

A Common *PCSK9* Haplotype, Encompassing the E670G Coding Single Nucleotide Polymorphism, Is a Novel Genetic Marker for Plasma Low-Density Lipoprotein Cholesterol Levels and Severity of Coronary Atherosclerosis

Suet N. Chen, MS,* Christie M. Ballantyne, MD, FACC,* Antonio M. Gotto, JR, MD, DPHIL, FACC,† Yanli Tan, RN,* James T. Willerson, MD, FACC,‡ Ali J. Marian, MD, FACC*

Houston, Texas; and New York, New York

OBJECTIVES	We sought to determine the effects of <i>PCSK9</i> variants on plasma low-density lipoprotein cholesterol (LDL-C) levels, severity of coronary atherosclerosis, and response to statin therapy in the Lipoprotein Coronary Atherosclerosis Study (LCAS) population.
BACKGROUND	Mutations in <i>PCSK9</i> cause autosomal-dominant hypercholesterolemia. We hypothesized that <i>PCSK9</i> variants could affect plasma LDL-C in individuals with polygenic hypercholesterolemia.
METHODS	We sequenced all 12 exons and boundaries to detect novel polymorphisms, and genotyped 372 subjects in LCAS and 319 subjects in a second independent population for six polymorphisms, including novel leucine repeats, by fluorescently tagged markers. We reconstructed haplotypes using a Bayesian algorithm.
RESULTS	Permutation test results showed statistically significant differences in global haplotype distribution among the tertiles of LDL-C (odds ratio [OR]: 2.36, 95% confidence interval [CI]: 1.90 to 4.32, $p = 0.005$) and minimum lumen diameter of coronary lesions (OR: 1.83, 95% CI: 1.01 to 3.55, $p = 0.045$). Regression analysis identified haplotype 3 as an independent determinant of LDL-C levels (adjusted $R^2 = 2.2\%$, $F = 9.37$, $p = 0.002$). Haplotype structure analysis identified E670G as the determinant variant, exerting a dose effect (GG > EG > EE) and accounting for 3.5% of plasma LDL-C variability ($F = 14.6$, $p < 0.001$). Plasma total cholesterol, apolipoprotein B, and lipoprotein (a) levels were also associated with the E670G variant. Distributions of the E670G genotypes in an independent normolipidemic and the hyperlipidemic LCAS populations were significantly different ($F = 7.2$, $p = 0.027$). No significant treatment-by-genotype interactions were detected. The false positive report probability was between 2% and 8%.
CONCLUSIONS	Haplotype 3 encompassing the E670G variant is an independent determinant of plasma LDL-C levels and the severity of coronary atherosclerosis. (J Am Coll Cardiol 2005;45: 1611–9) © 2005 by the American College of Cardiology Foundation

Autosomal-dominant hypercholesterolemia (ADH) is a relatively uncommon disorder (OMIM143890) characterized by elevated plasma low-density lipoprotein cholesterol (LDL-C) levels and premature atherosclerosis. The best-characterized causal genes for ADH are *LDLR* and *APOB*,

(S127R and F216L) in *PCSK9* (3). These findings indicate that *PCSK9* is involved in LDL-C homeostasis.

We hypothesized that genetic variants of *PCSK9* could also affect plasma levels of LDL-C in polygenic (non-Mendelian) hypercholesterolemia. To test this hypothesis, we screened the exons and exon-intron boundaries of *PCSK9*, identified novel polymorphisms, reconstructed haplotypes, and analyzed the association of the haplotypes and genotypes with plasma levels of lipids, severity of atherosclerosis, and their responses to fluvastatin therapy in a well-characterized population with high LDL-C levels. We then tested the main results for replication in an independent population with normal plasma LDL-C levels.

See page 1620

causing familial hypercholesterolemia and familial defective apoB-100, respectively (1). Recently, a third locus was mapped to chromosome 1p32 (2) and subsequent mutation screening led to identification of two missense mutations

MATERIALS AND METHODS

The Lipoprotein Coronary Atherosclerosis Study (LCAS) population. All subjects provided informed consent, and the institutional review board approved the study. The design of the LCAS (4), the primary result (5), and results of selected genetic studies (6–9) have been published. The LCAS population comprised 372 35- to 75-

From the *Sections of Cardiology and Atherosclerosis, Center for Preventive Cardiology, Department of Medicine, Baylor College of Medicine, Houston, Texas; †Weill Medical College of Cornell University, New York, New York; and the ‡University of Texas Health Science Center, Houston, Texas. Supported by grants from the National Heart, Lung, and Blood Institute, Specialized Centers of Research P50-HL54313, RO1 HL68884, and a TexGen grant from Greater Houston Community Foundation.

Manuscript received October 14, 2004; revised manuscript received January 4, 2005; accepted January 11, 2005.

Abbreviations and Acronyms

ADH	=	autosomal-dominant hypercholesterolemia
FPRP	=	false positive report probability
HDL-C	=	high-density lipoprotein cholesterol
HWE	=	Hardy-Weinberg equilibrium
LCAS	=	Lipoprotein Coronary Atherosclerosis Study
LDL-C	=	low-density lipoprotein cholesterol
Lp(a)	=	lipoprotein (a)
MAF	=	minor allele frequencies
MI	=	myocardial infarction
PCR	=	polymerase chain reaction
PL-EM	=	partition-ligation-expectation-maximization
SNP	=	single nucleotide polymorphism
STR	=	short tandem repeat
TC	=	total cholesterol
TG	=	triglycerides

year-old subjects who had plasma LDL-C levels of 115 to 190 mg/dl despite diet and one or more coronary lesions causing 30% to 75%-diameter stenosis. Phenotypic characterization included measurements of plasma total cholesterol (TC), LDL-C, high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), lipoprotein (a) (Lp[a]), and apolipoprotein levels and quantitative coronary angiography before and 2.5 years after randomization to fluvastatin (40 mg/day) or a placebo. Definite or probable myocardial infarction (MI), unstable angina requiring hospitalization, percutaneous coronary interventions, coronary artery bypass grafting, and death from any cause were recorded during the follow up.

Detection of novel single nucleotide polymorphisms (SNPs) in PCSK9. We screened all 12 exons of *PCSK9* and exon-intron boundaries for the presence of polymorphisms in 50 healthy Caucasians by polymerase chain reaction (PCR) and direct sequencing (See Appendix for online supplementary Table 1). We compared the sequence among the individuals and with the published GenBank sequence (NT_032977).

Genotyping. Genotyping was performed by 5' nuclease assay (allelic discrimination assays) and capillary electrophoresis for 13,097C > T, 13,248A > G, 19,018A > G (I474V), 23,968A > G (E670G), and 24,609T > C SNPs using the Applied Biosystems Prism 7900HT Sequence Detection System (See Appendix for online supplementary Table 2). The short tandem repeat (STR) polymorphism was genotyped using fluorescent-labeled primers, PCR amplification of a 243-bp encompassing fragment, and capillary electrophoresis on an Applied Biosystems 3100 genetic analyzer and analyzed using the GeneScan/Genotyper software (version 3.7, 2001, Applied Biosystems Inc.). An investigator without knowledge of the angiographic and clinical data performed the genotyping.

In silico haplotyping and population frequency of the haplotypes. We used Phase 2.0 (10,11) to reconstruct haplotypes and estimate population haplotype frequencies, as published (12). Phase 2 implements a Bayesian statistical

method to infer phase and to reconstruct haplotypes from population genetics by Markov Chain-Monte Carlo algorithm and coalescent theory. It has been shown to infer haplotypes more accurately than other Bayesian-based methods in real data sets (10). Other previously used algorithms, such as Partition-Ligation-Expectation-Maximization (PL-EM) and Haplotyper (13,14), did not permit haplotype reconstruction in the presence of three alleles of the STR marker. We determined the estimates of population haplotype frequencies for 20, 100, 500, 1,000, and 10,000 iterations. Pairwise linkage disequilibrium (LD) was calculated with known gametic phases, estimated by expectation-maximization algorithm, using Arlequin 2.0 (University of Geneva, Geneva, Switzerland) (15).

Replication study population (TexGen). To determine whether the main findings in the LCAS population could be replicated in an independent population with normal plasma LDL-C levels, we screened 1,010 individuals from our outpatient clinics and identified 319 subjects who had plasma LDL-C levels of <130 mg/dl. Demographic data were collected, and fasting plasma levels of total cholesterol, HDL-C, LDL-C, and TG were obtained. Subjects with advanced comorbid conditions, including malignancies, advanced heart failure, and valvular heart disease, were not included.

False positive report probability. We determined the probability of no true association between the variant and phenotype given a statistically significant finding using a previously described method as follows: False positive report probability (FPRP) = $\alpha(1 - \pi)/[\alpha(1 - \pi) + (1 - \beta)\pi]$, whereby α is the probability of a statistically significant finding, given that null hypothesis of no association is true; $1 - \beta$ is the statistical power, and π is the prior probability (16).

Statistical analysis. Mean and SD were used for analysis of the continuous variables unless otherwise specified. Permutation analysis (between 100 and 1,000 permutations and 10,000 iterations) was used to compare haplotype frequencies among the categorical phenotypes (Phase 2.0). Because Phase 2.0 does not permit a permutation test for the parametric phenotypes, they were discretized to categorical phenotypes according to the tertiles of the traits. Linear and logistic regression analyses were used to detect association of the parametric and nonparametric phenotypes with the haplotypes, respectively. Differences in parametric phenotypes between carriers and noncarriers of specific haplotypes or genotypes were compared by analysis of variance or Student *t* test and for nonparametric phenotypes by the Kruskal-Wallis test. Treatment-by-haplotype interactions were analyzed by permutation analysis as well as by analysis of variance (general linear model).

RESULTS

Population characteristics. Baseline demographics, plasma levels of lipids, and indices of severity of coronary athero-

Table 1. Baseline Characteristics of the Lipoprotein Coronary Atherosclerosis Study (LCAS) and TexGen Study Populations

	LCAS (N = 372)	TexGen (N = 319)
Demographic (N = 372)		
Age (yrs)	58.8 ± 7.7	60.3 ± 12.6
Gender (male/female)	310(83)/62(17)	210(66)/109(34)
White/African American/others	333(90)/26(7)/13(3)	253(79)/48(15)/18(6)
Height (m)	1.73 ± 0.08	1.72 ± 1.05
Weight (kg)	84.6 ± 15.4	86.2 ± 19.2
Body mass index	28.1 ± 4.4	29.2 ± 5.6
Systolic blood pressure (mm Hg)	124.4 ± 15.2	128.1 ± 21.6
Diastolic blood pressure (mm Hg)	77.0 ± 9.1	75.5 ± 12.8
Diabetes mellitus	14(4)	62(19)
Smoker	75(20)	31(10)
History of myocardial infarction	154(41)	99(31)
Baseline lipid levels (N = 372)		
Total cholesterol (mg/dl)	220.6 ± 24.3	172.5 ± 32.8
HDL-C (mg/dl)	44.0 ± 11.4	45.2 ± 17.7
LDL-C (mg/dl)	144.6 ± 20.0	106.7 ± 24.5
Triglycerides (mg/dl)	160.0 ± 56.9	125.2 ± 71.4
Apolipoprotein A-I	133.0 ± 27.3	NA
Apolipoprotein B	134.9 ± 20.9	NA
Apolipoprotein C-III	37.0 ± 12.1	NA
Baseline severity of atherosclerosis (N = 316)		
Mean number of lesions	3.0 ± 1.38	NA
Mean number of total occlusions	0.32 ± 0.59	NA
Mean minimal lumen diameter	1.67 ± 0.40	NA

Values are presented as n (%) for categorical variables and as mean ± standard deviation for continuous variables.
 HDL-C = high-density lipoprotein cholesterol, LDL-C = low-density lipoprotein cholesterol, MLD = minimal lumen diameter, NA = not available.

sclerosis are shown in Table 1 (4,5). Because exclusion of the non-Caucasian subjects (10% of the LCAS population) did not affect the results, the overall results were presented in the entire group. Whenever a specific association was detected, the analysis was repeated in the Caucasian population.

The TexGen population was slightly older ($p = 0.065$) and was composed of 34% female subjects as compared with 17% ($p < 0.0001$) and 79% Caucasians as compared with 90% ($p = 0.0003$) in the LCAS population. Diabetes mellitus was more common in the TexGen population ($p < 0.0001$), but prior history of MI ($p = 0.006$) and (current) smoking ($p = 0.0001$) were less prevalent. Per a priori selection criteria, the mean plasma total cholesterol ($p < 0.0001$), LDL-C ($p < 0.0001$), and TG ($p < 0.0001$) levels were lower, but HDL-C ($p = 0.299$) levels were comparable to those in the LCAS population.

Polymorphisms in PCSK9. The list and location of SNPs and STR markers, the reference number, and the minor allele frequencies (MAF) are shown in Figure 1. The STR polymorphism in exon 1 (CTG)_n leads to expansion of amino acid leucine repeats L_n of 9, 10, and 11.

Genotype distribution and Hardy-Weinberg equilibrium. The distribution of genotypes of each SNP followed the Hardy-Weinberg equilibrium (HWE) (See Appendix for online Table 3).

LD and frequency of haplotypes. Table 2 shows the D' and r² values for pairwise LD. Increasing the number of

iterations to more than 100 did not significantly affect the estimates of population haplotype frequencies (See Appendix for online supplementary Table 4). Thirty, including 11 relatively common, haplotypes ($f > 0.02$) were detected, with mean and median values of certainties of haplotypes assignment being 87% and 94%, respectively. Estimates of population haplotype frequencies (10,000 iterations) are shown in Table 3.

PCSK9 genotypes/haplotypes and demographic phenotypes. There were no significant associations between SNPs and baseline demographic variables including age, gender, ethnic background, weight, height, body mass index, systolic and diastolic blood pressure, history of diabetes, and history of smoking. Only L_n STR polymorphism was associated with age, and those with the L9L9 were younger (57.7 ± 7.7 , N = 268) as compared with those with L9L10 (61.2 ± 7.3 , N = 93), L9L11 (62.7 ± 6.9 , N = 7), or L10L10 (66.0 ± 4.0 , N = 7) groups ($F = 7.56$, $p < 0.001$). Therefore, the possible effect of the L_n polymorphism on the phenotypes was determined by interaction analysis.

PCSK9 genotypes/haplotypes and baseline plasma levels of lipids. A permutation test showed a significant association between PCSK9 haplotypes and the tertiles of plasma LDL-C levels ($p = 0.05$), and linear regression analysis identified haplotype 3 as an independent predictor (adjusted R² = 2.2%, $F = 9.37$, $p = 0.002$). Mean plasma LDL-C levels were 143.65 ± 19.7 , 152.21 ± 20.06 , and $172.25 \pm$

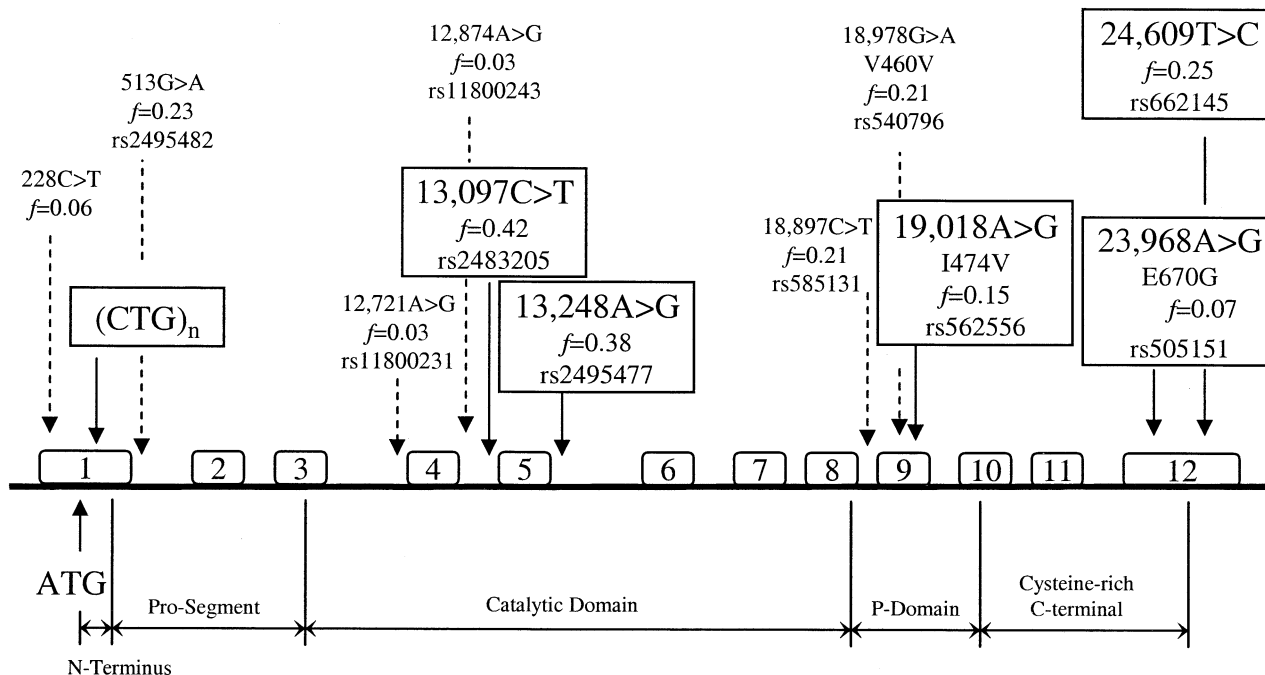


Figure 1. Schematic diagram showing structure of *PCSK9* and exons/introns, location of single nucleotide polymorphisms (SNPs), and polymorphisms. Segments of the gene corresponding to neural apoptosis-regulated convertase-1 domains are shown. The reference number (rs) and the frequency (*f*) of the less common allele for each SNP are also shown. Polymorphisms shown in boxes, representing the L_n short tandem repeat, nonsynonymous coding SNPs, and those located in intron-exon boundaries, were genotyped and analyzed for haplotype reconstruction and linkage disequilibrium studies.

15.91 (mg/dl) in those with zero copy (N = 335), one copy (N = 35), and two copies (N = 2) of haplotype 3 (F = 4.95, p = 0.008). Similarly, LDL-C levels were higher in those with haplotype 3 as compared with the referent haplotype (Fig. 2). Logistic regression analysis showed an association between the tertiles of LDL-C and haplotype 3 (OR 2.36, 95% CI 1.90 to 4.32, p = 0.005). Multivariate linear regression analysis that included age, gender, body mass index, smoking status, alcohol consumption, diabetes, and all 11 common haplotypes identified haplotype 3 (regression coefficient: 10.4, T = 2.58, p = 0.010, N = 39), and age (regression coefficient: -0.2841, T = -1.97, p = 0.050) as independent predictors of plasma LDL-C levels. Reconstruction of the haplotypes after exclusion of the L_n polymorphism, which was in linkage equilibrium with the SNPs, reduced the number of haplotypes from 30 to 19. In accord with the original analysis, haplotype 3 remained an independent predictor of plasma LDL-C levels (regression coefficient: 10.2 ± 4.0, T = 2.6, p = 0.011). Plasma LDL-C levels were associated with haplotype 3 in a copy-dependent manner (0 copies: 143.7 ± 19.6, N = 343; 1 copy: 153.9 ± 22.0, N = 27; 2 copies: 172.3 ± 15.9, n = 2; F = 5.3, p = 0.006). Regarding other lipids and lipoproteins, haplotype 3 (regression coefficient: 12.8, T = 2.78, p = 0.006) and gender (regression coefficient: 22.3, T = 7.0, p < 0.001) were independent predictors of plasma levels of TC (adjusted R²: 12.3%, F = 26.4, p < 0.001). Subjects with haplotype 3 had higher plasma TC levels than those without (256.0 ± 12.0 vs. 226.0 ± 25.4 vs. 219.8 ± 24.0, p = 0.042). There were no associations between

haplotypes and plasma levels of HDL-C and triglycerides. Finally, multivariate regression analysis in a Caucasian-only population showed a significant association between plasma LDL-C levels and haplotype 3 (F = 4.27, p = 0.015). Plasma levels of LDL-C (mg/dl) were higher in subjects with haplotype 3 (152.1 ± 20.1, N = 33) than those without (143.0 ± 19.7, N = 300, p = 0.018).

Analysis of haplotype structure (Table 3) showed that haplotype 3 is uniquely identified by the E670G cSNP, suggesting that E670G cSNP is responsible for the observed association of plasma LDL-C levels with haplotype 3. Indeed, multivariate regression analysis that included all SNPs, the STR marker, demographics, and the baseline variables showed that E670G (23,968A > G) cSNP was the only independent predictor of plasma LDL-C levels, accounting for 3.5% of its variability (F = 14.6, p < 0.001). Accordingly, LDL-C levels were higher in those with the GG, intermediate in those with the AG, and lowest in those with the AA genotypes (Fig. 2). Similarly, plasma TC, apoB, and Lp(a) levels were also associated with E670G cSNP (Table 4). Analysis for possible association of plasma LDL-C levels with each of the six polymorphisms showed that the 24,609T > C SNP in exon 12 was also associated with plasma LDL-C levels (TT: 142.5 ± 19.3, N = 207; TC: 146.2 ± 20.4, N = 142; CC: 154.0 ± 20.4, N = 23; F = 4.19, p = 0.016). However, the observed association reflected LD between the 24,609T > C and the E670G SNPs, as in multivariate analysis, only the E670G cSNP (regression coefficient: = 2.7, p = 0.004) remained an independent predictor of plasma LDL-C levels but not the

Table 2. Pairwise Linkage Disequilibrium and r^2 Between *PCSK9* Polymorphisms

Polymorphism	(CTG) _n L _n	13,097C>T	13,248A>G	19,018A>G I474V	23,968A>G E670G	24,609T>C
(CTG) _n		$r^2 = 0.01$	$r^2 = 0.02$	$r^2 = 0.02$	$r^2 = 0.0$	$r^2 = 0.01$
L _n		$p = 0.023$	$p = 0.199$	$p = 0.001$	$p = 0.498$	$p = 0.233$
13,097C>T	0.1387 (9L) $p = 0.0653$ -0.1977 (10L) $p = 0.0114$ 1.0 (11L) $p = 0.0224$		$r^2 = 0.15$ $p < 0.001$	$r^2 = 0.06$ $p < 0.001$	$r^2 = 0.04$ $p = 0.103$	$r^2 = 0.06$ $p = 0.076$
13,248A>G	0.1295 (9L) $p = 0.0583$ -0.1793 (10L) $p = 0.0116$ 1.0 (11L) $p = 0.0378$	0.6055 $p < 0.0000$		$r^2 = 0.10$ $p < 0.001$	$r^2 = 0.0$ $p = 0.371$	$r^2 = 0.06$ $p < 0.001$
19,018A>G I474V	0.1715 (9L) $p < 0.0000$ -0.1384 (10L) $p = 0.0003$ -0.6631 (11L) $p < 0.0000$	0.6308 $p < 0.0000$	0.7292 $p < 0.0000$		$r^2 = 0.0$ $p = 0.701$	$r^2 = 0.21$ $p < 0.001$
23,968A>G E670G	-0.5125 (9L) $p = 0.0981$ 0.4797 (10L) $p = 0.1361$ 1.0 (11L) $p = 0.4526$	-0.5719 $p = 0.0002$	-0.2805 $p = 0.0913$	-0.5212 $p = 0.0892$		$r^2 = 0.32$ $p < 0.001$
24,609T>C	-0.1079 (9L) $p = 0.0340$ 0.0736 (10L) $p = 0.1640$ 0.6177 (11L) $p = 0.0047$	-0.3065 $p = 0.0913$	-0.5032 $p < 0.0000$	-0.8697 $p < 0.0000$	-0.9757 $p < 0.0000$	

D' values are shown on the left half of the table, r^2 values on the right half of the table. The small r^2 values between polymorphisms suggest, despite the presence of significant linkage disequilibrium, that none could represent the other reliably.

24,609T > C SNP. Finally, when multivariate regression analysis was restricted to Caucasians, the E670G cSNP remained a significant determinant of plasma LDL-C levels (regression coefficient: 3.2, $p = 0.008$).

PCSK9 genotypes/haplotypes and severity of coronary atherosclerosis. A permutation test showed significant differences in global haplotype frequencies among minimum

lumen diameter (MLD) tertiles ($p = 0.02$). In logistic regression analysis, a modest association between MLD tertiles and haplotype 3 was present (OR: 1.83, 95% CI: 1.01 to 3.55, $p = 0.045$). Otherwise, no significant associations between quantitative indices of severity of coronary atherosclerosis at the baseline and the remaining haplotypes or SNPs were detected. In multivariate linear regression analysis, only body mass index

Table 3. *PCSK9* Common Haplotypes and Their Number and Estimated Population Frequencies

Identification Number	Haplotype	(CTG) _n	SNPs					<i>f</i>	N
			13,097 C>T	13,248 A>G	19,018 A>G I474V	23,968 A>G E670G	24,609 T>C		
2	9-C-A-A-A-T	9	C	A	A	A	T	0.3556 ± 0.0080	275
16	9-T-G-A-A-T	9	T	G	A	A	T	0.1215 ± 0.0066	100
13	9-T-A-A-A-T	9	T	A	A	A	T	0.1022 ± 0.0064	69
18	9-T-G-G-A-C	9	T	G	G	A	C	0.0722 ± 0.0063	54
3	9-C-A-A-G-C	9	C	A	A	G	C	0.0435 ± 0.0035	39
21	10-C-A-A-A-T	10	C	A	A	A	T	0.0396 ± 0.0057	27
8	9-C-G-A-A-T	9	C	G	A	A	T	0.0465 ± 0.0055	26
28	10-T-G-G-A-C	10	T	G	G	A	C	0.0271 ± 0.0040	26
25	10-T-A-A-A-T	10	T	A	A	A	T	0.0320 ± 0.0047	25
15	9-T-G-A-A-C	9	T	G	A	A	C	0.0260 ± 0.0032	21
23	10-C-G-A-A-T	10	C	G	A	A	T	0.0219 ± 0.0039	19

Several rare haplotypes with estimated frequencies of <0.02 (N between 1 and 9) were also detected, which are not included in this table. *f* = estimate of haplotype frequency and standard error; SNP = single-nucleotide polymorphism.

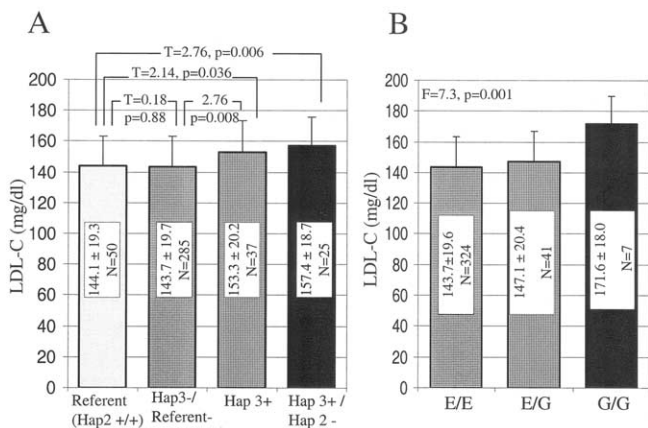


Figure 2. Plasma low-density lipoprotein cholesterol (LDL-C) levels according to haplotypes (A) and E670G genotypes (B). (A) Referent indicates subjects homozygous for the most common haplotype (haplotype 2+/+). Haplotype 3-/referent- indicates subjects with neither haplotype 3 nor the reference haplotype; haplotype 3+ identifies heterozygous (N = 35) and homozygous (N = 2) subjects with haplotype 3; haplotype 3+/haplotype 2- denotes subjects with haplotype 3 but no haplotype 2. Mean values of LDL-C, standard deviation of the mean, and the number of subjects are indicated in each graph. The results of pairwise *t* test are shown. (B) The mean values among the three genotypes were compared by analysis of variance and depicted. Pairwise comparisons showed statistically significant differences between AA and GG ($p = 0.007$) as well as between AG and GG ($p = 0.011$), but not between AA and AG ($p = 0.321$).

was an independent predictor of the number of coronary lesions at baseline ($R^2 = 2.0\%$, $F = 7.50$, $p = 0.007$).

Pharmacogenetics. Treatment with fluvastatin reduced plasma LDL-C levels by $25.3 \pm 17.5\%$ and apoB by $16.0 \pm 15.4\%$, and raised plasma levels of HDL-C and apoA-I by $9.3 \pm 15.7\%$ and $6.9 \pm 20.1\%$, respectively (all $p < 0.005$ as compared with placebo). Plasma TG levels remained unchanged ($0.8 \pm 41\%$) in the fluvastatin group, but were increased by $9.8 \pm 40\%$ in the placebo group ($p = 0.033$).

In multivariate regression analysis, only treatment with fluvastatin was the primary predictor of changes in plasma

LDL-C during the follow-up period. There were no significant haplotype- or genotype-by-treatment interactions regarding changes in TC, LDL-C, and HDL-C levels or change in MLD or development of new coronary lesions in response to treatment with fluvastatin (as compared with placebo).

Fifty-four clinical events occurred during the follow-up. There were no statistically significant interactions between SNPs or haplotypes and the incidence of new clinical events in the entire population.

FPRP. Because *PCSK9* mutations are known to affect LDL-C levels in ADH, the prior probability was considered moderate to high and multiple π values (0.1, 0.15, 0.20, and 0.25) were tested. The sample size of 37 subjects with haplotype 3 and 335 without provided $>99\%$ power to detect the observed 10-mg/dl differences in plasma LDL-C levels at an α value of 0.05 and SD of 20 mg/dl. Accordingly, FPRP was calculated for the observed p value of 0.008, π values ranging from 0.1 to 0.25, and β values ranging from 0.05 to 0.2. Even at the most relaxed β value of 0.2, the FPRP varied from 3% to 8% for π between 0.1 to 0.25. At the β value of 0.01, which was the calculated value, FPRP was between 2% and 7% for the prior probabilities ranging from 0.1 to 0.25. Calculation for the observed association of E670G cSNP with plasma LDL-C levels also showed a low rate of false positive probability.

Results in the TexGen population. Because the E670G cSNP was the main SNP determining plasma LDL-C levels, we genotyped 319 individuals for the E670G cSNP by the 5' nuclease assay. The AA genotype was present in 291 individuals; AG in 28 subjects and GG in 0 subjects, which followed the HWE (expected numbers were AA: 292; AG: 27; and GG: 0; $p = 0.888$). Compared with the LCAS population, the frequency of the G allele was lower (0.043 vs. 0.074, chi-squared = 5.5, $p = 0.019$) and the GG genotype was absent in the TexGen population. The overall

Table 4. Plasma Levels of Lipids and Severity of Coronary Atherosclerosis According to 23,968A>G (E670G) Genotypes (N = 372)

	AA (EE) (N = 324)	AG (EG) (N = 41)	GG (GG) (N = 7)	p
Lipoproteins and apolipoproteins				
TC (mg/dl)	220.0 ± 24.1	220.0 ± 24.3	253.1 ± 8.9	0.002
LDL-C (mg/dl)	143.7 ± 19.6	147.1 ± 20.4	171.6 ± 18.0	0.001*
HDL-C (mg/dl)	43.9 ± 11.3	43.4 ± 10.7	52.9 ± 16.8	0.111
TG (mg/dl)	162.0 ± 56.9	147.5 ± 54.4	142.6 ± 67.6	0.219
ApoB (mg/dl)	134.8 ± 20.6	132.3 ± 21.1	154.3 ± 26.5	0.035
ApoA-I (mg/dl)	133.3 ± 27.6	127.8 ± 21.4	150.4 ± 36.4	0.109
ApoCIII (mg/dl)	37.5 ± 12.3	33.2 ± 10.0	39.7 ± 13.7	0.089
Lp(a) (mg/dl)	36.9 ± 33.5	34.2 ± 31.8	67.4 ± 48.8	0.051
Severity of coronary atherosclerosis (QCA)				
Qualifying lesion (n)	3.0 ± 1.4	2.9 ± 1.4	3.0 ± 1.4	0.880
Total occlusion (n)	0.31 ± 0.58	0.34 ± 0.58	0.86 ± 1.07	0.048
MLD (mm)	1.66 ± 0.40	1.74 ± 0.43	1.89 ± 0.37	0.252

*Post-hoc Tukey test results show significant differences between AA and GG and between AG and GG, but not between AA and AG.

ApoA-I = apolipoprotein A1; ApoB = apolipoprotein B; ApoCIII = apolipoprotein C III; Lp(a) = lipoprotein (a); QCA = quantitative coronary angiogram; TC = total cholesterol; TG = triglycerides; other abbreviations as in Table 1.

distribution of genotypes differed between the LCAS and the TexGen populations (chi squared = 7.2, $p = 0.027$). Plasma LDL-C levels were 88.1 ± 25.3 in those with the AA and 94.1 ± 21.4 in those with the AG genotype ($p = 0.10$). Plasma levels of TC, HDL-C, and TG were not significantly different between the AA and AG groups (data not shown).

DISCUSSION

We genotyped six *PCSK9* polymorphisms, reconstructed haplotypes, and determined association of the biochemical, angiographic, and clinical phenotypes of coronary atherosclerosis with the haplotypes and genotypes in the LCAS population. The results are remarkable for the presence of significant copy-number-dependent association between haplotype 3 and plasma levels of LDL-C and TC and to a lesser extent MLD. Haplotype 3, which comprises the information content of L_n STR polymorphism and five SNPs, is the only haplotype with amino acid glycine at position 670 in the protein. Consequently, the E670G cSNP was identified as the risk variant and others had no discernible effect. The results in a second independent population (TexGen), showing lower frequency of the risk allele in those with normal LDL-C levels, provided indirect evidence of support for the findings in the LCAS population. The findings are also in accord with the results of recent linkage mapping and identification of two mutations in *PCSK9* in patients with ADH (2,3). Nonetheless, because of the differences in the characteristics of the LCAS and TexGen populations, the results require confirmation in additional replicates and through experimentation. Studies in a larger sample size could establish the clinical significance of the observed associations of plasma levels of lipids and the haplotypes.

The strengths of the study are the prospective and placebo-controlled randomized design of LCAS and comprehensive phenotypic characterization, an essential component of LD studies. The study includes detection and analysis of novel polymorphisms including the L_n polymorphism, and reconstruction of haplotypes comprising information on the content of six polymorphisms that collectively span the *PCSK9* locus. Statistical analyses used a permutation test, which is considered robust and less prone to spurious association. Cognizant of the relatively high rate of spurious results in genetic association studies, we also calculated the FPRP, which was less than 8%, even under the most relaxed conditions. Finally, the findings are in accord with recent genetic linkage mapping and detection of mutations in *PCSK9* in families with ADH (2,3). Collectively, these findings suggest the presence of a significant association between *PCSK9* E670G cSNP (haplotype 3) and plasma LDL-C and TC levels in non-Mendelian dyslipidemia.

Genetic linkage studies have established *PCSK9* as a causal gene for familial ADH (2,3,17), however, its function

and the mechanisms by which *PCSK9* mutations affect plasma LDL-C levels are largely unknown (18). *PCSK9* encodes neural apoptosis-regulated convertase (NARC-1 aka PCSK9), a novel 691-amino-acid proprotein convertase expressed predominantly in the liver and intestine (19,20). The protein is a member of subtilase subfamily with multiple domains (Fig. 1) (20) including signaling peptide (aa 1-30), pro-segment (aa 31-147), catalytic (aa 148-425), and cysteine-rich C-terminal (aa 526 to 691) domains. Human PCSK9 is synthesized as a zymogen and undergoes autocatalytic intramolecular processing in the endoplasmic reticulum, a step necessary for exiting the endoplasmic reticulum. The E670G cSNP is located in the cysteine-rich C-terminal domain, which is involved in regulation of autoprocessing, because deletion of this domain leads to accumulation of the processed PCSK9 (19). The biological role of PCSK9 in regulating plasma LDL-C levels also remains elusive. Recent studies suggest that PCSK9 negatively regulates expression of LDL-C receptors in the liver through a post-translational mechanism before the internalization and recycling of the receptor (21,22). Adenoviral-mediated overexpression of murine *Pcsk9* results in near-complete depletion of the LDL-C receptor, whereas inactivation of the catalytic activity of *Pcsk9* has no effect (22). Accordingly, PCSK9 mutations are expected to increase the activity of the enzyme (gain of function). Other effects of PCSK9 mutations comprise decreased zymogen processing of PCSK9, reduced LDL-C receptor density (23), and an increased production rate of apoB100 (24). Finally, PCSK9 could interfere with the ability of the LDL-C receptor to bind to LDL-C (25). Whether E670G cSNP impairs the effects of PCSK9 on LDL-C receptor abundance and/or activity or the production rate of apoB100 remains unknown.

Despite a strong association of the plasma LDL-C levels with haplotype 3 (E670G cSNP), MLD showed only a modest association. This is not surprising because the effect of gene variants is expected to be stronger on the immediate (gene products) and weaker on the remote phenotypes (such as atherosclerosis or death) because of the contribution of the competing factors to the distant phenotypes. A recent study in a Japanese population implicated an intronic SNP (C-161T) and a cSNP (I474V) in influencing plasma LDL-C levels but reported no association with MI (26). We genotyped the I474V SNP and detected a MAF of 0.15, which is considerably higher than that in the Japanese population (MAF = 0.03). The I474V cSNP was in partial LD with the E670G cSNP and was not an independent predictor of the phenotypes in the LCAS population. The MAF of the C(-161)T SNP, determined in 50 subjects, was 0.059. Because of the relatively low frequency and its location in a noncoding region, we did not genotype the C(-161)T SNP in the LCAS population.

Our approach for genetic studies of complex trait, pending the completion of the HapMap project and resolution of the superiority of the SNP-centric or haplotype-centric

approach, is based on analysis of common (common disease-common variant hypothesis) putatively functional SNPs (pfSNPs) in the gene/locus of interest, complemented by haplotype reconstruction. Accordingly, we analyzed six common SNPs, including two cSNPs, one amino acid repeat polymorphism, and one regulatory (3' untranslated region) SNP. There are at least six additional SNPs in *PCSK9*, as shown in Figure 1, that were genotyped only in 50 subjects. The MAFs of these SNPs are shown in Figure 1. We did not genotype the entire LCAS population for these SNPs because they were in near-complete LD with those analyzed and/or because they were located in introns. Thus, we cannot exclude the possible presence of additional common haplotypes in the LCAS population. Finally, because we analyzed only the common SNPs, based on common disease-common variant hypothesis, we cannot exclude the possibility of uncommon alleles contributing to plasma LDL-C levels in the LCAS population.

The sample size of LCAS provided 80% power, at an α value of 0.05, to detect a 20% difference in the mean baseline plasma LDL-C levels for haplotypes that were present in at least 25 heterozygous subjects. Considering the relatively low frequency of the approximately half of the haplotypes, a larger sample size would be necessary to detect effects of the rare haplotypes or smaller effects of the common haplotypes. In addition, mean MLD values in the placebo and fluvastatin groups changed only by -0.11 and -0.04 mm, respectively, which are consistent with the results of other angiographic regression/progression studies (27-30), but less than 10% of the baseline MLD. Thus, a much larger sample size and/or a longer duration of follow up may be necessary to detect the potential effects of genetic variants on MLD. Similarly, the number of clinical events could be too small to detect the potential impact of genotypes or haplotypes on the clinical events.

In conclusion, we have identified and analyzed novel and known polymorphisms in *PCSK9* locus, reconstructed haplotypes encompassing the information content of six polymorphisms, and shown that haplotype 3, representative of the E670G cSNP, is an important determinant of plasma LDL-C and TC levels and is associated with the severity of coronary atherosclerosis in the LCAS population. Identification of molecular mechanism(s) by which *PCSK9* variants affect plasma LDL-C levels could provide new insight into the pathogenesis of atherosclerosis and development of new drug targets.

Reprint requests and correspondence: Dr. Ali J. Marian, Baylor College of Medicine, Section of Cardiology, One Baylor Plaza, 519D, Houston, Texas 77030. E-mail: amarian@bcm.tmc.edu.

REFERENCES

- Rader DJ, Cohen J, Hobbs HH. Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. *J Clin Invest* 2003;111:1795-803.
- Varret M, Rabes JP, Saint-Jore B, et al. A third major locus for autosomal dominant hypercholesterolemia maps to 1p34.1-p32. *Am J Hum Genet* 1999;64:1378-87.
- Abifadel M, Varret M, Rabes JP, et al. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nat Genet* 2003;34:154-6.
- West MS, Herd JA, Ballantyne CM, et al. The Lipoprotein and Coronary Atherosclerosis Study (LCAS): design, methods, and baseline data of a trial of fluvastatin in patients without severe hypercholesterolemia. *Control Clin Trials* 1996;17:550-83.
- Herd JA, Ballantyne CM, Farmer JA, et al. Effects of fluvastatin on coronary atherosclerosis in patients with mild to moderate cholesterol elevations (Lipoprotein and Coronary Atherosclerosis Study [LCAS]). *Am J Cardiol* 1997;80:278-86.
- Ballantyne CM, Herd JA, Stein EA, et al. Apolipoprotein E genotypes and response of plasma lipids and progression-regression of coronary atherosclerosis to lipid-lowering drug therapy. *J Am Coll Cardiol* 2000;36:1572-8.
- Cahilly C, Ballantyne CM, Lim DS, Gotto A, Marian AJ. A variant of p22(phox), involved in generation of reactive oxygen species in the vessel wall, is associated with progression of coronary atherosclerosis. *Circ Res* 2000;86:391-5.
- Lutucuta S, Ballantyne CM, Elghannam H, Gotto AM, Jr., Marian AJ. Novel polymorphisms in promoter region of *atp* binding cassette transporter gene and plasma lipids, severity, progression, and regression of coronary atherosclerosis and response to therapy. *Circ Res* 2001;88:969-73.
- Ma K, Cilingiroglu M, Otvos JD, Ballantyne CM, Marian AJ, Chan L. Endothelial lipase is a major genetic determinant for high-density lipoprotein concentration, structure, and metabolism. *Proc Natl Acad Sci U S A* 2003;100:2748-53.
- Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population data. *Am J Hum Genet* 2003;73:1162-9.
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-89.
- Chen SN, Tsybouleva N, Ballantyne CM, Gotto AM, Jr., Marian AJ. Effects of PPAR alpha, gamma and delta haplotypes on plasma levels of lipids, severity and progression of coronary atherosclerosis and response to statin therapy in the Lipoprotein Coronary Atherosclerosis Study. *Pharmacogenetics* 2004;14:61-71.
- Qin ZS, Niu T, Liu JS. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* 2002;71:1242-7.
- Liu JS, Sabatti C, Teng J, Keats BJ, Risch N. Bayesian analysis of haplotypes for linkage disequilibrium mapping. *Genome Res* 2001;11:1716-24.
- Slatkin M, Excoffier L. Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity* 1996;76(Pt 4):377-83.
- Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434-42.
- Timms KM, Wagner S, Samuels ME, et al. A mutation in *PCSK9* causing autosomal-dominant hypercholesterolemia in a Utah pedigree. *Hum Genet* 2004;114:349-53.
- Maxwell KN, Breslow JL. Adenoviral-mediated expression of *Pcsk9* in mice results in a low-density lipoprotein receptor knockout phenotype. *Proc Natl Acad Sci U S A* 2004;101:7100-5.
- Naureckiene S, Ma L, Sreekumar K, et al. Functional characterization of *Narc 1*, a novel proteinase related to proteinase K. *Arch Biochem Biophys* 2003;420:55-67.
- Seidah NG, Benjannet S, Wickham L, et al. The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation. *Proc Natl Acad Sci U S A* 2003;100:928-33.
- Maxwell KN, Breslow JL. Adenoviral-mediated expression of *Pcsk9* in mice results in a low-density lipoprotein receptor knockout phenotype. *Proc Natl Acad Sci U S A* 2004;101:7100-5.
- Park SW, Moon YA, Horton JD. Post-transcriptional regulation of LDL receptor protein by proprotein convertase subtilisin/kexin type 9a (*PCSK9*) in mouse liver. *J Biol Chem* 2004;279:50630-8.

23. Benjannet S, Rhainds D, Essalmani R, et al. NARC-1/PCSK9 and its natural mutants: zymogen cleavage and effects on the LDLR and LDL-cholesterol. *J Biol Chem* 2004;279:48865-75.
24. Ouguerram K, Chetiveaux M, Zair Y, et al. Apolipoprotein B100 metabolism in autosomal-dominant hypercholesterolemia related to mutations in PCSK9. *Arterioscler Thromb Vasc Biol* 2004;24:1448-53.
25. Attie AD. The mystery of PCSK9. *Arterioscler Thromb Vasc Biol* 2004;24:1337-9.
26. Shioji K, Mannami T, Kokubo Y, et al. Genetic variants in PCSK9 affect the cholesterol level in Japanese. *J Hum Genet* 2004;49:109-14.
27. Ornish D, Scherwitz LW, Billings JH, et al. Intensive lifestyle changes for reversal of coronary heart disease. *JAMA* 1998;280:2001-7.
28. Jukema JW, Bruschke AV, van Boven AJ, et al. Effects of lipid lowering by pravastatin on progression and regression of coronary artery disease in symptomatic men with normal to moderately elevated serum cholesterol levels. The Regression Growth Evaluation Statin Study (REGRESS). *Circulation* 1995;91:2528-40.
29. Effect of simvastatin on coronary atheroma. the Multicentre Anti-Atheroma Study (MAAS). *Lancet* 1994;344:633-8.
30. Frick MH, Syvanne M, Nieminen MS, et al. Prevention of the angiographic progression of coronary and vein-graft atherosclerosis by gemfibrozil after coronary bypass surgery in men with low levels of HDL cholesterol. Lipid Coronary Angiography Trial (LOCAT) Study Group. *Circulation* 1997;96:2137-43.

APPENDIX

For Supplementary Tables 1 to 4, please see the May 17, 2005, issue of *JACC* at www.onlinejacc.org.