D., Thomas Meitinger, M.D., Thorsten Kessler, M.D., Adnan Kastrati, M.D., Karl-Ludwig Laugwitz, M.D., David Siscovick, M.D., M.P.H., Jerome I. Rotter, M.D., Stanley L. Hazen, M.D., Ph.D., Russell Tracy, Ph.D., Sharon Cresci, M.D., John Spertus, M.D., M.P.H., Rebecca Jackson, M.D., Stephen M. Schwartz, Ph.D., Pradeep Natarajan, M.D., Jacy Crosby, Ph.D., Donna Muzny, M.S., Christie Ballantyne, M.D., Stephen S. Rich, Ph.D., Christopher J. O'Donnell, M.D., Goncalo Abecasis, Ph.D., Shamil Sunyaev, Ph.D., Deborah A. Nickerson, Ph.D., Julie E. Buring, Sc.D., Paul M. Ridker, M.D., Daniel I. Chasman, Ph.D., Erin Austin, Ph.D., Zi Ye, M.D., Ph.D., Iftikhar J. Kullo, M.D., Peter E. Weeke, M.D., Christian M. Shaffer, B.S., Lisa A. Bastarache, M.S., Joshua C. Denny, M.D., Dan M. Roden, M.D., Colin Palmer, Ph.D., Panos Deloukas, Ph.D., Dan-Yu Lin, Ph.D., Zheng-zheng Tang, Ph.D., Jeanette Erdmann, Ph.D., Heribert Schunkert, M.D., John Danesh, M.B., Ch.B., D.Phil., Jaume Marrugat, M.D., Ph.D., Roberto Elosua, M.D., Ph.D., Diego Ardissino, M.D., Ruth McPherson, M.D., Hugh Watkins, M.D., Ph.D., Alex P. Reiner, M.D., James G. Wilson, M.D., David Altshuler, M.D., Ph.D., Richard A. Gibbs, Ph.D., Eric S. Lander, Ph.D., Eric Boerwinkle, Ph.D., Stacey Gabriel, Ph.D., and Sekar Kathiresan, M.D.

Drs. Stitziel and Won contributed equally to this article.

The authors' affiliations are as follows: the Cardiovascular Division, Department of Medicine (N.O.S., S.C.), Division of Statistical Genomics (N.O.S.), and Department of Genetics (S.C.), Washington University School of Medicine, St. Louis; Center for Human Genetic Research (H.-H.W., G.M.P., R.D., P.N., D. Altshuler, S.K.), Cardiovascular Research Center (H.-H.W., G.M.P., R.D., P.N., S.K.), and Cardiology Division (C.J.O., S.K.), Massachusetts General Hospital, the Department of Medicine, Harvard Medical School (H.-H.W., G.M.P., R.D., P.N., C.J.O., D. Altshuler, S.K.), and the Divisions of Genetics (S.S.) and Preventive Medicine (J.E.B., P.M.R., D.I.C.), Brigham and Women's Hospital and Harvard Medical School, Boston, the Program in Medical and Population Genetics, Broad Institute, Cambridge (H.-H.W., G.M.P., R.D., P. Fontanillas, N.G., P.N., S.S., D. Altshuler, E.S.L., S.G., S.K.), and the National Heart, Lung, and Blood Institute Framingham Heart Study, Framingham (C.J.O.) — all in Massachusetts; the Human Genetics Center, University of Texas Health Science Center at Houston (A.C.M., J.C., E.B.), and the Human Genome Sequencing Center (D.M., R.A.G., E.B.), and Section of Atherosclerosis and Vascular Medicine (C.B.), Baylor College of Medicine — both in Houston; the Departments of Genetics (L.A.L.), and Biostatistics (D.-Y.L.), University of North Carolina, Chapel Hill; Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milan (S.D., R.A.), Azienda Ospedaliera-Universitaria di Parma, Parma (P.A.M., N.M., M.F.N., D. Ardissino), and Associazione per lo Studio della Trombosi in Cardiologia, Pavia (P.A.M., D. Ardissino) — all in Italy; the Division of Cardiovascular Medicine, Radcliffe Department of Medicine and the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford (A.G., M.F., H.W.), Public Health and Primary Care, University of Cambridge (J.D.), and Wellcome Trust Sanger Institute (P.D.), Cambridge, Medical Research Institute, University of Dundee, Dundee (C.P.), and William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London (P.D.) — all in the United Kingdom; the Departments of Genetics (D.J.R.) and Biostatistics and Epidemiology (D. Saleheen) and the Cardiovascular Institute (M.R., R.W.), Perelman School of Medicine, University of Pennsylvania, Philadelphia; Center for Noncommunicable Diseases, Karachi, Pakistan (D. Saleheen); Institut für Medizinische Biometrie und Statistik (P. Ferrario, I.K.) and Institut für Integrative und Experimentelle Genomik (J.E.), Universität zu Lübeck, and DZHK (German Research Center for Cardiovascular Research) partner site Hamburg–Lübeck–Kiel (J.E.), Lübeck,

Institut für Humangenetik, Helmholtz Zentrum, Neuherberg (T.M.), and Institut für Humangenetik (T.M.) and Medizinische Klinik (K.-L.L.), Klinikum rechts der Isar, and Deutsches Herzzentrum München (T.K., A.K., H.S.),Technische Universität München, and German Center for Cardiovascular Research (DZHK), partner site Munich Heart Alliance (T.M., A.K., K.-L.L., H.S.), Munich — all in Germany; the Division of Public Health Sciences, Fred Hutchinson Cancer Research Center (U.S., S.M.S., A.P.R.), and the Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology (D. Siscovick), Department of Epidemiology (S.M.S., A.P.R.), and Department of Genome Sciences (D.A.N.), University of Washington — both in Seattle; School of Public Health, University of Wisconsin–Milwaukee, Milwaukee (P.A.); Stanford Cardiovascular Institute and the Division of Cardiovascular Medicine, Stanford University, Stanford (T.L.A.), and Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor–UCLA Medical Center, Torrance (J.I.R.) — both in California; the Department of Vascular Medicine, Academic Medical Center, Amsterdam (G.K.H.); the Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland (S.L.H.), and the Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Ohio State University, Columbus (R.J.) — both in Ohio; the Departments of Pathology and Biochemistry, University of Vermont College of Medicine, Burlington (R.T.); St. Luke's Mid America Heart Institute, University of Missouri–Kansas City, Kansas City (J.S.); Center for Public Health Genomics, University of Virginia, Charlottesville (S.S.R.); Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor (G.A.); the Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN (E.A., Z.Y., I.J.K.); the Departments of Medicine (P.E.W., C.M.S., J.C.D., D.M.R.), Biomedical Informatics (L.A.B., J.C.D.), Pharmacology (D.M.R.), and Biostatistics (Z.T.), Vanderbilt University, Nashville; the Department of Cardiology, Laboratory of Molecular Cardiology, Copenhagen University Hospital Rigshospitalet, Copenhagen (P.E.W.); King Abdulaziz University, Jeddah, Saudi Arabia (P.D.); Grupo de Epidemiología y Genética Cardiovascular, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona (J.M., R.E.); the Division of Cardiology, University of Ottawa Heart Institute, Ottawa (R.M.); and the Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson (J.G.W.).

REFERENCES

1. Garcia-Calvo M, Lisnock J, Bull HG, et al. The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1). Proc Natl Acad Sci U S A 2005;102:8132-7.
2. Davis HR Jr, Zhu LJ, Hoos LM, et al. Niemann-Pick C1 Like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole-body cholesterol homeostasis. J Biol Chem 2004;279:33586-92.
3. Altmann SW, Davis HR Jr, Zhu LJ, et al. Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. Science 2004;303:1201-4.
4. Sudhop T, Lütjohann D, Kodal A, et al. Inhibition of intestinal cholesterol absorption by ezetimibe in humans. Circulation 2002;106:1943-8.
5. Ezzet F, Wexler D, Statkevich P, et al. The plasma concentration and LDL-C relationship in patients receiving ezetimibe. J Clin Pharmacol 2001;41:943-9.
6. Kastelein JJ, Akdim F, Stroes ES, et al. Simvastatin with or without ezetimibe in familial hypercholesterolemia. N Engl J Med 2008;358:1431-43. [Erratum, N Engl J Med 2008;358:1977.]
7. 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.
8. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov 2013;12:581-94.
9. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707-13.
10. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 1992;90:41-54.
11. Do R, Stitziel NO, Won H-H, et al. Multiple rare alleles at *LDLR* and *APOA5* confer risk for early-onset myocardial infarction. Nature (in press).
12. Atherosclerosis, Thrombosis, and Vascular Biology Italian Study 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.
13. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007;316:1488-91.
14. 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.
15. Saleheen D, Zaidi M, Rasheed A, et al. The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. Eur J Epidemiol 2009;24:329-38.
16. Sentí M, Tomás M, Marrugat J, Elosua R. Paraoxonase1-192 polymorphism modulates the nonfatal myocardial infarction risk associated with decreased HDLs. Arterioscler Thromb Vasc Biol 2001;21:415-20.
17. Crosby J, Peloso GM, Auer PL, et al. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. N Engl J Med 2014;371:22-31.
18. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 1989;129:687-702.
19. Taylor HA Jr. The Jackson Heart Study: an overview. Ethn Dis 2005;15:Suppl 6: S6-1–S6-3.
20. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther 2008;84: 362-9.
21. Doney AS, Fischer B, Leese G, Morris AD, Palmer CN. Cardiovascular risk in type 2 diabetes is associated with variation at the PPARG locus: a Go-DARTS study. Arterioscler Thromb Vasc Biol 2004;24: 2403-7.
22. Erdmann J, Stark K, Esslinger UB, et al. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. Nature 2013;504:432-6.
23. Ye Z, Kalloo FS, Dalenberg AK, Kullo IJ. An electronic medical record-linked biorepository to identify novel biomarkers for atherosclerotic cardiovascular disease. Glob Cardiol Sci Pract 2013;2013:82-90.
24. Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. Clin Chem 2008;54:249-55.
25. 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.
26. Baigent C, Blackwell L, Emberson J, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet 2010;376: 1670-81.
27. Betters JL, Yu L. NPC1L1 and cholesterol transport. FEBS Lett 2010;584:2740-7.
28. Krumholz HM. Emphasizing the burden of proof: the American College of Cardiology 2008 Expert Panel Comments on the ENHANCE Trial. Circ Cardiovasc Qual Outcomes 2010;3:565-7.
29. Cannon CP, Giugliano RP, Blazing MA, et al. Rationale and design of IMPROVE-IT (IMProved Reduction of Outcomes: Vytorin Efficacy International Trial): comparison of ezetimibe/simvastatin versus simvastatin monotherapy on cardiovascular outcomes in patients with

acute coronary syndromes. Am Heart J 2008;156:826-32.
30. Cohen JC, Stender S, Hobbs HH. APOC3, coronary disease, and complexities of Mendelian randomization. Cell Metab 2014;20:387-9.
31. Cohen JC, Pertsemlidis A, Fahmi S, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci U S A 2006; 103:1810-5.
32. Fahmi S, Yang C, Esmail S, Hobbs HH, Cohen JC. Functional characterization of genetic variants in NPC1L1 supports the sequencing extremes strategy to identify complex trait genes. Hum Mol Genet 2008;17:2101-7.
33. Berge KE, Tian H, Graf GA, et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science 2000;290:1771-5.
34. Kraft P. Curses — winner's and otherwise — in genetic epidemiology. Epidemiology 2008;19:649-51.
35. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. Nature 2014;508:469-76.
36. Hegele RA, Guy J, Ban MR, Wang J. NPC1L1 haplotype is associated with inter-individual variation in plasma low-density lipoprotein response to ezetimibe. Lipids Health Dis 2005;4:16.
37. Wang J, Williams CM, Hegele RA. Compound heterozygosity for two nonsynonymous polymorphisms in NPC1L1 in a non-responder to ezetimibe. Clin Genet 2005;67:175-7.
38. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 2014;111:E455-64.

Supplementary Appendix

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

Supplement to: The Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in *NPC1L1* and protection from coronary heart disease. N Engl J Med. DOI: 10.1056/NEJMoa1405386

Supplementary Appendix for

Inactivating mutations in *NPC1L1* and protection from coronary heart disease

Nathan O. Stitziel, M.D., Ph.D.[1,2,*], Hong-Hee Won, Ph.D.[3,4,5,6,*], Alanna C. Morrison, Ph.D.[7], Gina M. Peloso, Ph.D.[3,4,5,6], Ron Do, Ph.D.[3,4,5,6], Leslie A. Lange, Ph.D.[8], Pierre Fontanillas, Ph.D.[6], Namrata Gupta, Ph.D.[6], Stefano Duga, Ph.D.[9], Anuj Goel, M.Sc.[10], Martin Farrall, F.R.C.Path.[10], Danish Saleheen, MBBS, Ph.D.[11], Paola Ferrario, Ph.D.[12], Inke König, Ph.D.[12], Rosanna Asselta, Ph.D.[9], Piera Angelica Merlini, M.D.[13,14], Nicola Marziliano, Ph.D.[13], Maria Francesca Notarangelo, M.D.[13], Ursula Schick, M.S.[15], Paul Auer, Ph.D.[16], Themistocles L. Assimes, M.D., Ph.D.[17], Muredach Reilly, M.D.[18], Robert Wilensky, M.D.[18], Daniel J. Rader, M.D.[19], G. Kees Hovingh, M.D., Ph.D.[20], Thomas Meitinger, M.D.[21,22], Thorsten Kessler, M.D.[23], Adnan Kastrati, M.D.[22,23], Karl-Ludwig Laugwitz, M.D.[22,24], David Siscovick, M.D., M.P.H.[25], Jerome I. Rotter, M.D.[26], Stanley L. Hazen, M.D., Ph.D.[27], Russell Tracy, Ph.D.[28], Sharon Cresci, M.D.[1,29], John Spertus, M.D., M.P.H.[30], Rebecca Jackson, M.D.[31], Stephen M. Schwartz, Ph.D.[15,32], Pradeep Natarajan, M.D.[3,4,5,6], Jacy Crosby, Ph.D.[7], Donna Muzny, M.S.[33], Christie Ballantyne, M.D.[34], Stephen S. Rich, Ph.D.[35], Christopher J. O'Donnell, M.D.[5,36,37], Goncalo Abecasis, Ph.D.[38], Shamil Sunyaev, Ph.D.[6,39], Deborah A. Nickerson, Ph.D.[40], Julie E. Buring, Sc.D.[41], Paul M. Ridker, M.D.[41], Daniel I. Chasman, Ph.D.[41], Erin Austin, Ph.D.[42], Zi Ye, M.D., Ph.D.[42], Iftikhar J. Kullo, M.D.[42], Peter E. Weeke, M.D.[43,44], Christian M. Shaffer, B.S.[43], Lisa A. Bastarache, M.S.[45], Joshua C. Denny, M.D., M.S.[43,45], Dan M. Roden, M.D.[43,46], Colin Palmer, Ph.D.[47], Panos Deloukas, Ph.D.[48], Dan-Yu Lin, Ph.D.[49], Zheng-zheng Tang, Ph.D.[50], Jeanette Erdmann, Ph.D.[51,52], Heribert Schunkert, M.D.[22,23], John Danesh, M.B., Ch.B., D.Phil.[53], Jaume Marrugat, M.D., Ph.D.[54], Roberto Elosua, M.D., Ph.D.[54], Diego Ardissino, M.D.[13,14], Ruth McPherson, M.D.[55], Hugh Watkins, M.D., Ph.D.[10], Alex P. Reiner, M.D., M.Sc.[15,32], James G. Wilson, M.D.[56], David Altshuler, M.D., Ph.D. [3,5,6], Richard A. Gibbs, Ph.D.[33], Eric S. Lander, Ph.D.[6], Eric Boerwinkle, Ph.D.[7,33], Stacey Gabriel, Ph.D.[6], Sekar Kathiresan, M.D.[3,4,5,6,37]

Address for Correspondence:
Sekar Kathiresan, M.D.
Cardiovascular Research Center and Center for Human Genetic Research
Massachusetts General Hospital
185 Cambridge Street, CPZN 5.252
Boston, MA 02114
skathiresan@partners.org

1. Cardiovascular Division, Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA
2. Division of Statistical Genomics, Washington University School of Medicine, Saint Louis, MO, USA
3. Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA
4. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
5. Department of Medicine, Harvard Medical School, Boston, MA, USA
6. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
7. Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX, USA
8. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
9. Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milan, Italy
10. Division of Cardiovascular Medicine, Radcliffe Department of Medicine and The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
11. Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA; Center for Non-Communicable Diseases, Karachi, Pakistan
12. Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany
13. Azienda Ospedaliero-Universitaria di Parma, Parma, Italy
14. Associazione per lo Studio della Trombosi in Cardiologia, Pavia, Italy
15. The Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
16. School of Public Heath, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
17. Stanford Cardiovascular Institute and the Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA
18. Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
19. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
20. Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands
21. Institut für Humangenetik, Helmholtz Zentrum, Neuherberg, Germany and Institut für Humangenetik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
22. German Centre for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany
23. Deutsches Herzzentrum München, Technische Universität München, Munich, Germany

24. 1. Medizinische Klinik, Klinikum rechts der Isar, Technische Universität, München, Munich, Germany
25. Cardiovascular Health Research Unit, Department of Medicine, and Department of Epidemiology, University of Washington, Seattle, WA, USA
26. Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA
27. Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH, USA
28. Departments of Pathology and Biochemistry, University of Vermont College of Medicine, Burlington, VT, USA
29. Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA
30. St. Luke's Mid America Heart Institute, University of Missouri-Kansas City, Kansas City, MO, USA
31. Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Ohio State University, Columbus, OH, USA
32. Department of Epidemiology, University of Washington, Seattle, Washington, USA
33. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
34. Section of Atherosclerosis and Vascular Medicine, Baylor College of Medicine, Houston, TX, USA
35. Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
36. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA
37. Cardiology Division, Massachusetts General Hospital, Boston, MA, USA
38. Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.
39. Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
40. Department of Genome Sciences, University of Washington, Seattle, WA, USA
41. Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
42. Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA
43. Department of Medicine, Vanderbilt University, Nashville, TN, USA
44. Department of Cardiology, Laboratory of Molecular Cardiology, Copenhagen University Hospital Rigshospitalet, Denmark
45. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA
46. Department of Pharmacology, Vanderbilt University, Nashville, TN, USA
47. Medical Research Institute, University of Dundee, Dundee, UK
48. 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
49. Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA
50. Department of Biostatistics, Vanderbilt University, Nashville, TN, USA
51. Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, 23562 Lübeck, Germany
52. DZHK (German Research Centre for Cardiovascular Research) partner site Hamberg/Lübeck /Kiel, Lübeck, Germany
53. Public Health and Primary Care, University of Cambridge, Cambridge, UK
54. Grupo de Epidemiología y Genética Cardiovascular, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain
55. Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada
56. Department of Physiology and Biophysics, the University of Mississippi Medical Center, Jackson, MS, USA

Table of Contents

Additional Acknowledgements

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Sequencing was carried out at the Baylor Genome Center (U54 HG003273). The German MI Family Studies I-V and the Munich-MI studies are supported by the German Federal Ministry of Education and Research (BMBF) in the context of the e:Med program (*e:AtheroSysMed*), the FP7 European Union project *CVgenes@target* (261123), and by the Fondation Leducq (CADgenomics: Understanding Coronary Artery Disease Genes, 12CVD02). The Italian ATVB Study was supported by grant RFPS-2007-3-644382 "*Think pink about genes. Gender-specific genetic variants associated with early-onset myocardial infarction: a genome-wide association study*" and supported by the Programma di ricerca Regione-Università, Regione Emilia-Romagna, bando "Ricerca Innovativa" 2010-2012 to Dr. Diego Ardissino, "*Cardiovascular genetics: from bench to bedside - Genomic & transcriptomic of ischemic heart disease*". CUP E35E09000880002.

The Jackson Heart Study is supported by contracts HHSN268201300046C,

HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities. The REGICOR study was supported by the Spanish Ministry of Economy and Innovation through the Carlos III Health Institute (Red HERACLES RD06/0009, PI09/90506), European Funds for Development (ERDF-FEDER), and by the Catalan Research and Technology Innovation Interdepartmental Commission (SGR 1195). TRIUMPH was funded by NIH Specialized Center for Clinically-Oriented Research (SCCOR) in Cardiac Dysfunction and Disease P50 HL077113. SC is funded by NIH RO1 NR013396. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). Exome chip data and analysis were supported through the Exome Sequencing Project (NHLBI RC2 HL-102924, RC2 HL-102925 and RC2 HL-102926), the Genetics and Epidemiology of Colorectal Cancer Consortium (NCI CA137088), the Genomics and Randomized Trials Network (NHGRI U01-HG005152) and a National Cancer Institute training grant (R25CA094880). The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible. The WGHS is supported by HL043851, HL080467, HL099355 from the National Heart, Lung, and Blood Institute, CA047988 from the National Cancer Institute, and the Donald W. Reynolds Foundation with collaborative scientific

support and funding for genotyping provided by Amgen. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHGRI, NHLBI, or NIH.

Supplementary Methods: Sequencing

Exome sequencing was performed in samples from the sequencing phase at the Broad Institute, the Baylor College of Medicine Human Genome Sequencing Center, or the University of Washington. The following is a summary of sequencing procedures followed at the Broad Institute; similar methods were utilized at the other sites. In brief, following DNA quantification using PicoGreen, we confirmed high-molecular weight DNA and performed fingerprint genotyping and gender determination using Illumina iSelect platform. From samples passing these initial quality steps, 3μg of genomic DNA was used to perform library construction and in-solution hybrid selection[1] to target 33Mb of genomic sequence. The resulting exome-enriched DNA was sequenced on either Genome Analyzer II using v3 and v4 Sequencing-by-Synthesis Kits, then analyzed using RTA v1.7.48 or on HiSeq 2,000 using HiSeq 2,000 v2 Sequencing-by-Synthesis Kits, then analyzed using RTA v1.10.15. Sequencing was performed using 76 cycle paired-end runs. Sequencing was considered complete when ≥ 80% of targeted bases were covered with ≥ 20 sequencing reads.

Raw sequence reads were aligned to the human reference genome (HG19) using the Burroughs-Wheeler Alignment tool[2] in paired-end mode. Duplicate reads and reads aligned outside of the exome target were removed.

The Genome Analysis ToolKit[3] (GATK) was then used to locally realign reads, recalibrate base qualities, identify and genotype single nucleotide variants (SNVs) and short insertion and deletion events (indels), and recalibrate the resulting variant quality scores. SnpEff was used to predict the functional consequences of the identified variants[4].

Quality Control for Sequencing

SNVs were flagged for removal if they had low quality (quality per depth score < 5), exhibited strand bias (strand bias ≥ 0.1), or were part of a homopolymer run greater than 4. Indels were flagged for removal if they had low quality (quality score < 30 or quality per depth score < 2) or exhibited strand bias (strand bias > -1.0). We then removed samples that were incompletely sequenced (i.e., did not reach ≥ 80% of targeted bases covered with ≥ 20 reads), those that were discordant between pre- and post-sequencing fingerprints, or were discordant between inferred and reported gender. We also removed samples that were statistical outliers for several sequencing metrics including heterozygosity, the ratio of transitions to transversions, singleton count, total variant count, and missingness. Finally, we removed samples that were statistical outliers in population clustering using principal components analysis and samples with discordance between inferred and self-reported ancestry. We then assessed additional sequencing metrics to refine the variant quality control process to remove those variants likely to be artifacts and removed variants that were statistical outliers for depth of coverage, quality over depth, Hardy Weinberg

equilibrium, and frequency of missing genotypes. SNVs and indels occurring in the protein coding regions of *NPC1L1* were then extracted for this study.

Supplementary Methods: Genotyping

Samples from the genotyping phase cohorts were genotyped for *NPC1L1* p.Arg406X. In brief, after DNA quantification and quality control, samples were processed on the Illumina HumanExome BeadChip array (Illumina, Inc., CA, USA) according to standard protocols suggested by the manufacturer. Genotypes were assigned using GenomeStudio and supplemented with the zCall algorithm[5] when possible. Using these same methods, we also genotyped 2,495 samples from ATVB and 1,597 samples from PROCARDIS that had undergone *NPC1L1* sequencing.

Quality Control for Genotyping

After ensuring genotypes were aligned to the positive strand we removed samples that had an excess of missing genotypes (≥ 5%), were discordant between inferred and reported gender, duplicate samples, statistical outliers for inbreeding coefficient, samples with high proportion of shared genotypes identical by descent, samples with cryptic relatedness, and statistical outliers in principal components analysis. From these samples that passed quality control we extracted genotypes for *NPC1L1* p.R406X (also known as rs145297799 and exm618018).

Technical Validation

We sought to technically validate the presence of inactivating mutations in individuals from the ATVB cohort using the orthogonal method of Sanger sequencing. These mutations were selected for validation: a) to ensure the frameshift indel (p.A296VfsX57) was not an artifact given a higher false discovery rate for indel detection using next-generation sequencing compared with single nucleotide substitutions; and b) to ensure the relatively higher carrier rate in ATVB compared with other sequencing-phase cohorts was not a technical artifact. As shown in Figure S1, the single nucleotide deletion inducing p.A296VfsX57 was also discovered using Sanger sequencing. This deletion and p.R406X were confirmed in all carriers from ATVB.

We also performed technical validation for p.R406X genotyping. Given the low frequency of p.R406X and potential difficulty in identifying carriers through genotyping, we sought to determine the presence of false positives (carriers spuriously identified through genotyping but not confirmed through sequencing) and false negatives (real carriers identified through sequencing but missed by genotyping). In 4,092 samples that underwent both *NPC1L1* sequencing and genotyping we detected zero false positives and zero false negatives.

Association Testing to Account for Confounding

Due to the rarity of inactivating mutations in *NPC1L1*, we had limited statistical power to detect an association with CHD for any single mutation. Therefore, we calculated the summary odds ratio and associated 95% confidence interval of

CHD for carriers of any inactivating mutation across studies using a Mantel-Haenszel meta-analysis. To eliminate the possibility of confounding from population stratification or cryptic relatedness, we performed an additional association test in a subset of samples sequenced at the Broad Institute using a mixed linear model as implemented by EMMAX[6]. The mixed linear model accounts for multiple layers of stratification in the sample sets by incorporating a kinship matrix in the association test to correct for population stratification and sample relatedness. The level of statistical significance for the mixed linear model was similar to the Mantel-Haenszel approach (data not shown), suggesting the association results are not a result of cryptic relatedness or stratification.

Table S1. Sample sets included in the sequencing phase of the study

Study	Study Design	N Cases	N Controls	CHD definition	CHD-free control definition	Ref
ARIC	Prospective cohort	1631	6923	Incident probable or definite MI, silent MI, definite CHD death, or coronary revascularization	Free of CHD during follow-up	7
ATVB	Case-control	1794	1745	MI in men or women ≤ 45 years of age	No history of thromboembolic disease	8
ESP-EOMI	Case-control	178	277	MI in men ≤ 50 years of age or women ≤ 60 years of age	Free of MI, coronary revascularization; men ≥ 50 years of age or women ≥ 60 years of age	9
JHS	Prospective cohort	235	2016	Combination of prevalent CHD (self-reported or electrocardiographic evidence of MI) and incident CHD (MI or coronary revascularization as previously described[10]	Free of CHD during follow-up	11
Munich-MI	Case-control	368	336	MI in men ≤ 40 years of age or women ≤ 55 years of age	Controls without CAD, men ≥ 65 years of age and women ≥ 75 years of age	12
OHS	Case-control	966	987	Angiographically confirmed coronary artery disease (>1 coronary artery with >50% stenosis) without history of diabetes at age ≤ 50 for men or ≤ 60 for women	Asymptomatic, men > 65, women > 70	13
PROCARDIS	Case-control	966	936	Symptomatic CAD before age 66. CAD was defined as clinically documented evidence of myocardial infarction, coronary artery bypass grafting, acute coronary syndrome, coronary angioplasty, or stable angina	No personal or sibling history of CAD before 66 years of age	14
PROMIS	Case-control	844	1107	MI in men and women ≤ 45 years of age	No history of cardiovascular disease	15
REGICOR	Case-control	382	401	MI in men ≤ 50 years of age or women ≤ 60 years of age	Controls from a population-based study; free of MI, coronary revascularization; ≥ 55 and <80 years of age	16

Table S2. Sample sets included in the genotyping phase of the study

Study	Study Design	N Cases	N Controls	CHD definition	CHD-free control definition	Ref
ARIC	Prospective cohort	1142	4095	Incident probable or definite MI, silent MI, definite CHD death, or coronary revascularization	Free of CHD during follow-up	[7]
BioVU	Case-control	4587	16556	Cases with were ascertained from the Vanderbilt University Medical Center (VUMC) Biorepository by searching the electronic medical record for ≥ 2 instances of ICD-9 codes 410.x – 414.x	Individuals from the VUMC Biorepository without any record of ICD-9 codes 410.x – 414.x	[17]
German North	Case-control	4464	2886	The German North cohort includes individuals from GerMIFS4, PopGen, and HNR with MI or CAD	Controls from population-based studies in Germany	[18,19,20]
German South	Case-control	5255	2921	The German South cohort includes samples from GerMIFS3 and Munich-MI with MI or CAD	Controls from population-based studies in Germany	[21,22]
GoDARTS	Case-control	997	2768	The GoDARTS (Genetics of Diabetes Audit and Research in Tayside 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 Tayside general practitioners with an interest in diabetes care. Cases were defined as fatal and non-fatal myocardial infarction.	Controls were free of CAD, stroke, and peripheral vascular disease	[23]
Mayo	Case-control	1177	1492	History of MI, coronary revascularization, angina with positive stress test, or >50% stenosis of an epicardial coronary artery	No history of MI or atherosclerotic vascular disease	[24]
PROCARDIS	Case-control	1132	1095	Symptomatic CAD before age 66. CAD was defined as clinically documented evidence of myocardial infarction, coronary artery bypass grafting, acute coronary syndrome, coronary angioplasty, or stable angina	No personal or sibling history of CAD before 66 years of age	[14]
WGHS	Prospective clinical trial	976	21641	Prospectively ascertained MI, coronary revascularization (PCTA or CABG), and cardiovascular death	Free of CHD events during follow-up	[25,26]
WHI	Prospective cohort	2860	14958	MI, coronary revascularization, hospitalized angina, or death due to CHD	Free of CHD during follow-up	[27]

Table S3. Clinical characteristics of *NPC1L1* inactivating mutation carriers compared with non-carriers

Cohort		Number	Male gender (%)	Age, mean (SD), years	BMI, mean (SD) kg/m2	Current or Former Smoker (%)	Diabetes (%)	Systolic blood pressure, mean (SD), mmHg
ARIC AA	C	8	12.5%	54.1 (5.9)	26.9 (3.7)	50%	12.5%	129.9 (25.2)
	NC	2828	36.2%	53.2 (5.8)	29.8 (6.3)	52.5%	17.2%	127.2 (19.0)
ARIC EA	C	10	50%	53.4 (4.4)	25.1 (5.7)	60%	0%	112.2 (14.2)
	NC	10945	46.9%	54.3 (5.7)	27.0 (4.9)	59.7%	8.8%	118.4 (17.0)
ATVB	C	7	86%	37.4 (6.4)	24.5 (1.9)	71%	0%	120.0 (10)
	NC	3532	89%	39.6 (4.9)	25.9 (3.9)	68%	3.2%	129.1 (18.7)
BioVU	C	12	50%	60.7 (18.0)	30.0 (9.2)	50%	33.3%	126.1 (10.4)
	NC	21131	46%	65.1 (17.0)	27.8 (6.0)	26.7%	14.1%	126.1 (13.0)
ESP-EOMI	C	1	100%	72.6 (NA)	28.8 (NA)	0%	0%	101.5 (NA)
	NC	454	26%	62.5 (13.5)	31.6 (7.2)	60%	31.7%	138.0 (31.8)
German North	C	1	100%	53.0 (NA)	28.2 (NA)	100%	NA	NA
	NC	7349	60%	54.7 (10.3)	27.0 (4.4)	55%	NA	NA
German South	C	3	67%	64.3 (7.1)	29.3 (3.7)	66%	33%	NA
	NC	8173	62%	58.8 (11.7)	27.35 (4.6)	33.6%	12.4%	NA
GoDARTS	C	4	25%	67.5 (5.6)	31.7 (9.8)	25%	75%	151.2 (10.4)
	NC	3761	60%	61.4 (11.1)	30.9 (5.7)	61%	45%	140.4 (17.2)
JHS	C	6	67%	56.8 (17.6)	32.3 (7.7)	33%	33%	120.4 (18.1)
	NC	2245	37%	50.8 (12.5)	32.6 (4.1)	30.5%	21.3%	124.8 (17.6)
Mayo	C	2	0%	63.3 (8.7)	29.7 (0.1)	0%	0%	132.5 (16.3)
	NC	2667	59.6%	64.5 (9,3)	29.3 (5.6)	60.5%	19.9%	123.9 (17.3)
Munich-MI	C	1	100%	71 (NA)	26.04 (NA)	100%	0%	142 (NA)
	NC	703	64.3%	56.9 (16.6)	27.5 (7.5)	52.2%	12.5%	136.1 (27.2)
OHS	C	1	0%	78.1 (NA)	24.9 (NA)	58.56%	0%	106 (NA)
	NC	1952	67%	64.4 (16.3)	27.3 (4.4)	0%	0%	135.2 (18.8)
PROCARDIS	C	2	100%	65.5 (0.71)	29.4 (5.4)	50%	50%	142.5 (17.7)
	NC	4127	59.9%	56.6 (11.9)	27.8 (4.6)	61.96%	10.2%	134.7 (18.8)
PROMIS	C	1	100%	44 (NA)	25.8 (NA)	100%	0%	140 (NA)
	NC	1950	87.8%	49.5 (10.6)	25.1 (3.9)	45.6%	14.5%	128.8 (19.5)
REGICOR	C	2	100%	68 (5.7)	27.5 (1.1)	50%	0%	152.5 (9.2)
	NC	781	78.4%	57.1 (12.1)	27.6 (3.9)	41.4%	21.5%	139.4 (18.5)
WGHS	C	11	0%	55.1 (5.9)	23.7 (3.5)	54.5%	0%	123.2 (11.7)
	NC	22606	0%	54.7 (7.1)	25.9 (5.0)	49.1%	2.5%	123.6 (13.7)
WHI	C	10	0%	64.8 (6.6)	27.1 (5.1)	0%	11.1%	130.1 (20.7)
	NC	17808	0%	66.4 (6.7)	28.3 (5.9)	7.7%	6.8%	129.9 (17.9)

C=Carrier of NPC1L1 inactivating mutation; NC=Noncarrier; BMI=Body mass index; NA=Not available; SD=Standard deviation; AA=African ancestry; EA=European ancestry

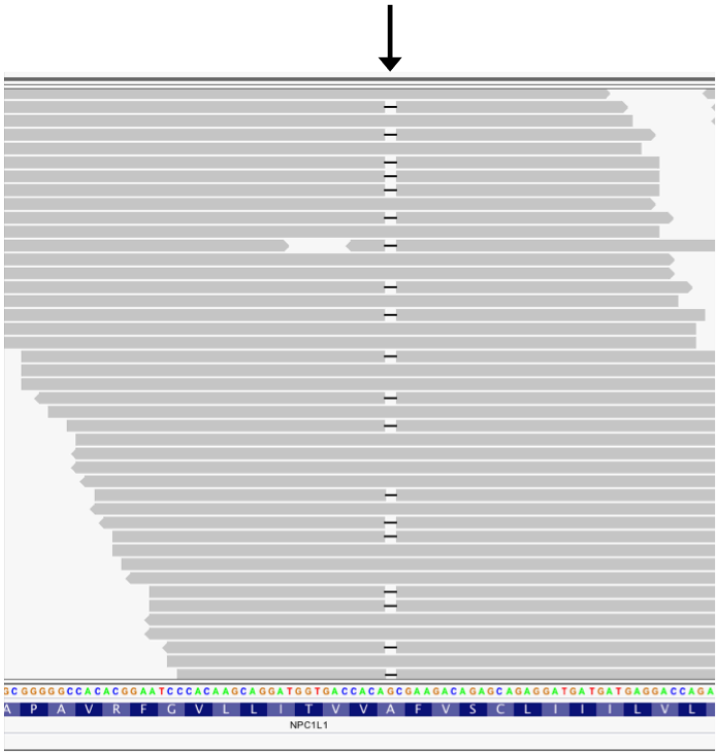
Table S4. Association of NPC1L1 inactivating mutations with risk of coronary heart disease by cohort

Cohort	Ancestry	Carriers		Total Number		Carrier frequency		OR* (95% CI)‡	P value
		Cases	Controls	Cases	Controls	Cases	Controls		
ARIC AA	AA	1	7	474	2362	0.21%	0.30%		
ARIC EA†	EA	3	7	2299	8656	0.13%	0.08%		
ATVB	EA	1	6	1794	1745	0.06%	0.34%		
ESP EOMI	AA	0	1	178	277	0%	0.36%		
OHS	EA	0	1	966	987	0%	0.10%		
PROCARDIS†	EA	0	2	2098	2031	0%	0.10%		
JHS	AA	1	5	235	2016	0.43%	0.25%		
Munich-MI	EA	0	1	368	336	0%	0.30%		
REGICOR	EA	0	2	382	401	0%	0.50%		
PROMIS	SA	0	1	844	1107	0%	0.09%		
BioVU	EA	1	11	4587	16556	0.022%	0.066%		
German North	EA	0	1	4464	2886	0%	0.03%		
German South	EA	1	2	5255	2921	0.02%	0.07%		
GoDARTS	EA	0	4	997	2768	0%	0.14%		
Mayo	EA	0	2	1177	1492	0%	0.13%		
WGHS	EA	2	9	976	21641	0.21%	0.042%		
WHI	EA	1	9	2860	14958	0.03%	0.06%		
Total		11	71	29954	83140	0.037%	0.085%	0.47 (0.25-0.87)	8.0×10^{-3}

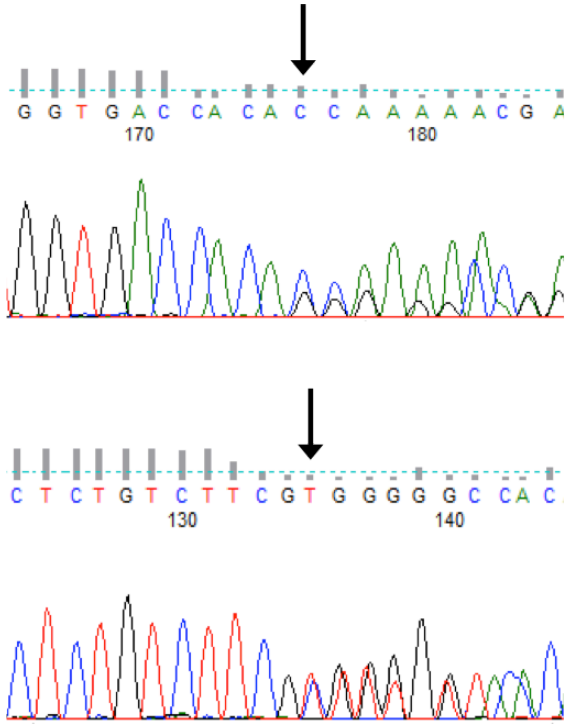
*OR=Odds Ratio; CI=Confidence interval; †Independent samples from sequencing and genotyping combined; ‡Test for heterogeneity P=0.43; AA=African ancestry; EA=European ancestry; SA=South Asian ancestry

Figure S1. Technical validation of frameshift indel observed in ATVB. A) Integrated genomics viewer screenshot of next-generation sequencing data supporting the p.A296VfsX57 frameshift in one individual. Horizontal gray bars represent individual sequencing reads aligned to the human reference genome below. The arrow indicates the alignment gap induced by the heterozygous single base-pair deletion. B) Chromatograms from Sanger confirmation sequencing of the forward (top) and reverse (bottom) strands. Arrows indicate the position of the deletion and the beginning of the frameshift.

A)



B)



Supplementary References

1. 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.
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. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 2012;6:80-92.
5. 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.
6. Kang HM, Sul JH, Service SK, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet 2010;42:348-54.
7. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 1989;129:687-702.
8. Atherosclerosis, Thrombosis, and Vascular Biology Italian Study 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.
9. Do R, Stitziel NO, Won H-H, et al. Exome sequencing identifies multiple rare alleles at *LDLR* and *APOA5* that confer risk for myocardial infarction. Nature 2014;In Press.
10. White AD, Folsom AR, Chambless LE, et al. Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. Journal of clinical epidemiology 1996;49:223-33.
11. Taylor HA, Jr. The Jackson Heart Study: an overview. Ethnicity & disease 2005;15:S6-1-3.
12. Crosby J, Peloso GM, Auer PL, et al. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. N Engl J Med 2014;371:22-31.
13. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007;316:1488-91.
14. 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.
15. Saleheen D, Zaidi M, Rasheed A, et al. The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other

determinants of myocardial infarction in South Asia. European journal of epidemiology 2009;24:329-38.
16. Senti M, Tomas M, Marrugat J, Elosua R, Investigators R. Paraoxonase1-192 polymorphism modulates the nonfatal myocardial infarction risk associated with decreased HDLs. Arterioscler Thromb Vasc Biol 2001;21:415-20.
17. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clinical pharmacology and therapeutics 2008;84:362-9.
18. Erdmann J, Stark K, Esslinger UB, et al. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. Nature 2013;504:432-6.
19. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community genetics 2006;9:55-61.
20. Schmermund A, Mohlenkamp S, Stang A, et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. American heart journal 2002;144:212-8.
21. 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.
22. Koch W, Turk S, Erl A, et al. The chromosome 9p21 region and myocardial infarction in a European population. Atherosclerosis 2011;217:220-6.
23. 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.
24. Ye Z, Kalloo FS, Dalenberg AK, Kullo IJ. An electronic medical record-linked biorepository to identify novel biomarkers for atherosclerotic cardiovascular disease. Global cardiology science & practice 2013;2013:82-90.
25. Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. Clin Chem 2008;54:249-55.
26. Ridker PM, Cook NR, Lee IM, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med 2005;352:1293-304.
27. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Controlled clinical trials 1998;19:61-109.