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 2014;371:2072-82. 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

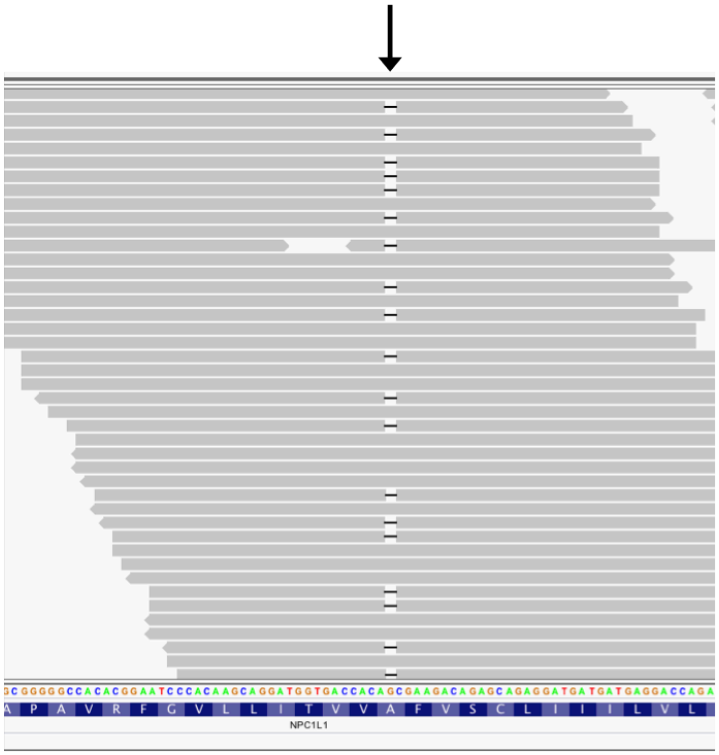
Table S4. Association of NPC1L1 inactivating mutations with risk of coronary heart disease by cohort

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}

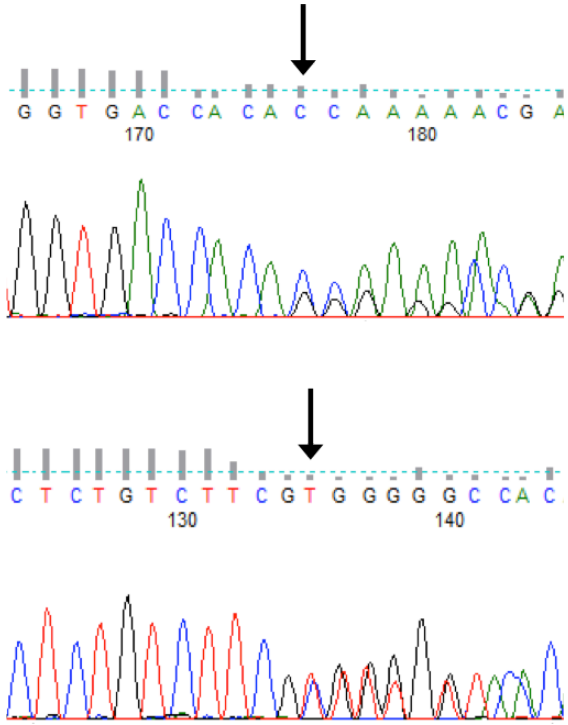
*OR=Odds Ratio; CI=Confidence interval; †Independent samples from sequencing and genotyping combined; ‡Test for heterogeneity P=0.43; AA=African ancestry; EA=European ancestry; SA=South Asian ancestry

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.