

MATERIAL AND METHODS

Study Design and Study Participants

The ARIC study is a prospective investigation of cardiovascular disease incidence involving 15,792 men and women aged 45 to 64 years and recruited from four U.S. communities in 1987–1989. Participants underwent a baseline exam and up to 4 follow-up visits. A detailed description of the ARIC study design and methods has been published elsewhere.¹ The current study was conducted among individuals who participated in ARIC study visit 4 (1996-1998). Of the 11,656 eligible individuals who participated in visit 4, we excluded those without sdLDL-C data (n=168), with self-reported race being neither white nor black (n=31), and black race at centers in Minneapolis or Washington County (n=38). For incident CHD analyses, individuals with prevalent CHD at visit 4 (n=972), or those with missing covariate data for the multivariable models (n=565) were excluded, resulting in 9,882 individuals who were included in our final analysis. Prevalent CHD was defined as self-reported myocardial infarction before visit 1 or silent myocardial infarction (diagnosed by electrocardiographic changes), validated myocardial infarction, or revascularization between visits 1 and 4.

The ascertainment procedure for incident CHD events has been described previously.² Briefly, incident CHD was defined as those participants with hospitalized myocardial infarction, fatal CHD, or cardiac procedure by 2008.

Participant Examination

Medical history, demographic data, anthropometric data, blood pressure measurements (ARIC Manual 11 visit 4, NHLBI 1997), fasting glucose, and fasting lipids (ARIC Manual 8, NHLBI 1994) obtained during visit 4 were used for this analysis. Cigarette smoking and the use of antihypertensive and lipid-lowering medications were ascertained from a standardized questionnaire.

Hypertension was defined as systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, prior physician diagnosis of hypertension, or use of antihypertensive medication during the previous 2 weeks. Diabetes mellitus was defined as a fasting glucose level ≥ 126 mg/dL, a nonfasting glucose level ≥ 200 mg/dL, or a self-reported history of physician-diagnosed diabetes or treatment for diabetes. The study was approved by the institutional review committees of all participating centers, and all participants provided informed consent.

Laboratory Analyses

Plasma total cholesterol, triglycerides, high-density lipoprotein cholesterol (HDL-C) were measured using enzymatic methods; low-density lipoprotein (LDL) cholesterol was calculated according to the Friedewald formula.³ Large buoyant LDL cholesterol (lbLDL-C) was estimated by subtracting the sdLDL-C concentration from the LDL-C concentration. Non-HDL-C was calculated by subtracting the HDL-C concentration from the total cholesterol concentration. The fraction of LDL-C which was sdLDL-C was calculated by dividing the sdLDL-C concentration by the LDL-C concentration. Plasma apolipoproteins AI and B and high-sensitivity C-reactive protein (hs-CRP) were measured by an immunonephelometric assay using a BNII nephelometer (Siemens Healthcare Diagnostics, Deerfield, IL).

A homogeneous assay method was used for the direct measurement of sdLDL-C in plasma (sd-LDL-EX "Seiken", Denka Seiken, Tokyo, Japan) on a Hitachi 917 automated chemistry analyzer. This method has been previously validated and shown to be in good agreement with the ultracentrifugal method used to isolate LDL in the 1.044–1.063 g/ml density range used by many investigators for sdLDL, with an $r^2 = 0.91$.⁴ The intra-assay and inter-assay coefficients of variation for the sdLDL-C assay were 1.3% and 3.1%, respectively. The reliability coefficient for the sdLDL-C assay based on 435 blinded quality control replicates was 0.92. It is important to note that the lbLDL-C fraction may include intermediate-density lipoprotein (IDL)

cholesterol since we did not use an ultracentrifugation method to isolate this specific lbLDL-C fraction.

Genotyping

Genome-wide genotyping of single-nucleotide polymorphisms (SNPs) was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, CA). Study participants who refused DNA testing, had high missing rates, suspected contaminated samples, samples with genotype mismatch with 47 previously genotyped SNPs, and genetic outliers based on identity-by-state statistics and EIGENSTRAT principal components analysis were excluded.

Additionally, monomorphic SNPs, SNPs with no chromosome location, and SNPs with call rate <95%, minor allele frequency <1%, or Hardy–Weinberg equilibrium $p < 10^{-6}$ were also excluded.

The Affymetrix 6.0 genotypes and a cosmopolitan set of HapMap haplotypes were used to impute 2.4 million autosomal SNPs. Imputation results were summarized as an allele dosage, which was defined as the expected number of copies of the minor allele at each SNP. We applied an *a priori* threshold of 5.0×10^{-8} for statistical significance for these genome-wide association analyses. When more than 1 genome-wide significant SNP clustered at a locus, we took the SNP with the smallest P value as the lead SNP.

Genome-wide association study (GWAS) analyses were adjusted for age and gender, using a co-dominant model. For the GWAS, linear regression analyses of sdLDL-C, lbLDL-C, and the ratio of sdLDL-C to LDL-C were carried out using PLINK (version 1.07) and ProbABEL, and Cox proportional hazards modeling was used to test for an independent association of select genetic variants and the presence of incident CHD. The association between the genome-wide significant SNP rs508487 located in the PCSK7 gene and CHD was examined from results in 40,260 cases and 60,790 controls from the Coronary Artery Disease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study.⁵ Although details differed among

the contributing studies in CARDIoGRAM, the definition of CHD included clinically defined myocardial infarction or angiographically accessed coronary artery disease.

Using ARIC data from the CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Exome Chip project, we analyzed rare variants designated as nonsynonymous, splicing, or stop gain in the PCSK7 gene for association with sdLDL-C as previously described.⁶ These variants were first tested individually for association with sdLDL-C using linear regression. Since these variants had minor allele frequencies below 1%, they were also tested collectively in two separate gene-based tests. The first gene-based test was the T1 burden test used to detect an association between variation in each gene and sdLDL-C. This test is the most powerful when all variants have the same direction of effect on the phenotype. The second gene-based was the SKAT test which allowed for different directions of effect between the variants included. All analyses were adjusted for age and gender.

Statistical Analyses

The distribution of sdLDL-C and all other clinically relevant continuous variables measured in our analysis population were evaluated to assess normality. For this analysis, we modeled sdLDL-C both as a continuous and categorical variable. As a categorical variable, quartile measures were used as cut-points to obtain four separate groups. The cut-points were obtained from the distribution of sdLDL-C in the whole analysis population (25th, 50th, and 75th percentile values were 28.0, 39.7, and 54.7 mg/dL, respectively). Means or proportions of demographic characteristics and traditional cardiovascular risk factors of the study participants were reported by sdLDL-C quartiles. The p-values for trends were evaluated with linear or logistic regression using quartile number adjusted for age, race, and gender. Pearson's correlation coefficient was used to assess the correlation of sdLDL-C and traditional or novel cardiovascular risk factors. Triglycerides and hs-CRP were log-transformed to account for their non-Gaussian distributions. Associations between sdLDL-C and incident CHD were determined using Cox proportional

hazards modeling, in both unadjusted and adjusted models. The basic model (Model 1) adjusted for age, gender, and race as potential confounders. Model 2 was additionally adjusted for smoking status (current versus not current), body mass index (BMI), hypertension, antihypertensive medication use, HDL-C, log triglycerides, lipid-lowering medication use, presence of diabetes mellitus (defined as a fasting glucose level ≥ 126 mg/dL, a nonfasting glucose level ≥ 200 mg/dL, or a self-reported history of physician-diagnosed diabetes), diabetes medication use, and log hs-CRP. In all models, the 2nd, 3rd and 4th quartiles were compared to the 1st quartile (the referent group). Statistical analysis was performed using SAS version 9.2 (SAS Institute Inc., Cary, NC) and STATA version 11 (StataCorp LP, College Station, TX). All tests were 2-sided with a p-value < 0.05 considered significant.

References

1. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 1989;129:687-702.
2. White AD, Folsom AR, Chambless LE, Sharret AR, Yang K, Conwill D, Higgins M, Williams OD, Tyroler HA. Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. *J Clin Epidemiol* 1996;49:223-33.
3. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972;18:499-502.
4. Albers JJ, Kennedy H, Marcovina SM. Evaluation of a new homogenous method for detection of small dense LDL cholesterol: Comparison with the LDL cholesterol profile obtained by density gradient ultracentrifugation. *Clin Chim Acta* 2011;412:556-61.
5. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011;43:333-8.
6. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* 2013;8:e68095.