SUPPLEMENTAL MATERIAL

Table of Contents

Supplemental Methods

Study Participants

Nine sitosterolemia pedigrees with phytosterol and lipid profiles derived from the Kanazawa University Mendelian Disease (KUMD) Registry were evaluated for *ABCG5* and *ABCG8* variants (**Supplemental Figure**).[1, 2] Sitosterolemia was diagnosed by 1) plasma sitosterol concentration ≥10 μg/mL and 2) presence of tendon or tuberous xanthomas and/or history of premature coronary artery disease (CAD).[3] Causative loss-of-function (LoF) variants in the pedigrees were defined as pathogenic or likely pathogenic by the American College of Medical Genetics standard criteria,[4] and were identified by Sanger sequencing or whole exome sequencing. Controls were unaffected relatives without any causative LoF variants in the sitosterolemia pedigrees.

Next, *ABCG5* and *ABCG8* were sequenced in the Myocardial Infarction Genetics consortium (MIGen), UK Biobank and the TruSeq Custom Amplicon target resequencing studies (TSCA). MIGen case-control studies included total of 48,576 participants from the Italian Atherosclerosis Thrombosis and Vascular Biology study,[5] the Deutsches Herzzentrum München Myocardial Infarction Study,[6] the Exome Sequencing Project Early-Onset Myocardial Infarction study,[7] the Jackson Heart Study,[8] the Leicester Acute Myocardial Infarction Study,[9] the Lubeck Myocardial Infarction Study,[10] the Ottawa Heart Study,[11] the Precocious Coronary Artery Disease study,[12] the Pakistan Risk of Myocardial Infarction Study,[13] and the Registre Gironi del COR study.[14] In addition, 43,223 individuals in UK Biobank were sequenced by the Broad Institute (8,965 individuals) and the Regeneron Genetics Centers (34,258 individuals) as previously described.[15] TSCA included 1,714 participants from the Duke CATHGEN study,[16] the MedStar study,[17] and the PennCath study.[17] An additional 293,134 individuals in UK Biobank underwent array-based genotyping for the *ABCG5* stop variant rs199689137 (p.R446X) and were included in the analysis.[15]

Lipid measurement and coronary artery disease ascertainment

In sitosterolemia pedigree-based analysis, all blood samples were obtained after a 12-hour overnight fast, either before initiation of lipid-lowering treatment or after discontinuation of medication for at least 4 weeks. Plasma levels of non-cholesterol sterols were determined using gas–liquid chromatography–mass spectrometry.[2] In addition to absolute non-cholesterol sterol levels, cholesterol adjusted ratios (each non-cholesterol sterol level to total cholesterol [TC] level ratio) were also evaluated.

In large cohort analysis of MIGen, TSCA, and UK Biobank, plasma concentrations of TC, triglyceride, and high-density lipoprotein cholesterol (HDL-C) were determined using enzymatic

assays. In MIGen and TSCA, low-density lipoprotein cholesterol (LDL-C) level was calculated using the Friedewald equation for those with triglycerides <400 mg/dL. In the UK Biobank, LDL-C levels were directly measured using an antibody-based assay. Also, we took into account the effect of lipid-lowering therapy at the time of lipid measurement, calculating estimated untreated TC and LDL-C by dividing the measured TC and LDL-C levels by 0.8 and 0.7, respectively.[18] We did not adjust HDL-C or triglyceride levels for use of lipid-lowering medications.

In MIGen, CAD cases were identified as early-onset (premature) myocardial infarction (defined as ≤ 50 years old in male and ≤ 60 years old in female). In the UK Biobank, CAD cases were defined as myocardial infarction at any age. In TSCA, coronary artery disease was defined as stenosis on angiography (≥ 1 coronary vessel with > 50% stenosis) at ages ≤ 50 years old for men and ≤ 60 years old for women.

Gene sequencing

Whole exome sequencing or targeted sequencing for KUMD, MIGen, and TSCA was performed at the Broad Institute as previously described.[18, 19] In brief, genomic deoxyribonucleic acid was captured on protein-coding regions using the NimbleGen Sequencing Capture Array or the Illumina TrueSeq Custom Amplicon. Sequencing reads were aligned to a human reference genome (build 37) using the Burrows–Wheeler Aligner-Maximal Exact Match algorithm. Aligned non-duplicate reads were locally realigned, and base qualities were recalibrated using the Genome Analysis ToolKit (GATK) software.[20] Variants were jointly called using the GATK HaplotypeCaller software. Whole exome sequencing in UK Biobank was performed by the Regeneron Genetics Center as previously described.[7]

In large cohort studies, rare LoF variants were defined as those with minor allele frequency less than 0.1% and which caused: (1) insertions or deletions of DNA that modified the reading frame of protein translation (frameshift); (2) point mutations at conserved splice site regions that altered the splicing process (splice-site); or (3) point mutations that changed an amino acid codon to a stop codon, leading to the truncation of a protein (nonsense). LoF variants were identified using the LOFTEE plugin of the Variant Effect Predictor software (version 82).[21]

Statistical Analysis

In sitosterolemia family-based analyses, the differences in lipid levels, lipoproteins and non-cholesterol sterols by *ABCG5* or *ABCG8* LoF variant carrier status were assessed by linear regression adjusted by kinship matrix within each family using the log-transformed values. The effects of *ABCG5* and *ABCG8* rare LoF variants on lipid profiles in MIGen, UK Biobank and TSCA was evaluated using linear regression, adjusting for age, gender, study, and first five principal components of ancestry. A

Cochran–Mantel–Haenszel statistic meta-analysis for stratified 2-by-2 tables was used to associate *ABCG5* and *ABCG8* rare LoF variants with risk of CAD. As a sensitivity analysis, we performed an inverse variance weighted fixed effects meta-analysis of the adjusted odds ratio, derived in each cohort using logistic regression adjusted for age, sex, study and five principal components of ancestry. In an exploratory analysis, we assessed whether the effects of *ABCG5 or ABCG8* rare LoF variants on LDL-C could explain their association with CAD risk. We derived a dose-response line between genetic LDL-C level and CAD risk by estimating the effect of LoF variants in 2 LDL-C related pathway genes (*PCSK9* and *LDLR*) on both LDL-C and CAD risk. The results of *PCSK9* and *LDLR* were derived from samples in MIGen and UK Biobank whole exome sequencing datasets. P values of less than 0.025 were considered to indicate statistical significance (i.e., Bonferroni correction for the testing of two genes). Statistical analyses were performed using R software version 3.4.3 (The R Project for Statistical Computing, Vienna, Austria).

Supplemental Table 1. List of *ABCG5* and *ABCG8* rare LoF variants from sitosterolemia family analysis.

Variant (GRCh37)	Consequence	Family #	ACMG guideline classification
ABCG5			
2: 44040449_C/T	Splice acceptor	6	Pathogenic (PVS1+PS3)
2: 44041700_TAAAAG/T	Frameshift: P558fsTer	1,3	Pathogenic (PVS1+PS3)
2: 44050063_G/A	Premature stop: R446X	3	Pathogenic (PVS1+PS3)
2: 44051120_C/T	Missense: R419H	6	Pathogenic (PS1+PS3)
2: 44051210_C/T	Missense: R389H	1,2,5,7	Pathogenic (PS1+PS3)
2: 44052028_T/C	Missense: M302V	5	Pathogenic (PS3+PM2+PM5+PP1+PP4)
2: 44053544_G/A	Premature stop: Q251X	8	Pathogenic (PVS1+PS3)
2: 44065689_A/C	Missense: S44A	2	Pathogenic (PS3+PM2+PM5+PP1+PP4)
ABCG8			
2:44100970_TC/AA	Missense: I419K	4	Pathogenic (PS3+PM2+PM5+PP1+PP4)
2:44100970_T/A	Missense: I419N	9	Pathogenic (PS3+PM2+PM5+PP3+PP4)
2:44100999_A/G	Missense: M429V	4	Likely pathogenic (PS3+PM5+PP1+PP4)

Supplemental Table 2. List of rare LoF variants identified in *ABCG5*.

Variant	Consequence	MIGen, n	UK Biobank, n	TSCA, n
2:44040325_G/C	Premature stop: S629X	0	10	0
2:44040325_G/T	Premature stop: E629X	0	4	0
2:44041643_C/A	Premature stop: E579X	1	0	1
2:44041730_T/C	Splice Acceptor	0	1	0
2:44041730_T/A	Splice Acceptor	2	0	0
2:44047174_T/TG	Frameshift: H510fsTer	0	0	1
2:44047211_G/A	Premature stop: R498X	1	2	0
2:44047240_C/T	Splice Acceptor	0	1	0
2:44047241_T/C	Splice Acceptor	1	0	0
2:44049953_GA/G	Frameshift: F482fsTer	1	0	0
2:44050063_G/A	Premature stop: R446X	14	6	0
2:44050075_C/A	Splice Acceptor	0	2	0
2:44050076_T/C	Splice Acceptor	0	1	0
2:44051036_GCACGTGGG CACTTACA/G	Splice Donor	4	0	0
2:44051051_C/A	Splice Donor	1	0	0
2:44051357_C/G	Splice Donor	1	0	0
2:44051431_T/A	Premature stop: K349X	1	0	0
2:44052027_C/T	Splice Donor	0	1	0
2:44052071_G/T	Premature stop: C287X	2	0	0
2: 44052096_TC/T	Frameshift: Q279fsTer	0	0	1
2:44052158_C/T	Splice Acceptor	1	9	0
2:44053520_C/T	Splice Donor	0	1	0
2:44053544_G/A	Premature stop: Q251X	0	1	0
2:44053568_G/A	Premature stop: R243X	3	0	0
2:44055121_C/G	Splice Donor	1	0	0
2:44055121_C/T	Splice Donor	1	0	0
2:44055180_GC/G	Frameshift: G192fsTer	1	0	0
2:44055209_G/A	Premature stop: R183X	3	2	0
2:44058973_C/A	Premature stop: E146X	1	3	0
2:44059095_G/T	Premature stop: Y131X	1	0	0
2:44059154_C/A	Premature stop: E131X	0	1	0
2:44059166_TC/T	Frameshift: G107Ter	4	0	0
2:44059223_C/T	Splice Acceptor	0	1	0
2:44064975_G/C	Premature stop: S88X	1	2	0

2:44065009_C/A	Premature stop: E77X	1	0	0
2:44065013_G/C	Premature stop: Y75X	0	2	0
2:44065051_G/A	Premature stop: Q63X	1	1	0
2:44065755_G/A	Premature stop: Y45X	5	0	0
2:44065684_G/C	Premature stop: Q22X	0	2	0
Total		53	53	3

Supplemental Table 3. List of rare LoF variants identified in *ABCG8*.

Variant	Consequence	MIGen, n	UK Biobank, n	TSCA, n
2: 44071644_A/G	Splice Acceptor	0	1	0
2:44071744_C/A	Premature stop: Y54X	1	0	0
2:44073393_C/T	Premature stop: Q89X	1	0	0
2:44073448_C/G	Premature stop: S107X	1	0	0
2: 44078866_C/T	Premature stop: Q156X	0	1	0
2:44078890_C/T	Premature stop: R164X	0	1	0
2:44078962_G/C	Splice Donor	1	0	0
2:44079924_T/G	Premature stop: L294X	1	0	0
2:44099233_G/A	Premature stop: W361X	38	61	2
2:44099360_A/G	Splice Acceptor	0	1	0
2:44100948_C/T	Premature stop: R412X	11	5	0
2:44101038_C/T	Premature stop: Q442X	1	0	0
2:44101552_CAGAG/C	Frameshift: S473Ter	1	0	0
2:44101577_AC/A	Frameshift: L482Ter	1	0	0
2: 44101595_C/A	Premature stop: Y487X	0	1	0
2:44101610_T/A	Premature stop: Y492X	4	0	1
2:44102474_TC/T	Frameshift: S560Ter	2	0	0
2:44102515_CG/C	Frameshift: G574Ter	2	0	0
2:44102548_G/A	Premature stop: W584X	1	0	0
2:44102554_T/A	Splice Donor	1	0	0
2:44104934_TG/T	Frameshift: L635Ter	1	0	0
Total		68	71	3

Supplemental Table 4. Associations of *ABCG8* LoF variants with blood lipids and CAD risk among three WES cohorts.

Trait	LoF variant carriers (N)	Total N	Effect (95% CI), P value
Total cholesterol	135	93049	Beta 0.03 (-0.12 to 0.17), P = 0.71
LDL cholesterol	135	93049	Beta 0.06 (-0.09 to 0.22), P = 0.47
HDL cholesterol	135	93049	Beta -0.08 (-0.22 to 0.07), P = 0.31
Triglycerides	135	93049	Beta 0.03 (-0.08 to 0.14), P = 0.61
CAD	142	82915	OR 0.79 (0.47 to 1.31), P = 0.36

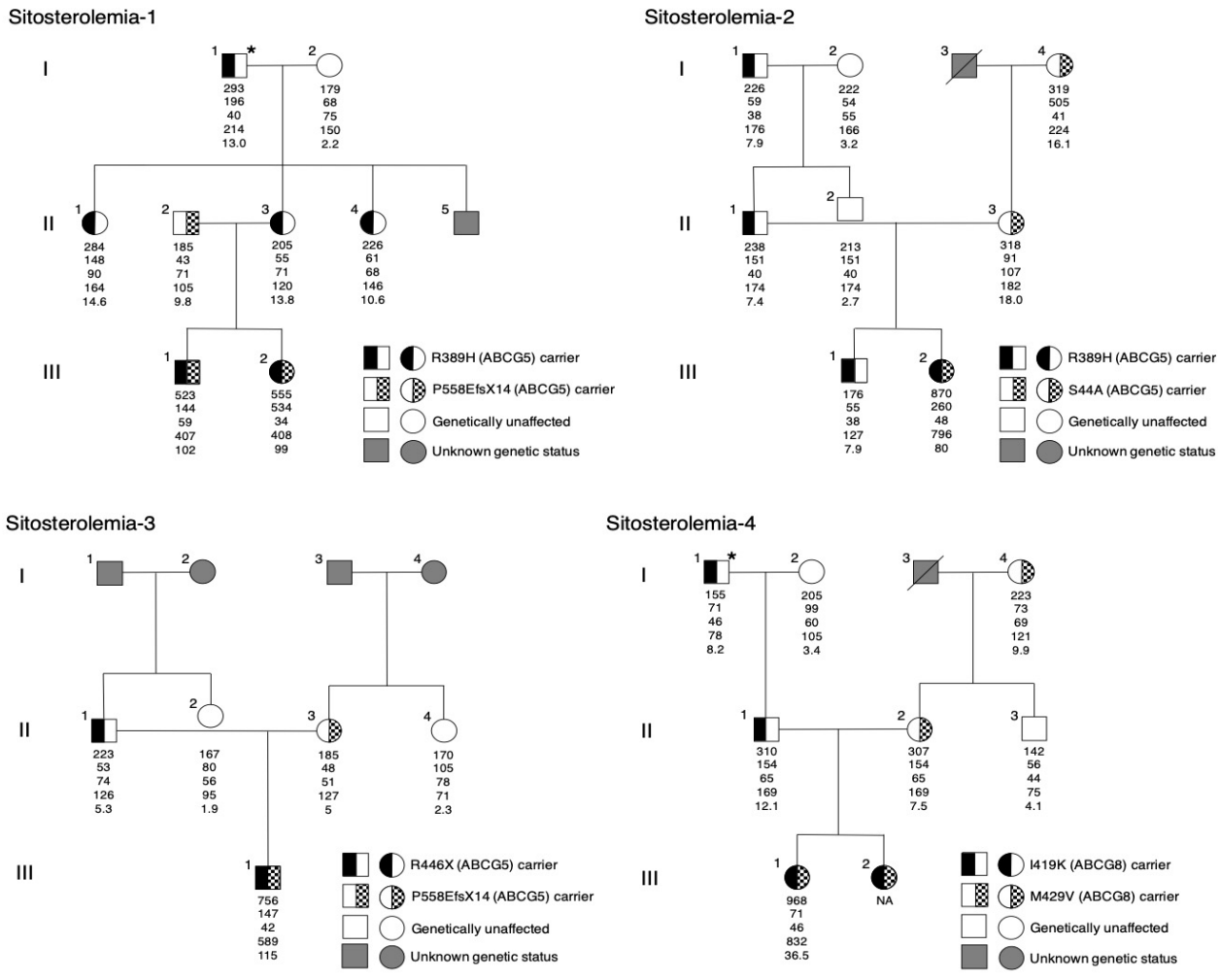
Estimates include exome sequencing data from the Myocardial Infarction Genetics consortium and UK Biobank.
Abbreviations: CAD, coronary artery disease; CI, 95% confidence interval; HDL, high-density-lipoprotein; LDL, low-density lipoprotein; LoF, loss-of-function; OR, odds ratio; SD, standard deviation; WES, whole-exome-sequencing.

Supplemental Table 5. Associations of *ABCG5, ABCG8, PCKS9* and *LDLR* LoF variants with LDL-C and CAD risk.

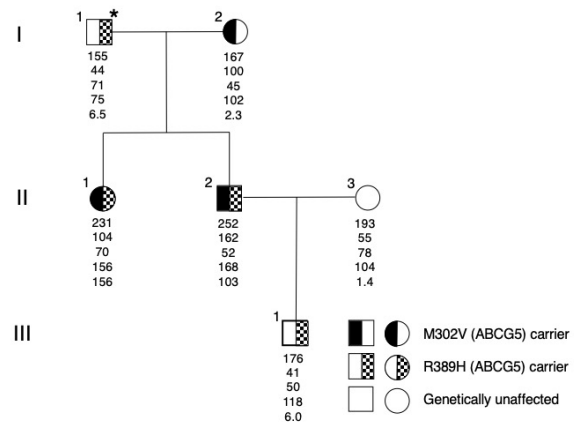
Gene	LoF carriers in CAD cases	LoF carriers in CAD-free controls	Effect on LDL-C (95% CI), P value	Effect on CAD risk (95% CI), P value
ABCG5	34 / 29,321	63 / 357,326	Beta 0.42 (0.24 to 0.59) $P = 1.1x10^{-6}$	OR 2.06 (1.27 to 3.35) P = 0.004
ABCG8	20 / 18,309	116 / 66,320	Beta 0.06 (-0.09 to 0.22) P = 0.47	OR 0.79 (0.47 to 1.31) P = 0.36
LDLR	47 / 17,125	19 / 65,795	Beta 2.87 (2.64 to 3.08) $P = 1.2x10^{-144}$	OR 5.36 (2.95 to 9.75) $P = 3.8x10^{-8}$
PCSK9	4 / 17,125	89 / 65,795	Beta -0.86 (-1.02 to 0.70) $P = 1.5x10^{-25}$	OR 0.33 (0.12 to 0.93) P = 0.04

Estimates include exome sequencing data from the Myocardial Infarction Genetics consortium and UK Biobank. Abbreviations: same as Supplemental Table 4.

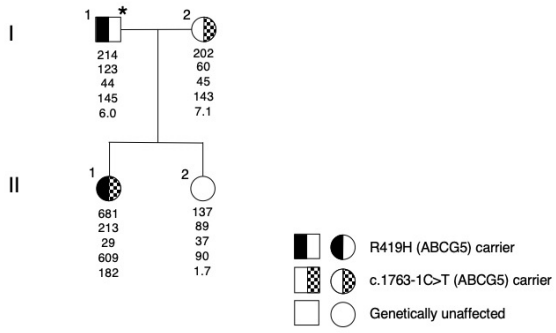
Supplemental Figure. Pedigrees of sitosterolemia families.
Square and circle indicate male and female, respectively. Black or lattice pattern full half indicates heterozygote subjects; black, homozygote subjects; gray shading, genetically unknown subjects; and white, genetically unaffected subjects. Total cholesterol (mg/dL), triglycerides (mg/dL), high-density lipoprotein cholesterol (mg/dL), low-density lipoprotein cholesterol (mg/dL), and sitosterol (μg/mL) levels are displayed below each individual identifier. Patients who had a history of coronary artery disease are indicated with an asterisk "*".



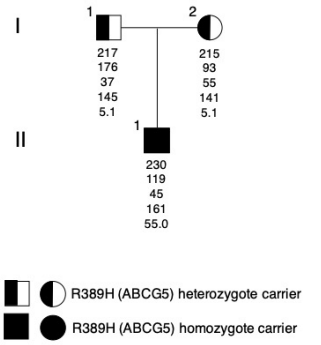
Sitosterolemia-5



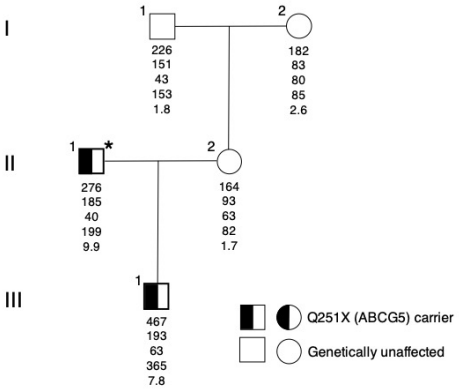
Sitosterolemia-6



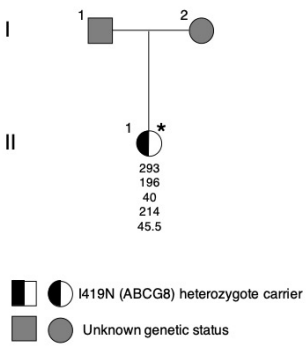
Sitosterolemia-7



Sitosterolemia-8



Sitosterolemia-9



Supplemental References

1. Katayama T, Satoh T, Yagi T, Hirose N, Kurita Y, Anzai T, Asakura Y, Yoshikawa T, Mitamura H, Ogawa S. A 19-year-old man with myocardial infarction and sitosterolemia. *Intern Med*. 2003;42:591-4.

2. Tada H, Nohara A, Inazu A, Sakuma N, Mabuchi H, Kawashiri MA. Sitosterolemia, Hypercholesterolemia, and Coronary Artery Disease. *J Atheroscler Thromb*. 2018;25:783-789.

3. Merkens LS, Myrie SB, Steiner RD, Mymin D. Sitosterolemia. 2013;2019.

4. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405-24.

5. 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.

6. Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med*. 2014;371:2072-82.

7. Do R, Stitziel NO, Won HH, Jorgensen AB, Duga S, Angelica Merlini P, Kiezun A, Farrall M, Goel A, Zuk O, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2015;518:102-6.

8. Taylor HA, Jr. The Jackson Heart Study: an overview. *Ethn Dis*. 2005;15:S6-1-3.

9. Dhillon OS, Khan SQ, Narayan HK, Ng KH, Struck J, Quinn PA, Morgenthaler NG, Squire IB, Davies JE, Bergmann A, et al. Prognostic value of mid-regional pro-adrenomedullin levels taken on admission and discharge in non-ST-elevation myocardial infarction: the LAMP (Leicester Acute Myocardial Infarction Peptide) II study. *J Am Coll Cardiol*. 2010;56:125-33.

10. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortica Investigators. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med*. 2016;374:1134-44.

11. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488-91.

12. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med*. 2009;361:2518-28.

13. Saleheen D, Zaidi M, Rasheed A, Ahmad U, Hakeem A, Murtaza M, Kayani W, Faruqui A, Kundi A, Zaman KS, 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.

14. Senti M, Tomas M, Marrugat J, Elosua R, REGICOR Investigators. Paraoxonase1-192 polymorphism modulates the nonfatal myocardial infarction risk associated with decreased HDLs. *Arterioscler Thromb Vasc Biol*. 2001;21:415-20.

15. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203-209.

16. Davies RW, Wells GA, Stewart AF, Erdmann J, Shah SH, Ferguson JF, Hall AS, Anand SS, Burnett MS, Epstein SE, et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet*. 2012;5:217-25.

17. Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, Burnett MS, Devaney JM, Knouff CW, Thompson JR, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet*. 2011;377:383-92.

18. Nomura A, Won HH, Khera AV, Takeuchi F, Ito K, McCarthy S, Emdin CA, Klarin D, Natarajan P, Zekavat SM, et al. Protein-Truncating Variants at the Cholesteryl Ester Transfer Protein Gene and Risk for Coronary Heart Disease. *Circ Res*. 2017;121:81-88.

19. Peloso GM, Nomura A, Khera AV, Chaffin M, Won HH, Ardissino D, Danesh J, Schunkert H, Wilson JG, Samani N, et al. Rare Protein-Truncating Variants in APOB, Lower Low-Density Lipoprotein Cholesterol, and Protection Against Coronary Heart Disease. *Circ Genom Precis Med*. 2019;12:e002376.

20. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-8.

21. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26:2069-70.