# Association of Low-Frequency and Rare Coding-Sequence Variants with Blood Lipids and Coronary Heart Disease in 56,000 Whites and Blacks

Gina M. Peloso,[1,2,3,4] Paul L. Auer,[5,6] Joshua C. Bis,[7] Arend Voorman,[8] Alanna C. Morrison,[9] Nathan O. Stitziel,[10,11] Jennifer A. Brody,[7] Sumeet A. Khetarpal,[12] Jacy R. Crosby,[9,13] Myriam Fornage,[9,14] Aaron Isaacs,[15] Johanna Jakobsdottir,[16] Mary F. Feitosa,[11] Gail Davies,[17,18] Jennifer E. Huffman,[19] Ani Manichaikul,[20] Brian Davis,[9] Kurt Lohman,[21] Aron Y. Joon,[14] Albert V. Smith,[16,22] Megan L. Grove,[9] Paolo Zanoni,[12] Valeska Redon,[12] Serkalem Demissie,[23,24] Kim Lawson,[9] Ulrike Peters,[5] Christopher Carlson,[5] Rebecca D. Jackson,[25] Kelli K. Ryckman,[26] Rachel H. Mackey,[27] Jennifer G. Robinson,[26] David S. Siscovick,[7,28] Pamela J. Schreiner,[29] Josyf C. Mychaleckyj,[20] James S. Pankow,[29] Albert Hofman,[30] Andre G. Uitterlinden,[30] Tamara B. Harris,[31] Kent D. Taylor,[32] Jeanette M. Stafford,[21] Lindsay M. Reynolds,[21] Riccardo E. Marioni,[17,18] Abbas Dehghan,[30] Oscar H. Franco,[30] Aniruddh P. Patel,[1,4,33] Yingchang Lu,[34,35] George Hindy,[36] Omri Gottesman,[34] Erwin P. Bottinger,[34] Olle Melander,[36] Marju Orho-Melander,[37] Ruth J.F. Loos,[34,35,38] Stefano Duga,[39] Piera Angelica Merlini,[40,41] Martin Farrall,[42] Anuj Goel,[42] Rosanna Asselta,[39] Domenico Girelli,[43] Nicola Martinelli,[43] Svati H. Shah,[44,45] William E. Kraus,[45,46] Mingyao Li,[47] Daniel J. Rader,[12] Muredach P. Reilly,[12] Ruth McPherson,[48] Hugh Watkins,[42,49] Diego Ardissino,[40,50] NHLBI GO Exome Sequencing Project, Qunyuan Zhang,[11] Judy Wang,[11] Michael Y. Tsai,[51] Herman A. Taylor,[52,53,54] Adolfo Correa,[54] Michael E. Griswold,[54] Leslie A. Lange,[55] John M. Starr,[17,56] Igor Rudan,[57,58] Gudny Eiriksdottir,[16] Lenore J. Launer,[31] Jose M. Ordovas,[59,60,61] Daniel Levy,[24,62] Y.-D. Ida Chen,[32] Alexander P. Reiner,[5,28] Caroline Hayward,[19] Ozren Polasek,[63] Ian J. Deary,[17,18] Ingrid B. Borecki,[11] Yongmei Liu,[21] Vilmundur Gudnason,[16,22] James G. Wilson,[64] Cornelia M. van Duijn,[15] Charles Kooperberg,[5] Stephen S. Rich,[20] Bruce M. Psaty,[7,28,65,66] Jerome I. Rotter,[32] Christopher J. O'Donnell,[3,24,62,67] Kenneth Rice,[8,69] Eric Boerwinkle,[9,68,69] Sekar Kathiresan,[1,2,3,4,67,69,*] and L. Adrienne Cupples[23,24,69,*]

[1]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; [2]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA; [3]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; [4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA; [5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [6]School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA; [7]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA; [8]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [9]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [10]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA; [11]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA; [12]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [13]Department of Biostatistics, Bioinformatics, and Systems Biology, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, TX 77030, USA; [14]Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [15]Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3015CN, the Netherlands; [16]Icelandic Heart Association, Kopavogur 201, Iceland; [17]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK; [18]Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK; [19]MRC Human Genetics, MRC IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK; [20]Center for Public Health Genomics and Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA; [21]Wake Forest School of Medicine, Winston-Salem, NC 27106, USA; [22]Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland; [23]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; [24]National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, MA 01702, USA; [25]Division of Endocrinology, Diabetes and Metabolism, Department of Internal Medicine, The Ohio State University, Columbus, OH 43210, USA; [26]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA 52242, USA; [27]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [28]Department of Epidemiology, University of Washington, Seattle, WA 98195, USA; [29]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55454, USA; [30]Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3015CN, the Netherlands; [31]National Institute on Aging, NIH, Bethesda, MD 20892, USA; [32]Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [33]School of Medicine, Yale University, New Haven, CT 06510, USA; [34]The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [35]The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [36]Department of Clinical Sciences in Malmö, Lund University, Clinical Research Center, Malmö 20502, Sweden; [37]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö 20502, Sweden; [38]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [39]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milano 20133, Italy; [40]Associazione per lo Studio della Trombosi in Cardiologia (ASTC), Pavia 27100, Italy; [41]Divisione di Cardiologia, Ospedale Niguarda, Milano 20162, Italy; [42]Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; [43]Department of Medicine, University of Verona School of Medicine, Verona 37134, Italy; [44]Department of Obstetrics and Gynecology, Division of Urogynecology, Duke University, Durham, NC 27710, USA; [45]Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC 27710, USA; [46]Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC 27701, USA; [47]Department of Biostatistics and Epidemiology, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [48]Atherogenomics Laboratory, Division of Cardiology, University of Ottawa Heart Institute, Ottawa,

Low-frequency coding DNA sequence variants in the proprotein convertase subtilisin/kexin type 9 gene (*PCSK9*) lower plasma low-density lipoprotein cholesterol (LDL-C), protect against risk of coronary heart disease (CHD), and have prompted the development of a new class of therapeutics. It is uncertain whether the *PCSK9* example represents a paradigm or an isolated exception. We used the "Exome Array" to genotype >200,000 low-frequency and rare coding sequence variants across the genome in 56,538 individuals (42,208 European ancestry [EA] and 14,330 African ancestry [AA]) and tested these variants for association with LDL-C, high-density lipoprotein cholesterol (HDL-C), and triglycerides. Although we did not identify new genes associated with LDL-C, we did identify four low-frequency (frequencies between 0.1% and 2%) variants (*ANGPTL8* rs145464906 [c.361C>T; p.Gln121*], *PAFAH1B2* rs186808413 [c.482C>T; p.Ser161Leu], *COL18A1* rs114139997 [c.331G>A; p.Gly111Arg], and *PCSK7* rs142953140 [c.1511G>A; p.Arg504His]) with large effects on HDL-C and/or triglycerides. None of these four variants was associated with risk for CHD, suggesting that examples of low-frequency coding variants with robust effects on *both* lipids and CHD will be limited.

## Introduction

Recently, a compelling new therapeutic target for lowering low-density lipoprotein cholesterol (LDL-C) emerged from human genetics: the proprotein convertase subtilisin/kexin type 9 gene (*PCSK9* [MIM 607786]).[1,2] *PCSK9* protein-coding sequence variants that are low in frequency (defined here as allele frequencies from 0.1% to 5%) have been associated with lower plasma LDL-C[3] and reduced risk for coronary heart disease (CHD).[4] With the identification of low-frequency variants that protected against CHD, many pharmaceutical companies have established drug development programs targeting PCSK9. These observations have raised the question of whether the *PCSK9* example is a paradigm for complex diseases like CHD or an exception.[5]

Low-frequency DNA sequence variants and rare alleles (defined here as <0.1% allele frequency) are poorly characterized by earlier generations of genome-wide association study (GWAS) genotyping arrays.[6] Sequencing across the exome or genome can directly assay low-frequency and rare variants but such an approach is currently too costly to study tens of thousands of individuals. One proposed method for testing low-frequency and rare DNA variation is to first sequence the exome to discover variation and, subsequently, to genotype the discovered variants in a larger number of individuals from the same or similar populations to test for association with phenotype. Based on this principle, the Illumina HumanExome genotyping array ("the Exome Array" or "Exome Chip") was designed based on coding sequence variants discovered from sequencing the exomes of ~12,000 individuals.

Here, we set out to address two questions: (1) are there novel low-frequency nonsynonymous and splice site variants associated with lipid levels in the population? and (2) if so, will these coding sequence variants also associate with risk of clinical CHD? To address these questions, we first genotyped and analyzed the Exome Array in 42,208 European ancestry (EA) individuals and 14,330 African ancestry (AA) individuals from 13 cohorts with blood levels of fasting low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and fasting triglycerides (TG). We subsequently tested validated lipid variants for association with CHD. We discovered four variants associated with HDL-C and/or TG, but these variants did not associate with CHD.

## Subjects and Methods

### Estimation of Proportion of Variation Captured by Exome Array

We identified EA (n = 3,173) and AA (n = 2,408) participants from the Atherosclerosis Risk in Communities (ARIC) Study with available exome sequence who were not among the ~12,000 individuals utilized to design the Exome Array. Separately by ancestry, we identified all variants with a minor allele frequency (MAF) >0.1% from exome sequencing of these independent ARIC individuals. These variants were compared with the variants available on the Illumina HumanExome v.1.0 array.

### Study Participants

Thirteen studies genotyped the Exome Array on a total of 56,538 participants (Table S1 available online). 42,208 individuals were of European ancestry from ARIC, AGES, CHS, FHS, RS, MESA, WHI, CARDIA, Health ABC, FamHS, LBC1936, and Korcula. A total of 14,330 subjects were of African ancestry from ARIC, CHS, MESA, Health ABC, FamHS, JHS, WHI, and CARDIA. All participants provided informed consent and each study was approved by their governing ethics committee.

ON K1Y 4W7, Canada; [49]Merck Sharp & Dohme Corp., Rahway, NJ 07065, USA; [50]Divisione di Cardiologia, Azienda Ospedaliero-Universitaria di Parma, Parma 43100, Italy; [51]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55454, USA; [52]Jackson State University, Jackson, MS 39217, USA; [53]Tougaloo College, Tougaloo, MS 39174, USA; [54]Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; [55]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; [56]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK; [57]Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh EH8 9AG, UK; [58]Croatian Centre for Global Health, Faculty of Medicine, University of Split, Split 21000, Croatia; [59]Department of Cardiovascular Epidemiology and Population Genetics, National Center for Cardiovascular Investigation, Madrid 28049, Spain; [60]IMDEA-Alimentacion, Madrid 28049, Spain; [61]Nutrition and Genomics Laboratory, Jean Mayer-USDA Human Nutrition Research Center on Aging at Tufts University, Medford, MA 02155, USA; [62]Division of Intramural Research, NHLBI, NIH, Bethesda, MD 20892, USA; [63]Department of Public Health, Faculty of Medicine, University of Split, Split 21000, Croatia; [64]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; [65]Group Health Research Institute, Group Health Cooperative, Seattle, WA 98101, USA; [66]Department of Health Services, University of Washington, Seattle, WA 98195, USA; [67]Cardiology Division, Massachusetts General Hospital, Boston, MA 02114, USA; [68]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [69]These authors contributed equally to this work
*Correspondence: skathiresan@partners.org (S.K.), adrienne@bu.edu (L.A.C.)

## Genotyping and Quality Control

All study participants were genotyped on the HumanExome Bead-Chip v.1.0 (Illumina). ARIC, AGES, CHS, FHS, RS, MESA, CARDIA, Health ABC, FamHS, and JHS were jointly called[7,8] at the University of Texas Health Science Center at Houston. LBC1936 and Korcula were called in GenomeStudio (Illumina) via the CHARGE Consortium joint calling cluster file. WHI used zCalls.[9] Variant quality control (QC) was performed at the University of Texas Health Science Center and by the individual cohorts. QC involved checking concordance to GWAS data and excluding those individuals missing >5% genotypes, population clustering outliers, individuals with high inbreeding coefficients or heterozygote rates, individuals with gender mismatches, one individual from duplicate pairs, and individuals with an unexpectedly high proportion of identity-by-descent sharing, with consideration for family studies, based on high-quality variants. All contributing studies used an additive coding of variants to the minor allele observed in the jointly called data set.

## Phenotypes

Association analyses were performed for three blood lipid traits in mg/dl: LDL-C, HDL-C, and TG. Fasting lipids values were used from the earliest available exam in each study. Studies with only nonfasting lipid levels contributed only to the HDL-C analysis. To account for lipid-lowering therapy in selected participants, we sought to estimate the untreated lipid value in each participant. Such an approach has been demonstrated to perform well in accounting for treatment effects in studies of quantitative traits.[10] Statins are the most widely used treatment to lower plasma lipids and a statin at average dose reduces total cholesterol (TC) by 20% and LDL-C by 30%.[11] Statins became routinely used after the publication of the seminal 4S randomized control trial in 1994.[12]

If the sample was collected after 1994, we accounted for lipid-lowering medication in the following manner: the treated TC value was divided by 0.8. No adjustment was done on data collected before 1994 unless specific information on statin use was available. LDL-C was calculated via the Friedewald equation[13,14] (LDL-C = TC − HDL-C − (TG/5)) for those with TG <400 mg/dl. If TG >400 mg/dl, calculated LDL-C was set to missing. If TC was modified as described above for medication use, the modified total cholesterol was used to calculate LDL-C. If only measured LDL-C was available in a study, we accounted for lipid-lowering medication in the following manner: the treated LDL-C value was divided by 0.7. No adjustments for medication use were made for HDL-C and TG. TG was natural log transformed for analysis.

## Statistical Analyses

Single-variant, conditional, and gene-based analyses were conducted with the seqMeta package in R. In brief, seqMeta implements efficient single-variant and gene-based meta-analysis. Each study computes association-specific statistics for each variant and trait assuming an additive model, and genotypic covariance matrices within predefined gene regions. Score statistics and genotypic covariance matrices are combined across studies to obtain an overall score statistic for each variant and a combined covariance matrix across studies. The combined score statistic is used for the single-variant results. For each gene, the score statistics for the variants within the gene along with the combined covariance matrix across those variants are used to construct the gene-based tests. An additive effect of each variant is assumed for individuals carrying more than one rare variant. All analyses were performed separately by race. We chose to meta-analyze EA and AA participants separately to minimize population biases and to increase the number of variants at specific MAF thresholds where there could be vastly different frequencies in the two ancestries.

### Single-Variant Analysis

Though we defined low-frequency variants as having a MAF between 0.1% and 5%, we restricted single-variant tests to nonsynonymous and splice site variants with a MAF >0.02% in EA individuals ($n_{variants}$ = 54,837) and >0.07% in AA individuals ($n_{variants}$ = 90,218), corresponding to approximately 20 copies of the minor allele in each ancestry sample. Twenty copies of the minor allele provided stable estimates of the standard errors of the beta estimates and well-calibrated quantile-quantile plots.

### Gene-Based Analysis

We constructed gene-based tests from the cohort-combined variant statistics and LD matrices. We only included nonsynonymous and splice site variants when producing the gene-based statistics. We performed three gene-based tests: (1) T1, where all variants with a MAF <1% were collapsed into gene-based score;[15] (2) T0.1, where all variants with a MAF <0.1% were collapsed into gene-based score; and (3) sequence kernel association test (SKAT)[16] at a MAF <5%. SKAT is more powerful than T1 when there are both protective and deleterious variants with different magnitudes in the same gene whereas T1 and T0.1 are more powerful when the magnitudes and directions of effect for the individual variants in a gene are consistent. We filtered gene-based results with a cumulative MAF <0.05% in the EA individuals and <0.14% in the AA individuals for SKAT analyses, corresponding to approximately 40 copies of the minor allele. For the T1 and T0.1 tests, we used a cumulative MAF >0.04% for the EA individuals and >0.11%, corresponding to 30 copies of the minor allele across all cohorts. Simulations under the null (i.e., no associations) were performed to determine the thresholds to control type I error for the burden tests (results not shown). The gene-based T1 and T0.1 analyses were limited to genes with at least two variants contributing to the test. For EA, T1 analysis included 12,351 genes with a cumulative MAF >0.04% and T0.1 with 8,357 genes; for AA participants, T1 analysis included 13,574 genes with a cumulative MAF >0.11% and T0.1 with 7,579 genes.

All analyses included age, age², sex, and up to the first four principal components of ancestry as covariates. Cohort statistics were adjusted for related individuals, where appropriate. Beta coefficients were checked for consistency across the individual studies contributing to the meta-analysis. Hardy-Weinberg equilibrium p values were calculated and cluster plots from the joint calling were verified for the reported associations (Figure S1). Fasting glucose at baseline (between 1987 and 1989) in ARIC EA and exam 5 (between 1991 and 1995) in FHS was tested for association with rs145464906 in *ANGPTL8* by linear regression adjusting for age, sex, and PCs.

### Population Structure

Each study identified components to correct for population structure in their study and used these components as covariates in all association analyses.

### Conditional Analysis

The allele dosages (0–2 copies) of the top previously identified GWAS variants in each region were used as additional covariates in the association model.

### Statistical Significance

For single-variant association, we set the significance threshold to <3 × $10^{-8}$, corresponding to a Bonferroni correction for 1,485,864 tests (3 phenotypes × 247,644 variants on the

array × 2 ancestries). For the gene-based association, we set the significance threshold to $<2 \times 10^{-7}$, corresponding to a Bonferroni correction for 316,332 tests (3 phenotypes × 17,574 genes on the array × 3 tests × 2 ancestries).

*Annotation*
We used dbNSFP v.2.0 to annotate the variants.[17]

## Clinical CHD Association Analysis

Individuals from AGES, ARIC, CHS, FHS, MESA, Health ABC, RS1, WHI, MDC-CVA, IPM, ATVB/VHS, Ottawa, Procardis, Duke, and Penn contributed to the CHD analysis in EA. Individuals from ARIC, CHS, MESA, Health ABC, WHI, and IPM contributed to the CHD analysis in AA. All individuals were genotyped on the HumanExome BeadChip. Counts of number of CHD events in carriers and noncarriers of the four reported lipid variants were obtained from each available study. A Cochran-Mantel-Haenszel test was performed to associate the four reported lipid variants with CHD by using the R metafor package across the contributing studies.

## Power Calculation

Power for the CHD association was calculated via the case-control for discrete traits option in the Genetic Power Calculator.

## Mouse Studies Involving PCSK7

We created an adeno-associated virus 8 (AAV8) vector encoding the human *PCSK7* coding region (cDNA) driven by the liver-specific thyroxine-binding globulin (TBG) promoter. An AAV8 vector lacking any transgene was used as a control. Separate groups of 10- to 12-week-old wild-type C57BL/6J male mice (six to seven per group) were injected via the peritoneal route with $1 \times 10^{12}$ vector genomes/mouse of the relevant vectors. Mice were fed a regular chow diet (Purina LabDiet 5010) throughout the study. Plasma samples after 4 hr of fasting were taken from the mice immediately prior to vector administration, as well as 7 and 14 days after AAV injection for analysis of lipids. Lipid measurements were performed on a Roche Cobas-Mira autoanalyzer (Roche Diagnostic Systems) with Wako Chemicals reagents. Mice were sacrificed at 14 days after AAV administration. All animal procedures were performed according to the regulations of, and with the prior approval of, the University of Pennsylvania Animal Care and Use Committee (IACUC). We repeated this experiment in a separate design comparing plasma lipids at baseline and 14 days after injection of an AAV8 vector encoding the mouse *Pcsk7* cDNA versus control in wild-type C57BL/6J male mice (data not shown).

## Results

### Coverage of Exome Array

We evaluated the coverage of Exome Array in samples independent from those used to design the array. Based on exome sequences from 3,173 EA and 2,408 AA independent participants, we estimate that the Illumina Exome Array captures 78% of nonsynonymous coding and splice site variation with >1:1,000 allele frequency in EA individuals and 71% in AA individuals. Therefore, the array provides good coverage for low-frequency DNA variation and because of its low cost can be assayed in a large number of individuals.

Exome-sequencing variants not captured by the Exome Array may have been dropped during the design of the array. To address this possibility, we compared the variants discovered from exome sequencing to proposed content of the array (before variants failed design). We estimate that 95% of the variation in EA participants (and 85% of the variation in AA) with frequency >1:1,000 is captured by the proposed content of the array.

### Study Participants and Genotypes

Clinical characteristics of the 56,538 analyzed study subjects are summarized in Table S1. Of the 247,644 variants that passed quality control criteria, 85% (n = 209,756) are polymorphic in the EA participants and 82% (n = 202,255) are polymorphic in the AA participants. Approximately 90% (n = 188,480) of the polymorphic variants are annotated[17] as nonsynonymous (nonsense or missense) or splice variants in the EA participants, and 89% (n = 180,908) of the polymorphic variants are annotated as such in the AA participants (Table 1).

### Single-Variant Association

We tested each variant individually for association with blood lipid levels separately in each cohort and association summary statistics across cohorts were combined by fixed effects meta-analysis, separately within each ancestry group. There was no evidence of inflation in the association test statistics (Figure S2).

We replicated previously reported associations with common and low-frequency variants (Table 2) in both EA and AA participants. For example, we found that variants in *PCSK9* (c.137G>T [p.Arg46Leu], [rs11591147] in EA and c.2037C>A [p.Cys679*], [rs28362286] in AA) were associated with lower LDL-C ($-17$ mg/dl; p = $3 \times 10^{-59}$ and $-40$ mg/dl; p = $2 \times 10^{-57}$, respectively).

We discovered four associations of low-frequency variants with either HDL-C and/or TG that met our a priori significance threshold of p < $3 \times 10^{-8}$ and not previously reported in the literature. Two signals emerged from EA participants and two from AA participants (Table 3). In contrast to HDL-C and TG, we did not discover any new genes where low-frequency or rare DNA sequence variants significantly associated with LDL-C. Below, we describe each of the four new associations.

First, in EA participants, we found an association of blood HDL-C with a 0.1% premature stop codon at the chromosome 19 open reading frame 80 gene (*C19orf80*, aliases include *ANGPTL8*, lipasin, and betatrophin; c.361C>T [p.Gln121*]; rs145464906). Carriers of *ANGPTL8* rs145464906 had 10 mg/dl higher HDL-C (p = $5 \times 10^{-11}$), lower TG ($-15\%$; p = 0.003), and nonsignificantly lower LDL-C ($-5.8$ mg/dl; p = 0.13) than did noncarriers of this variant (Table 4). *ANGPTL8* rs145464906 is seen at a 0.01% frequency in AA participants and we were not able to reliably estimate its effect on lipid measures in this group.

**Table 1.** Number of Exome Array Variants Available for Analysis by Variant Type

| Variant Type | All Polymorphic Sites | MAF < 0.1% | MAF 0.1%–5% | MAF > 5% |
|---|---|---|---|---|
| **42,208 European Ancestry Samples** | | | | |
| Nonsynonymous | 175,444 | 130,553 | 32,501 | 12,390 |
| Nonsense/splice | 13,036 | 10,694 | 1,686 | 656 |
| Synonymous | 5,444 | 3,710 | 978 | 756 |
| Intronic/Intergenic | 14,205 | 139 | 1,186 | 12,880 |
| Other | 1,627 | 156 | 199 | 1,272 |
| **14,330 African Ancestry Samples** | | | | |
| Nonsynonymous | 169,140 | 91,083 | 61,809 | 16,248 |
| Nonsense/splice | 11,768 | 7,729 | 3,177 | 862 |
| Synonymous | 5,398 | 2,341 | 2,098 | 959 |
| Intronic/Intergenic | 14,169 | 56 | 650 | 13,463 |
| Other | 1,780 | 116 | 356 | 1,308 |

Variant type was determined based on dbNSFP v.2.0 annotations. Abbreviation is as follows: MAF, minor allele frequency. Other category includes variants labeled as downstream (n = 187), ncRNA_exonic (n = 111), ncRNA_intronic (n = 447), ncRNA_splicing (n = 1), ncRNA_UTR3 (n = 8), _UTR5 (n = 1), upstream (n = 181), upstream;downstream (n = 8), UTR3 (n = 518), UTR5 (n = 77).

*ANGPTL8* rs145464906 is located in a genomic region previously associated with lipid levels in GWASs (rs737337 at the *DOCK6* [MIM 614194] locus; c.2136A>G [p.(=)]).[18] An analysis conditioning on rs737337 in *DOCK6* revealed rs145464906 to be an independent association ($p_{conditional} = 5.5 \times 10^{-13}$). Another more common coding variant in *ANGPTL8* (8% frequency; rs2278426; c.175C>T [p.Arg59Trp]) has been associated with low LDL-C and low HDL-C.[19] This variant is not present on the exome array, but the correlation between rs2278426 and rs145464906 is low at 0.08 and conditional analyses confirm that rs2278426 and rs145464906 are independent association signals (data not shown).

*Angplt8*-null mice display lower blood TG and functional studies have suggested that ANGPTL8 may activate ANGPTL3 and thereby regulate lipoprotein metabolism.[19,20] Of note, based on studies in mice, another proposed role for ANGPTL8 is as a hormone that promotes pancreatic β cell proliferation, expands β cell mass, and improves glucose tolerance.[21] These studies raise the hypothesis that loss of ANGPTL8 function might worsen measures of glucose tolerance. We studied the association of ANGPTL8 rs145464906 with fasting plasma glucose levels among 10,642 EA participants in the ARIC study and 3,112 EA participants from FHS and observed no evidence for association in either study. In ARIC, plasma glucose did not differ between the carriers of the rs145464906 allele (n = 23) and noncarriers (n = 10,620) (beta = −3.3 mg/dl; SE = 5.8 mg/dl; p = 0.57). In FHS, carriers of rs145464906 (n = 13) had a mean fasting glucose of 88.2 mg/dl whereas noncarriers had mean glucose of 99.9 mg/dl (p = 0.13).

A second discovery among EA participants is a significant association between a 1% frequency variant in the Platelet-Activating Factor Acetylhydrolase 1b, Catalytic Subunit 2 gene (*PAFAH1B2* [MIM 602508]; c.482C>T [p.Ser161Leu]; rs186808413) and HDL-C. Carriers of rs186808413 had 3 mg/dl higher HDL-C (p = 2.2 × $10^{-10}$), lower TG (−10%; p = 2.3 × $10^{-9}$), and marginally lower LDL-C (−3 mg/dl; p = 0.01) than did noncarriers (Table 4). In AA participants, the direction of association between HDL-C and *PAFAH1B2* rs186808413 (0.2% frequency in AA participants) was consistent with the EA results, but not significant (1 mg/dl; p = 0.58). *PAFAH1B2* is located near the chromosome 11 gene cluster involving *APOA1/C3/A4/A5*, a region where several genes regulate TG. Conditional analysis adjusting for six previously studied variants in this region (rs964184, rs3135506, rs2266788, rs76353203, rs147210663, rs140621530) showed *PAFAH1B2* rs186808413 to be an independent association signal ($p_{conditional} = 2.1 \times 10^{-11}$). Platelet-activating factor (PAF) is a lipid messenger functioning in many cellular processes.[22] Plasma platelet-activating factor-acetyl hydrolase (PAF-AH) belongs to a subfamily of phospholipases A2 that remove the sn-2 acetyl group and inactivate a number of oxidized lipids,[23,24] but intracellular type I (PAF-AH-Ib) is structurally distinct and may form a part of a signal transduction pathway.[22]

Third, among AA participants, we found an association of blood TG with a 2% frequency variant in the Collagen, Type XVIII, Alpha 1 gene (*COL18A1* [MIM 120328]; c.331G>A [p.Gly111Arg]; rs114139997), where carriers of rs114139997 had 16% lower TG than did noncarriers (p = 2 × $10^{-16}$). *COL18A1* rs114139997 was also marginally associated with higher HDL-C (2 mg/dl; p = 4 × $10^{-4}$) (Table 4). The frequency of *COL18A1* rs114139997 is 0.003% in EA participants and we were not able to reliably estimate the effect on lipid measures in this group. COL18A1 is a basement membrane proteoglycan with a

**Table 2. Top Association Results for Low-Frequency Variants Based on Level of Statistical Significance**

| Gene | dbSNP ID | Mutation (Substitution) | Chr: Position[a] | MAF[b] | Beta[c] | p Value |
|---|---|---|---|---|---|---|
| **HDL-C – EA (n = 42,208)** | | | | | | |
| *ANGPTL4* | rs116843064 | c.118G>A (p.Glu40Lys) | 19: 8,429,323 | 2.01% | 4 mg/dl | $1.7 \times 10^{-29}$ |
| *LIPG* | rs77960347 | c.1187A>G (p.Asn396Ser) | 18: 47,109,955 | 1.27% | 5 mg/dl | $5.1 \times 10^{-27}$ |
| *LPL* | rs268 | c.953A>G (p.Asn318Ser) | 8: 19,813,529 | 1.78% | −3 mg/dl | $2.4 \times 10^{-19}$ |
| **LDL-C – EA (n = 39,186)** | | | | | | |
| *PCSK9* | rs11591147 | c.137G>T (p.Arg46Leu) | 1: 55,505,647 | 1.54% | −17 mg/dl | $2.7 \times 10^{-59}$ |
| *APOB* | rs5742904 | c.10580G>A (p.Arg3527Gln) | 2: 21,229,160 | 0.05% | 71 mg/dl | $5.6 \times 10^{-34}$ |
| *CBLC* | rs3208856 | c.1075C>T (p.His359Tyr) | 19: 45,296,806 | 3.43% | −8 mg/dl | $6.2 \times 10^{-30}$ |
| **Triglycerides – EA (n = 39,859)** | | | | | | |
| *ANGPTL4* | rs116843064 | c.118G>A (p.Glu40Lys) | 19: 8,429,323 | 2.02% | −15% | $2.9 \times 10^{-37}$ |
| *LPL* | rs268 | c.953A>G (p.Asn318Ser) | 8: 19,813,529 | 1.76% | 14% | $1.2 \times 10^{-22}$ |
| *MAP1A* | rs55707100 | c.7046C>T (p.Pro2349Leu) | 15: 43,820,717 | 3.25% | 9% | $1.4 \times 10^{-17}$ |
| **HDL-C – AA (n = 14,330)** | | | | | | |
| *PCSK7* | rs142953140 | c.1511G>A (p.Arg504His) | 11: 117,089,205 | 0.2% | 17 mg/dl | $3.4 \times 10^{-20}$ |
| *CETP* | rs34119551 | c.17T>A (p.Val6Asp) | 16: 56,995,908 | 0.24% | 16 mg/dl | $2.0 \times 10^{-19}$ |
| *LCAT* | rs35673026 | c.340G>A (p.Val114Met) | 16: 67,976,851 | 0.34% | 10 mg/dl | $2.0 \times 10^{-10}$ |
| **LDL-C – AA (n = 14,227)** | | | | | | |
| *PCSK9* | rs28362286 | c.2037C>A (p.Cys679*) | 1: 55,529,215 | 0.87% | −40 mg/dl | $1.5 \times 10^{-57}$ |
| *PCSK9* | rs67608943 | c.426C>G (p.Tyr142*) | 1: 55,512,222 | 0.30% | −35 mg/dl | $2.6 \times 10^{-16}$ |
| *APOE* | rs769455 | c.487C>T (p.Arg163Cys) | 19: 45,412,040 | 1.93% | −12 mg/dl | $3.6 \times 10^{-12}$ |
| **Triglycerides – AA (n = 14,351)** | | | | | | |
| *APOE* | rs769455 | c.487C>T (p.Arg163Cys) | 19: 45,412,040 | 1.93% | 21% | $2.3 \times 10^{-18}$ |
| *COL18A1* | rs114139997 | c.331G>A (p.Gly111Arg) | 21: 46,875,775 | 1.93% | −16% | $1.6 \times 10^{-11}$ |
| *ZNF259* | rs35120633 | c.791C>T (p.Ala264Val) | 11: 116,655,600 | 2.96% | 13% | $2.7 \times 10^{-12}$ |

Results are based on 42,208 European ancestry (EA) individuals and 14,330 African ancestry (AA) individuals.
[a]Chr: Position is reported in UCSC Genome Browser build hg19.
[b]MAF, minor allele frequency.
[c]Beta is based on the geometric mean for triglycerides.

recently established role in presenting the enzyme lipoprotein lipase to the luminal side of the vascular endothelium.[25] Mice deleted for *Col18a1* (*Col18a1*$^{-/-}$) were shown to have hypertriglyceridemia resulting from decreased lipoprotein lipase activity, and humans with homozygous deficiency of COL18A1 (Knobloch syndrome [MIM 267750]) have been observed to have higher blood TG than do normal individuals.[25]

Finally, in AA participants, we also identified an association of blood HDL-C with a 0.2% frequency variant at the proprotein convertase subtilisin/kexin type 7 serine protease gene (*PCSK7* [MIM 604874]; c.1511G>A [p.Arg504His]; rs142953140). Carriers of the *PCSK7* rs142953140 mutation had 17 mg/dl higher HDL-C levels than did noncarriers (p = $5 \times 10^{-20}$). *PCSK7* rs142953140 was also associated with lower TG (−30%; p = $2 \times 10^{-9}$) and nominally associated with lower LDL-C (−11.5 mg/dl; p = 0.02) among the AA participants (Table 4). *PCSK7* is located

near the chromosome 11 gene cluster involving *APOA1/C3/A4/A5*, a region where several genes regulate HDL-C and TG. Conditional analysis adjusting for six known variants in this region (rs964184, rs3135506, rs2266788, rs76353203, rs147210663, rs140621530) showed *PCSK7* rs142953140 to be an independent association signal ($p_{conditional}$ = $3.4 \times 10^{-20}$). *PCSK7* rs142953140 was not polymorphic in individuals of European ancestry. *PCSK7* belongs to the same family of serine proteases as does *PCSK9*, and a putative link between *PCSK7* and lipoprotein metabolism was recently established in vitro.[26,27] In cultured cells, PCSK7 cleaved angiopoietin-like protein 4 (ANGPTL4) and activated its ability to inhibit lipoprotein lipase.

We studied the functional consequence of overexpression of human *PCSK7* in mice and found that mice overexpressing *PCSK7* in liver have lower HDL-C and higher TG compared to controls (Figure 1), suggesting that gain

**Table 3. Variant Associations with Blood Lipids that Have Not Been Previously Reported**

| Trait | Gene | rsID | Mutation (Substitution) | Beta[a] | p Value | Conditional Variants | $p_{conditional}$ | EA MAF | AA MAF |
|-------|------|------|------------------------|---------|---------|---------------------|-------------------|--------|--------|
| **Ancestry: EA** | | | | | | | | | |
| HDL-C | ANGPTL8 | rs145464906 | c.361C>T (p.Gln121*) | 10 mg/dl | $5.1 \times 10^{-11}$ | rs737337 | $5.5 \times 10^{-13}$ | 0.1% | 0.01% |
| HDL-C | PAFAH1B2 | rs186808413 | c.482C>T (p.Ser161Leu) | 3 mg/dl | $2.2 \times 10^{-10}$ | rs964184, rs3135506, rs2266788, rs76353203, rs14721066, rs140621530 | $2.1 \times 10^{-11}$ | 1.1% | 0.2% |
| **Ancestry: AA** | | | | | | | | | |
| TG | COL18A1 | rs114139997 | c.331G>A (p.Gly111Arg) | −16% | $1.6 \times 10^{-16}$ | – | – | 0.003% | 1.9% |
| HDL-C | PCSK7 | rs142953140 | c.1511G>A (p.Arg504His) | 17 mg/dl | $4.9 \times 10^{-20}$ | rs964184, rs3135506, rs2266788, rs76353203, rs14721066, rs140621530 | $3.4 \times 10^{-20}$ | 0% | 0.2% |

Results are based on 42,208 European ancestry (EA) individuals and 14,330 African ancestry (AA) individuals. Abbreviations are as follows: TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; MAF, minor allele frequency.
[a]Effect estimate for triglycerides is based on the geometric mean.

of PCSK7 function decreases HDL-C and increases TG in vivo. Further studies are warranted to determine whether reduction of hepatic PCSK7, for example through *Pcsk7* knockdown in mice, will result in the same lipid profile as that of human subjects with the *PCSK7* rs142953140 mutation.

### Gene-Based Association

In contrast to low-frequency variants, the alternative alleles at rare variants occur too infrequently to be analyzed individually and must be evaluated by aggregating into sets. Here, typically, the unit of analysis is a gene. We used two methods to perform gene-based analyses incorporating variants annotated as nonsynonymous or splice site: the sum of the number of minor alleles[15] including variants with a MAF <1% or MAF <0.1% and the sequence kernel association test (SKAT),[16] a test that considers effect sizes in both directions, including variants with a MAF <5%. Gene-based association results were well calibrated (Figure S3). We were able to replicate genes previously reported to be associated with blood lipid levels (Table 5). For example, multiple rare alleles at *LDLR* (MIM 606945) were strongly associated with blood LDL-C (p = 5 × $10^{-10}$). No new genetic associations were identified through the gene-based analyses.

### Association with CHD

To assess whether the four newly associated low-frequency coding variants also relate to clinical cardiovascular disease, we tested association of each variant with CHD in 63,470 individuals from 15 EA studies and 13,772 individuals from 7 AA studies (Table S2). Of these participants, 14,201 EA and 2,380 AA individuals were affected with CHD.

**Table 4. Association Results in Other Lipid Traits for the Reported Associations**

| Gene | rsID | Trait | EA Beta | EA p Value | AA Beta | AA p Value |
|------|------|-------|---------|------------|---------|------------|
| ANGPTL8 | rs145464906 | HDL-C | 10 mg/dl | $5.1 \times 10^{-11}$ | – | – |
| | | TG | −15% | 0.003 | – | – |
| | | LDL-C | −5.8 mg/dl | 0.13 | – | – |
| PAFAH1B2 | rs186808413 | HDL-C | 3 mg/dl | $2.2 \times 10^{-10}$ | 1 mg/dl | 0.58 |
| | | TG | −10% | $2.3 \times 10^{-09}$ | −4% | 0.48 |
| | | LDL-C | −3 mg/dl | 0.01 | −2.6 md/dl | 0.58 |
| COL18A1 | rs114139997 | HDL-C | – | – | 2.2 mg/dl | $4.1 \times 10^{-4}$ |
| | | TG | – | – | −16% | $1.6 \times 10^{-16}$ |
| | | LDL-C | – | – | −2.6 mg/dl | 0.12 |
| PCSK7 | rs142953140 | HDL-C | – | – | 17 mg/dl | $4.9 \times 10^{-20}$ |
| | | TG | – | – | −30% | $1.5 \times 10^{-09}$ |
| | | LDL-C | – | – | −11.5 mg/dl | 0.02 |

Abbreviations are as follows: TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; EA, European ancestry; AA, African ancestry; Beta, effect size.
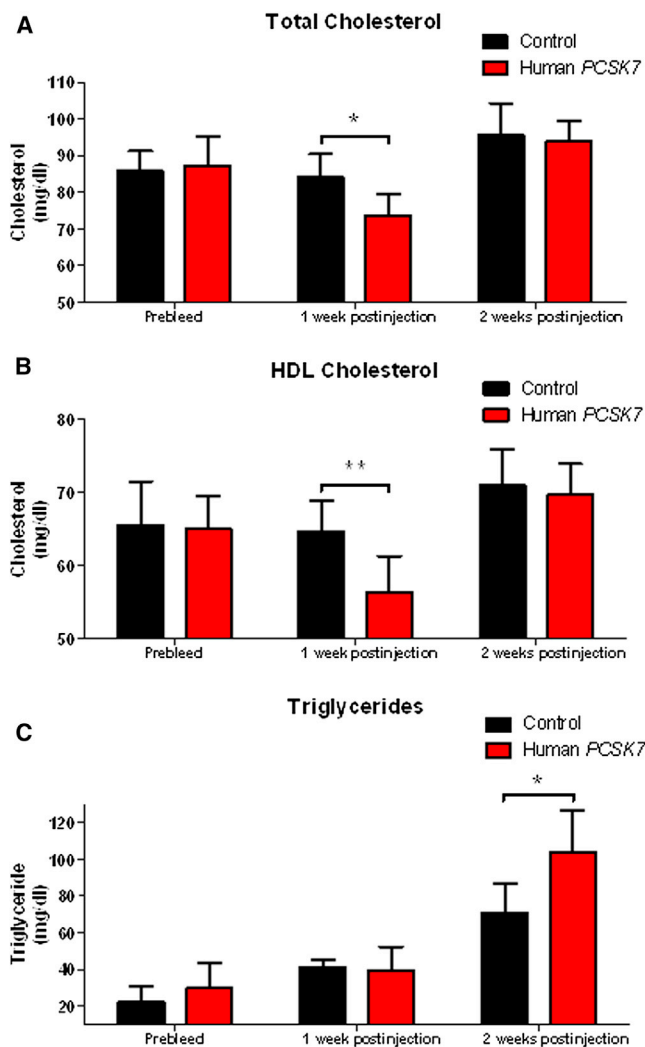
**Figure 1. Effects of Human *PCSK7* Overexpression in Mouse Liver on Plasma Lipids**

Plasma TC (A), HDL-C (B), and TG (C) were measured at baseline and at 7 and 14 days after injection of the human *PCSK7* or control AAV8 vectors in C57BL/6J male mice. Error bars show standard deviations. *p < 0.05 and **p < 0.01, Student's unpaired t test.

As a positive control, we confirmed that the *PCSK9* c.137T (c.137G>T [p.Arg46Leu]; rs11591147) allele leading to lower LDL-C was associated with reduced risk for CHD among EA participants (OR = 0.84, 95% CI = 0.74–0.95, p = 0.007) and *PCSK9* c.2037A (c.2037C>A [p.Cys679*]; rs28362286) allele leading to lower LDL-C was similarly associated with reduced risk for CHD among AA participants (OR = 0.40, 95% CI = 0.23–0.68, p = 5.4 × 10$^{-4}$).

In contrast to the *PCSK9* variants, none of the four coding sequence variants at *ANGPTL8*, *PAFAH1B2*, *COL18A1*, or *PCSK7* were associated with risk for CHD (Table 6).

## Discussion

The contribution of low-frequency DNA sequence variation to complex phenotypes such as blood lipids and

**Table 5. Top Gene-Based Association Results Based on Level of Statistical Significance**

| Gene | p Value | Best Test | Beta[a] | SE | CMAF[b] | No. Variants[c] |
|---|---|---|---|---|---|---|
| **LDL-EA** | | | | | | |
| PCSK9 | 1.81 × 10$^{-62}$ | SKAT | – | – | 0.0626 | 22 |
| LDLR | 4.9 × 10$^{-10}$ | T1 | 15.36 | 2.47 | 0.0027 | 28 |
| **HDL-EA** | | | | | | |
| APOC3 | 3.2 × 10$^{-19}$ | T1 | 11.75 | 1.31 | 0.0035 | 4 |
| ANGPTL4 | 5.41 × 10$^{-30}$ | SKAT | – | – | 0.0283 | 11 |
| LIPG | 2.55 × 10$^{-29}$ | SKAT | – | – | 0.0162 | 12 |
| LPL | 5.76 × 10$^{-29}$ | SKAT | – | – | 0.0372 | 10 |
| **Triglycerides-EA** | | | | | | |
| APOC3 | 4.0 × 10$^{-29}$ | T1 | −0.41 | 0.05 | 0.0036 | 4 |
| ANGPTL4 | 2.70 × 10$^{-37}$ | SKAT | – | – | 0.0286 | 11 |
| LPL | 1.56 × 10$^{-31}$ | SKAT | – | – | 0.037 | 10 |
| **LDL-AA** | | | | | | |
| PCSK9 | 1.40 × 10$^{-71}$ | SKAT | – | – | 0.1295 | 25 |
| APOE | 2.65 × 10$^{-12}$ | SKAT | – | – | 0.0205 | 3 |
| **HDL-AA** | | | | | | |
| APOC3 | 9.5 × 10$^{-12}$ | T1 | 10.14 | 1.49 | 0.0055 | 4 |
| CETP | 7.7 × 10$^{-09}$ | T1 | 6.18 | 1.07 | 0.0065 | 7 |
| **Triglycerides-AA** | | | | | | |
| APOC3 | 2.9 × 10$^{-19}$ | T1 | −0.34 | 0.05 | 0.0055 | 4 |
| APOE | 2.26 × 10$^{-18}$ | SKAT | – | – | 0.0205 | 3 |

Abbreviations are as follows: TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; EA, European ancestry; AA, African ancestry; Beta, effect size; SE, standard error.
[a]Beta is in mg/dl units for LDL and HDL and in percent change for triglycerides.
[b]CMAF, cumulative minor allele frequency of the variants contributing to the test.
[c]No. variants, number of variants contributing to the test.

CHD has been the subject of much debate. Some have argued that low-frequency variants with large effects may be instrumental in discovering genes and that such discoveries can lead to novel therapeutic targets. By using the Exome Array in >42,000 EA individuals and >14,000 AA individuals, we discovered four low-frequency variants with large effects on lipids, but none of these variants associated with risk for CHD.

Several limitations of our study deserve mention. We were not able to evaluate extremely rare variants that may be unique to individuals; exome or whole-genome sequencing is needed to capture this type of variation. As such, it remains possible that a burden of such very rare mutations could contribute to plasma lipid variation. If so, the sample sizes required to yield new rare variant discoveries are likely to be extraordinarily large.

The Exome Array is constrained to the coding and splice site variation observed in the ~12,000 individuals who comprised the initial exome sequencing discovery set.

**Table 6. Association of Lipid Variants with Coronary Heart Disease**

| Gene | Mutation (Substitution) | rsID | Case Frequency | Control Frequency | Odds Ratio | p Value |
|---|---|---|---|---|---|---|
| **Ancestry: EA** | | | | | | |
| *ANGPTL8* | c.361C>T (p.Gln121*l) | rs145464906 | 0.29% | 0.27% | 1.023 | 0.985 |
| *PAFAH1B2* | c.482C>T (p.Ser161Leu) | rs186808413 | 2.05% | 2.23% | 0.905 | 0.170 |
| **Ancestry: AA** | | | | | | |
| *PCSK7* | c.1511G>A (p.Arg504His) | rs142953140 | 0.53% | 0.47% | 1.258 | 0.592 |
| *COL18A1* | c.331G>A (p.Gly111Arg) | rs114139997 | 3.39% | 3.38% | 0.996 | 0.971 |

Association with CHD was performed in a total of 63,470 EA individuals and 13,772 AA individuals. 14,201 EA and 2,380 AA individuals developed CHD.

Furthermore, ~20% of the content contributed for design failed to be converted into genotyping assays and thus these variants are not present on the Exome Array.

The lack of significant association of our reported low-frequency variants with CHD may be due to insufficient statistical power. For the CHD association testing among EA participants, we had >80% power at $\alpha = 0.05$ to detect an odds ratio outside the range (0.86, 1.14) with a 2% frequency variant, and an odds ratio outside the range (0.74, 1.29) with a 0.5% frequency variant. Among AA participants, we had >80% power at $\alpha = 0.05$ to detect an odds ratio outside the range (0.69, 1.35) with a 2% frequency variant, and an odds ratio outside the range (0.43, 1.73) with a 0.5% frequency variant.

Of note, all four of our reported loci were associated with HDL-C, which may not be a causal risk factor for clinical CHD, and this observation could explain the lack of association with CHD.[28] Finally, we do not yet fully understand the biologic mechanism by which the genes reported here affect blood lipoproteins.

In summary, by using the Exome Array, we identified four low-frequency coding variants in *ANGPTL8, PAFAH1B2, COL18A1*, and *PCSK7* that altered plasma HDL-C and/or TG but did not affect risk for CHD. These results suggest that the example of *PCSK9* with low-frequency alleles that affect both plasma lipids and CHD is likely to be an exception rather than a paradigm.

## Supplemental Data

Supplemental Data include descriptions of study samples, Supplemental Acknowledgments, three figures, and two tables and can be found with this article online at http://www.cell.com/AJHG/.

## Acknowledgments

## Web Resources

The URLs for data presented herein are as follows:

dbNSFP v.2.0, https://sites.google.com/site/jpopgen/dbNSFP

Exome Array site information, ftp://share.sph.umich.edu/exomeChip/

Exome Chip Design, http://genome.sph.umich.edu/wiki/Exome_Chip_Design

ExomeChip – Charge Consortium, http://www.chargeconsortium.com/main/exomechip

Genetic Power Calculator, http://pngu.mgh.harvard.edu/~purcell/gpc/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

R metafor package, http://cran.r-project.org/web/packages/metafor/index.html

seqMeta, http://cran.r-project.org/web/packages/seqMeta/

UCSC Genome Browser, http://genome.ucsc.edu

## References

1. Stein, E.A., Mellis, S., Yancopoulos, G.D., Stahl, N., Logan, D., Smith, W.B., Lisbon, E., Gutierrez, M., Webb, C., Wu, R., et al. (2012). Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. N. Engl. J. Med. *366*, 1108–1118.

2. Abifadel, M., Varret, M., Rabès, J.P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., et al. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat. Genet. *34*, 154–156.

3. Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K., and Hobbs, H.H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet. *37*, 161–165.

4. Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. *354*, 1264–1272.

5. Cohen, J.C., and Hobbs, H.H. (2013). Genetics. Simple genetics for a complex disease. Science *340*, 689–690.

6. Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L., et al. (2012). Exome sequencing and the genetic basis of complex traits. Nat. Genet. *44*, 623–630.

7. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C., and Boerwinkle, E.; CHARGE Consortium (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet *2*, 73–80.

8. Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples, L.A., Fornage, M., et al. (2013). Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS ONE *8*, e68095.

9. Goldstein, J.I., Crenshaw, A., Carey, J., Grant, G.B., Maguire, J., Fromer, M., O'Dushlaine, C., Moran, J.L., Chambert, K., Stevens, C., et al.; Swedish Schizophrenia Consortium; ARRA Autism Sequencing Consortium (2012). zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics *28*, 2543–2545.

10. Tobin, M.D., Sheehan, N.A., Scurrah, K.J., and Burton, P.R. (2005). Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Stat. Med. *24*, 2911–2935.

11. Baigent, C., Keech, A., Kearney, P.M., Blackwell, L., Buck, G., Pollicino, C., Kirby, A., Sourjina, T., Peto, R., Collins, R., and Simes, R.; Cholesterol Treatment Trialists' (CTT) Collaborators (2005). Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet *366*, 1267–1278.

12. Scandinavian Simvastatin Survival Study Group (1994). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet *344*, 1383–1389.

13. Friedewald, W.T., Levy, R.I., and Fredrickson, D.S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin. Chem. *18*, 499–502.

14. Warnick, G.R., Knopp, R.H., Fitzpatrick, V., and Branson, L. (1990). Estimating low-density lipoprotein cholesterol by the Friedewald equation is adequate for classifying patients on the basis of nationally recommended cutpoints. Clin. Chem. *36*, 15–19.

15. Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am. J. Hum. Genet. *83*, 311–321.

16. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.

17. Liu, X., Jian, X., and Boerwinkle, E. (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum. Mutat. *34*, E2393–E2402.

18. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

19. Quagliarini, F., Wang, Y., Kozlitina, J., Grishin, N.V., Hyde, R., Boerwinkle, E., Valenzuela, D.M., Murphy, A.J., Cohen, J.C., and Hobbs, H.H. (2012). Atypical angiopoietin-like protein that regulates ANGPTL3. Proc. Natl. Acad. Sci. USA *109*, 19751–19756.

20. Zhang, R., and Abou-Samra, A.B. (2013). Emerging roles of Lipasin as a critical lipid regulator. Biochem. Biophys. Res. Commun. *432*, 401–405.

21. Yi, P., Park, J.S., and Melton, D.A. (2013). Betatrophin: a hormone that controls pancreatic β cell proliferation. Cell *153*, 747–758.

22. Ho, Y.S., Swenson, L., Derewenda, U., Serre, L., Wei, Y., Dauter, Z., Hattori, M., Adachi, T., Aoki, J., Arai, H., et al. (1997). Brain acetylhydrolase that inactivates platelet-activating factor is a G-protein-like trimer. Nature *385*, 89–93.

23. Stafforini, D.M., McIntyre, T.M., Zimmerman, G.A., and Prescott, S.M. (1997). Platelet-activating factor acetylhydrolases. J. Biol. Chem. *272*, 17895–17898.

24. Derewenda, Z.S., and Derewenda, U. (1998). The structure and function of platelet-activating factor acetylhydrolases. Cell. Mol. Life Sci. *54*, 446–455.

25. Bishop, J.R., Passos-Bueno, M.R., Fong, L., Stanford, K.I., Gonzales, J.C., Yeh, E., Young, S.G., Bensadoun, A., Witztum, J.L., Esko, J.D., and Moulton, K.S. (2010). Deletion of the basement membrane heparan sulfate proteoglycan type XVIII collagen causes hypertriglyceridemia in mice and humans. PLoS ONE *5*, e13919.

26. Oexle, K., Ried, J.S., Hicks, A.A., Tanaka, T., Hayward, C., Bruegel, M., Gögele, M., Lichtner, P., Müller-Myhsok, B., Döring, A., et al. (2011). Novel association to the proprotein convertase PCSK7 gene locus revealed by analysing soluble transferrin receptor (sTfR) levels. Hum. Mol. Genet. *20*, 1042–1047.

27. Seidah, N.G., Khatib, A.M., and Prat, A. (2006). The proprotein convertases and their implication in sterol and/or lipid metabolism. Biol. Chem. *387*, 871–877.

28. Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Hólm, H., Ding, E.L., Johnson, T., et al. (2012). Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet *380*, 572–580.