

ORIGINAL ARTICLE

Polymorphisms Associated with Cholesterol and Risk of Cardiovascular Events

Sekar Kathiresan, M.D., Olle Melander, M.D., Ph.D., Dragi Anevski, Ph.D., Candace Guiducci, B.S., Noël P. Burt, B.S., Charlotta Roos, M.Sc., Joel N. Hirschhorn, M.D., Ph.D., Göran Berglund, M.D., Ph.D., Bo Hedblad, M.D., Ph.D., Leif Groop, M.D., Ph.D., David M. Altshuler, M.D., Ph.D., Christopher Newton-Cheh, M.D., M.P.H., and Marju Orho-Melander, Ph.D.

ABSTRACT

BACKGROUND

From the Cardiology Division (S.K., C.N.-C.) and the Center for Human Genetic Research and Department of Molecular Biology (D.M.A.), Massachusetts General Hospital; the Departments of Medicine (S.K., D.M.A., C.N.-C.), Pediatrics (J.N.H.), and Genetics (J.N.H., D.M.A.), Harvard Medical School; and the Division of Endocrinology and Genetics, Children's Hospital (J.N.H.) — all in Boston; the Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA (S.K., C.G., N.P.B., J.N.H., D.M.A., C.N.-C.); the Departments of Clinical Sciences, Hypertension, and Cardiovascular Diseases (O.M.), Diabetes and Endocrinology (D.A., C.R., L.G., M.O.-M.), Internal Medicine (G.B.), and Epidemiological Research (B.H.), University Hospital Malmö, Lund University, Malmö, and Chalmers University of Technology, Göteborg (D.A.) — both in Sweden; and Helsinki University Hospital, Helsinki (L.G.). Address reprint requests to Dr. Kathiresan at the Cardiovascular Disease Prevention Center, Cardiology Division, Massachusetts General Hospital, 25 New Chardon St., Suite 301, Boston, MA 02114, or at skathiresan@partners.org.

Drs. Kathiresan and Orho-Melander contributed equally to this article.

N Engl J Med 2008;358:1240-9.
Copyright © 2008 Massachusetts Medical Society.

Common single-nucleotide polymorphisms (SNPs) that are associated with blood low-density lipoprotein (LDL) or high-density lipoprotein (HDL) cholesterol modestly affect lipid levels. We tested the hypothesis that a combination of such SNPs contributes to the risk of cardiovascular disease.

METHODS

We studied SNPs at nine loci in 5414 subjects from the cardiovascular cohort of the Malmö Diet and Cancer Study. We first validated the association between SNPs and either LDL or HDL cholesterol and subsequently created a genotype score on the basis of the number of unfavorable alleles. We used Cox proportional-hazards models to determine the time to the first cardiovascular event in relation to the genotype score.

RESULTS

All nine SNPs showed replication of an association with levels of either LDL or HDL cholesterol. With increasing genotype scores, the level of LDL cholesterol increased from 152 mg to 171 mg per deciliter (3.9 to 4.4 mmol per liter), whereas HDL cholesterol decreased from 60 mg to 51 mg per deciliter (1.6 to 1.3 mmol per liter). During follow-up (median, 10.6 years), 238 subjects had a first cardiovascular event. The genotype score was associated with incident cardiovascular disease in models adjusted for covariates including baseline lipid levels ($P < 0.001$). The use of the genotype score did not improve the clinical risk prediction, as assessed by the C statistic. However, there was a significant improvement in risk classification with the use of models that included the genotype score, as compared with those that did not include the genotype score.

CONCLUSIONS

A genotype score of nine validated SNPs that are associated with modulation in levels of LDL or HDL cholesterol was an independent risk factor for incident cardiovascular disease. The score did not improve risk discrimination but did modestly improve clinical risk reclassification for individual subjects beyond standard clinical factors.

PLASMA LEVELS OF LOW-DENSITY LIPOprotein (LDL) and high-density lipoprotein (HDL) cholesterol are associated with a future risk of cardiovascular disease.^{1,2} It has been estimated that roughly 50% of variation in LDL and HDL cholesterol levels is heritable.³ Several common DNA sequence variants have been related to blood LDL or HDL cholesterol levels.⁴⁻¹¹ In light of the increased practicality of conducting genome-wide association studies, it is likely that additional polymorphisms associated with lipid levels will be identified.

These observations suggest two hypotheses. First, a DNA sequence variant that is related to blood lipoprotein levels may influence the risk of cardiovascular disease. Second, since DNA sequence variants represent an index of lifelong exposure to altered lipoprotein levels,¹² whereas plasma measurements vary, DNA sequence variants may add predictive information beyond a single measurement of blood lipids.

The potential of such genetic testing is limited by the fact that each sequence variant that has been discovered to date explains only a modest fraction of the variance (2% or less) in lipid levels.¹³ However, the combination of several lipid-related polymorphisms could improve risk prediction. Accordingly, we sought first to validate the association between single-nucleotide polymorphisms (SNPs) at nine loci with the level of LDL or HDL cholesterol and then to evaluate the ability of a panel of validated lipid-modulating SNPs to predict a first cardiovascular event in a large, community-based cohort.

METHODS

STUDY SAMPLE AND RISK FACTORS

The Malmö Diet and Cancer Study (MDCS) is a community-based prospective epidemiologic cohort of 28,449 subjects who were recruited for baseline examination between 1991 and 1996.¹⁴ From this cohort, 6103 subjects were randomly selected to participate in a cardiovascular cohort (MDCS-CC), which seeks to investigate risk factors for cardiovascular disease. The MDCS-CC study protocols were approved by the ethics committee at Lund University. All subjects provided written informed consent.

All subjects underwent a review of their medical history, a physical examination, and a labora-

tory assessment of cardiovascular risk factors. The status of cigarette smoking was elicited by a self-administered questionnaire and was coded as never, former, or current. Current cigarette smoking was defined as any smoking within the past year. A parental or sibling history of myocardial infarction (regardless of the age at onset) was also elicited by questionnaire. Blood pressure was measured once after 10 minutes of rest in the supine position. Diabetes mellitus was defined by self-report on the basis of a physician's diagnosis, a fasting blood glucose level of more than 6.0 mmol per liter (108 mg per deciliter), or the use of antidiabetes medication. C-reactive protein was measured by a high-sensitivity assay (Tina-quant CRP, Roche Diagnostics). In fasting venous blood samples, we measured total cholesterol, HDL cholesterol, and triglycerides according to standard procedures at the Department of Clinical Chemistry, University Hospital Malmö. Levels of LDL cholesterol were calculated according to Friedewald's formula, with the assignment of a missing value to subjects with a triglyceride level of more than 400 mg per deciliter (4.5 mmol per liter). The use of lipid-lowering and antihypertensive therapy was recorded.

SNP SELECTION AND GENOTYPING

We selected for study 11 SNPs in 9 genes on the basis of literature reports of individual associations and from a genomewide association study. We selected six SNPs from six loci on the basis of their association with levels of LDL or HDL cholesterol in at least one previous study. These six SNPs were, for association with LDL cholesterol, *APOB* (apolipoprotein B, rs693),⁴ *PCSK9* (proprotein convertase subtilisin/kexin type 9, rs11591147),⁶ and *LDLR* (low-density lipoprotein receptor, rs688)⁵; and for association with HDL cholesterol, *CETP* (cholesteryl ester transfer protein, rs1800775),⁸ *LIPC* (hepatic lipase, rs1800588),⁹ and *LPL* (lipoprotein lipase, rs328).¹⁵ We selected five SNPs from a genomewide association study of levels of blood LDL and HDL cholesterol in 2931 subjects from Sweden and Finland (a sample distinct from the subjects in the MDCS-CC).¹³ In the genomewide association study, we found SNPs in *APOB* (rs7575840), *APOE* cluster (apolipoprotein E, rs4420638), *HMGCR* (3-hydroxy-3-methylglutaryl-coenzyme A reductase, rs12654264), and *LDLR* (rs1529729) to be nominally associated with LDL cholesterol ($P=1\times 10^{-4}$,

$P=3.4\times 10^{-13}$, $P=4\times 10^{-4}$, and $P=0.002$, respectively) and *ABCA1* (ATP-binding cassette subfamily A member 1, rs3890182) with HDL cholesterol ($P=3.3\times 10^{-5}$).¹⁶

Genotyping was performed with the use of matrix-assisted laser desorption–ionization time-of-flight mass spectrometry on a MassARRAY platform (Sequenom), as described previously.¹⁷ All SNPs were in Hardy–Weinberg equilibrium ($P>0.001$). The genotyping success rate was 96%. Using 15 samples analyzed in quadruplicate, we found the genotyping error rate to be <0.7%.

OUTCOMES

Cardiovascular events were ascertained through linkage of the 10-digit personal identification number of each Swedish citizen with three registries: the Swedish Hospital Discharge Register, the Swedish Cause of Death Register, and the Stroke Register of Malmö.¹⁸ The prespecified composite end point of cardiovascular events was defined as myocardial infarction, ischemic stroke, and death from coronary heart disease. Myocardial infarction was defined on the basis of codes 410 and I21 in the *International Classification of Diseases, 9th Revision and 10th Revision* (ICD-9 and ICD-10), respectively. Ischemic stroke was defined on the basis of codes 434 or 436 (ICD-9) and I63 or I64 (ICD-10). Before inclusion, ischemic stroke events in the Swedish Hospital Discharge Register were validated with the use of the Stroke Register of Malmö, in which original medical records (including imaging studies, when available) were examined.¹⁸ We considered only ischemic strokes and excluded cases of intracerebral or subarachnoid hemorrhage. Death from coronary heart disease was defined on the basis of codes 412 and 414 (ICD-9) or I22–I23 and I25 (ICD-10) in the Swedish Cause of Death Register. Follow-up extended to December 31, 2003.

STATISTICAL ANALYSIS

For each subject, levels of LDL or HDL cholesterol were adjusted for age, sex, and diabetes status to create a multivariable-adjusted residual level of LDL or HDL cholesterol. The multivariable-adjusted residual lipid level for each subject served as the phenotype in genotype–phenotype association analyses. We conducted multivariable linear regression analyses to test the null hypothesis that LDL or HDL cholesterol residuals did not differ according

to an increasing copy number in the minor allele. Study subjects who were receiving lipid-lowering medication at the time of the baseline examination were excluded from these analyses.

On the basis of the results of the genotype–phenotype association analyses, we identified nine SNPs that were independently associated with lipid levels. Two SNPs that we initially tested were correlated with other SNPs in the panel and thus excluded.

We assumed that an individual SNP with a modest effect on lipid levels would have limited power to show an effect on incident cardiovascular events. We therefore constructed a genotype score on the basis of the number of unfavorable alleles (those associated with higher LDL cholesterol levels or lower HDL cholesterol levels) that were carried by each subject for each of the nine SNPs. We calculated crude incidence rates of first cardiovascular events according to strata of genotype scores.

We constructed multivariable proportional-hazards models to examine the association between the genotype score and the time to the first cardiovascular event, excluding subjects who had had a previous myocardial infarction or ischemic stroke. We first confirmed that the proportional-hazards assumption was met.¹⁹ The hazard ratio for the genotype score as a continuous measure was estimated in a model adjusting for all 14 available baseline covariates. Cumulative incidence curves were constructed according to the genotype score with the use of Cox regression analysis.

To evaluate the ability of the genotype score to classify risk, we plotted receiver-operating-characteristic (ROC) curves for the 14 baseline covariates with or without the genotype score.²⁰ We used the incidence of cardiovascular events at 10 years as the outcome. The C statistic, a measure of the area under the ROC curve, was calculated with and without the genotype score. We also evaluated the ability of the genotype score to reclassify risk across categories outlined by the Adult Treatment Panel III of the National Cholesterol Education Program²¹ with the use of two separate approaches suggested by Cook²² and Pencina et al.²³

All genetic association analyses were performed with the use of SPSS software, version 14.0. The tests for the proportional-hazards assumption were performed with the use of the survival pack-

age for R; the ROC curves and C statistics were generated with the use of the ROCR package for R (www.r-project.org).

RESULTS

BASELINE CHARACTERISTICS

Of the 6103 MDCS-CC subjects, 563 did not provide a baseline plasma sample, 56 did not have data regarding levels of either LDL or HDL cholesterol, 18 did not have DNA available, and 52 did not have genotypes for at least one SNP.²⁴ This left 5414 subjects who were eligible for the present investigation.

Baseline characteristics of the subjects are shown in Table 1. The subjects were all of self-reported European ancestry. The mean (\pm SD) age was 58 ± 6 years, and 59% were women. A small percentage of the subjects were receiving lipid-lowering therapy (2.3%) or had had a prevalent myocardial infarction or stroke (2.3%).

SNP—LIPID ASSOCIATIONS AND GENOTYPE SCORE

For validation of the association between SNPs and lipid levels, 5287 subjects who were not receiving lipid-lowering therapy at baseline were available. We validated the association between each of 11 SNPs with levels of either LDL cholesterol or HDL cholesterol (Table 2). The proportion of residual trait variance that was explained by each SNP was small, ranging from 0.1 to 2.5%. Among the SNPs that were associated with LDL cholesterol levels, the difference in LDL cholesterol level between homozygote classes ranged from 3 mg to 72 mg per deciliter (0.1 to 1.9 mmol per liter). Among the SNPs related to HDL cholesterol, the difference ranged from 3 mg to 5 mg per deciliter.

For the *APOB* and *LDLR* loci, there were two significant correlated SNPs at each locus. We selected the SNP with the stronger evidence for association at each locus (rs693 and rs1529729, respectively) for inclusion in the genotype score. Thus, the analyses relating genotypes with cardiovascular-event prediction consisted of nine SNPs from nine loci (Table 2).

We then constructed a genotype score for each subject by counting the number of unfavorable alleles carried by that subject for each SNP. With zero, one, or two unfavorable alleles possible for

Table 1. Baseline Characteristics of the Subjects.*

Characteristic	Men (N=2246)	Women (N=3168)
Mean age — yr	57.8 \pm 6.0	57.5 \pm 5.9
Parent or sibling with history of myocardial infarction — no. (%)	734 (32.7)	1288 (40.7)
Cholesterol — mg/dl		
Total	232 \pm 38	242 \pm 44
LDL	159 \pm 35	162 \pm 40
HDL	47 \pm 12	58 \pm 14
Triglycerides — mg/dl	134 \pm 72	112 \pm 59
Blood pressure — mm Hg		
Systolic	144 \pm 19	140 \pm 19
Diastolic	89 \pm 10	86 \pm 9
Diabetes mellitus — no. (%)	260 (11.6)	192 (6.1)
Body-mass index	26.2 \pm 3.5	25.6 \pm 4.2
Status of cigarette smoking — no. (%)		
Former	916 (40.8)	930 (29.4)
Never	608 (27.1)	1399 (44.2)
Current	464 (20.7)	630 (19.9)
C-reactive protein — mg/liter	2.7 \pm 4.5	2.5 \pm 4.3
Current drug therapy — no. (%)		
Lipid-lowering	67 (3.0)	60 (1.9)
Antihypertensive	308 (13.7)	434 (13.7)
Prevalent cardiovascular disease — no. (%) [†]	89 (4.0)	35 (1.1)

* Plus-minus values are means \pm SD. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for triglycerides to millimoles per liter, multiply by 0.01129. The body-mass index is the weight in kilograms divided by the square of the height in meters. LDL denotes low-density lipoprotein, and HDL high-density lipoprotein.

[†] Prevalent cardiovascular disease includes myocardial infarction or stroke before the baseline visit.

each of nine SNPs, the potential genotype score for each subject ranged from 0 to 18. Since only a small number of subjects had a score of 6 or less, these groups were a priori collapsed into one group for the purposes of data display and analysis. In a similar way, subjects with a score of 13 or more were collapsed into one group.

SNPs AND CARDIOVASCULAR EVENTS

Complete data for all covariates in the multivariable Cox regression analyses were available for 4232 subjects who had not had prevalent myocardial infarction or stroke. During a median of 10.6 years of follow-up, 137 men and 101 women in

Table 2. Association between Single-Nucleotide Polymorphisms (SNPs) and Low-Density Lipoprotein (LDL) or High-Density Lipoprotein (HDL) Cholesterol.*

SNP	Gene	SNP Type	MAF	Major-Allele Homozygotes			Heterozygotes			Minor-Allele Homozygotes			Percentage of Variance Explained	P Value†
				No. of Subjects	Genotype	Mean Cholesterol Level mg/dl	No. of Subjects	Genotype	Mean Cholesterol Level mg/dl	No. of Subjects	Genotype	Mean Cholesterol Level mg/dl		
LDL cholesterol														
rs693‡	APOB	Coding	0.48	1349	AA	167±38	2462	AG	160±37	1173	GG	157±38	0.9	2×10 ⁻¹¹
rs7575840	APOB	3' Downstream	0.35	1989	GG	158±37	2176	GT	161±38	568	TT	166±39	0.5	8×10 ⁻⁷
rs4420638‡	APOE cluster	5' Upstream	0.20	3291	AA	157±37	1621	AG	167±38	224	GG	173±38	1.7	3×10 ⁻²¹
rs12654264‡	HMGCR	Intronic	0.39	1911	AA	158±38	2405	AT	163±38	764	TT	162±38	0.2	0.002
rs1529729‡	LDLR	Intronic	0.44	1610	TT	160±38	2481	TC	161±38	993	CC	164±38	0.2	0.003
rs688	LDLR	Coding	0.42	1717	CC	160±39	2551	CT	161±38	860	TT	163±37	0.1	0.04
rs11591147‡	PCSK9	Coding	0.01	4885	GG	161±38	114	GT	146±30	2	TT	89±22	0.5	7×10 ⁻⁷
HDL cholesterol														
rs3890182‡	ABCA1	Intronic	0.13	3818	GG	54±15	1163	GA	53±14	82	AA	51±12	0.2	0.003
rs1800775‡	CETP	5' Upstream	0.49	1397	CC	51±13	2456	CA	54±14	1245	AA	56±15	2.5	2×10 ⁻²⁹
rs1800588‡	LIPC	5' Upstream	0.21	3157	CC	53±14	1754	CT	54±14	247	TT	57±16	0.8	4×10 ⁻¹⁰
rs328‡	LPL	Coding	0.09	4219	CC	53±14	863	CG	56±15	49	GG	58±16	0.9	3×10 ⁻¹²

* Plus-minus values are means ±SD. A total of 5287 subjects were available for analysis of the association between the SNP and the lipid level after the exclusion of subjects who were receiving lipid-lowering therapy. Listed are alleles for the SNP on the forward strand of the human genome reference sequence (from National Center for Biotechnology Information Build 35). To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. MAF denotes minor-allele frequency.

† Association analyses were conducted with the use of the multivariable-adjusted lipid level (adjusted for age, sex, and diabetes status) as the phenotype.

‡ This SNP was one of nine that were included in the genotype score. APOB rs7575840 and LDLR rs688 were not included in the genotype score, since these SNPs provided information redundant with that provided by APOB rs693 and LDLR rs1529729.

Table 3. Lipid Levels and Crude Incidence Rates of Cardiovascular Disease, According to Genotype Score.*

Variable	Genotype Score [†]								P for Trend
	≤6 (N=122)	7 (N=309)	8 (N=574)	9 (N=894)	10 (N=913)	11 (N=726)	12 (N=465)	≥13 (N=229)	
LDL cholesterol (mg/dl)	152±41	152±37	158±38	159±38	163±37	165±38	168±41	171±36	2×10 ⁻¹⁸
HDL cholesterol (mg/dl)	60±16	57±14	56±14	55±14	53±14	53±14	51±13	51±13	3×10 ⁻²⁴
Crude incidence rate per 1000 person-years	3.1	2.7	5.1	3.9	5.3	6.8	5.7	11.0	

* Plus-minus values are means ±SD. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. LDL denotes low-density lipoprotein, and HDL high-density lipoprotein.

[†] The genotype score represents the number of unfavorable alleles (the allele associated with higher LDL cholesterol or lower HDL cholesterol) at each of nine SNPs. These nine SNPs were *APOB* rs693, *APOE* cluster rs4420638, *HMGCR* rs12654264, *LDLR* rs1529729, *PCSK9* rs11591147, *ABCA1* rs3890182, *CETP* rs1800775, *LIPC* rs1800588, and *LPL* rs328.

this cohort had an incident cardiovascular event, prespecified as myocardial infarction (131 subjects), ischemic stroke (96), or death from coronary heart disease (11).

Table 3 shows the lipid levels and crude incidence rates for the first cardiovascular event stratified according to genotype score. LDL cholesterol levels increased from 152 mg per deciliter (3.9 mmol per liter) for those with a score of 6 or less to 171 mg per deciliter (4.4 mmol per liter) for those with a score of 13 or more (P for trend, 2×10⁻¹⁸). HDL cholesterol levels decreased from 60 mg per deciliter (1.6 mmol per liter) for those with a score of 6 or less to 51 mg per deciliter (1.3 mmol per liter) for those with a score of 13 or more (P for trend, 3×10⁻²⁴). There was an increase in the crude incidence rate of first cardiovascular events according to genotype score from 3.1 per 1000 person-years (among those with a score ≤6) to 11.0 per 1000 person-years (among those with a score ≥13).

MULTIVARIABLE ANALYSIS OF OUTCOME

After accounting for baseline lipid levels and a full set of available risk factors, we found that the genotype score remained associated with the first cardiovascular event (adjusted hazard ratio, 1.15 per copy of an unfavorable allele; 95% confidence interval [CI], 1.07 to 1.24; P=3×10⁻⁴ by multivariable analysis) (Table 4). As compared with subjects with a genotype score of 9 or less (43.2% of the sample), subjects with a score of 11 or more (33.6%) had an increased risk of a cardiovascular event by a factor of 1.63 (95% CI, 1.21 to 2.19; P=0.001). The predicted cumulative freedom from

Table 4. Multivariable Analysis of the Association between Genotype Score and the Time to First Cardiovascular Event.*

	Multivariable-Adjusted Hazard Ratio (95% CI)	P Value
Age, per SD	1.77 (1.52–2.07)	<0.001
Male sex	1.61 (1.20–2.17)	0.002
Parent or sibling with history of myocardial infarction	1.52 (1.17–1.97)	0.002
Cholesterol, per SD		
LDL	1.13 (0.99–1.29)	0.08
HDL	0.75 (0.61–0.91)	0.003
Log triglycerides, per SD	0.87 (0.73–1.04)	0.12
Blood pressure, per SD		
Systolic	1.29 (1.08–1.54)	0.005
Diastolic	1.16 (0.97–1.38)	0.11
Body-mass index, per SD	1.09 (0.94–1.25)	0.26
Diabetes mellitus	1.47 (1.02–2.13)	0.04
Status of cigarette smoking		<0.001 [†]
Former versus never	1.17 (0.85–1.59)	
Current versus never	2.00 (1.41–2.83)	
Log C-reactive protein, per SD	1.14 (0.99–1.30)	0.06
Drug therapy		
Lipid-lowering	1.29 (0.63–2.64)	0.48
Antihypertensive	1.46 (1.08–1.97)	0.01
Genotype score, per single unfavorable allele	1.15 (1.07–1.24)	<0.001

* A cardiovascular event was prespecified as myocardial infarction, ischemic stroke, or death from coronary heart disease. The body-mass index is the weight in kilograms divided by the square of the height in meters.

[†] The P value is for the overall comparison among the smoking-status groups.

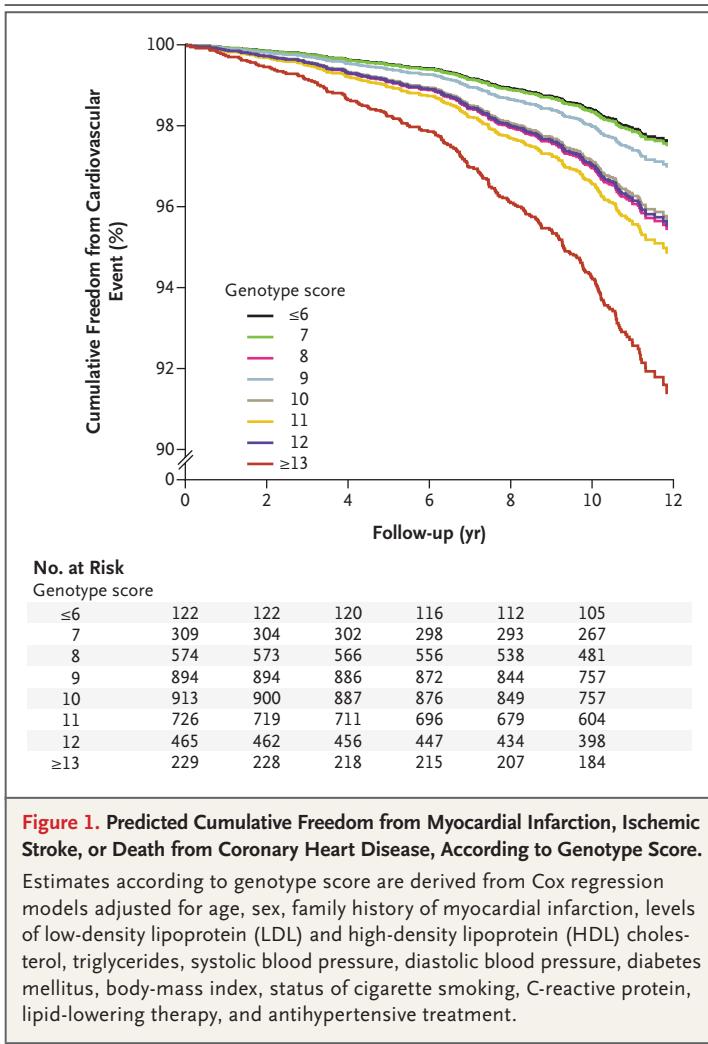


Figure 1. Predicted Cumulative Freedom from Myocardial Infarction, Ischemic Stroke, or Death from Coronary Heart Disease, According to Genotype Score.

Estimates according to genotype score are derived from Cox regression models adjusted for age, sex, family history of myocardial infarction, levels of low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol, triglycerides, systolic blood pressure, diastolic blood pressure, diabetes mellitus, body-mass index, status of cigarette smoking, C-reactive protein, lipid-lowering therapy, and antihypertensive treatment.

cardiovascular events according to the genotype score, with adjustment for baseline covariates, is shown in Figure 1.

We constructed ROC curves for the model incorporating established risk factors with and without inclusion of the genotype score, using the incidence of cardiovascular events at 10 years as the outcome (Fig. 2). The C statistic (area under the ROC curve) for total cardiovascular events was 0.80 for the risk model either with or without the genotype score.

Among subjects in the Adult Treatment Panel III intermediate-risk category (9% of the study sample), 26% were reclassified into a higher or lower risk category in the risk model containing the genotype score, as compared with the model without the genotype score (Table 1 of the Supplementary Appendix, available with the full text of this article at www.nejm.org). When we used

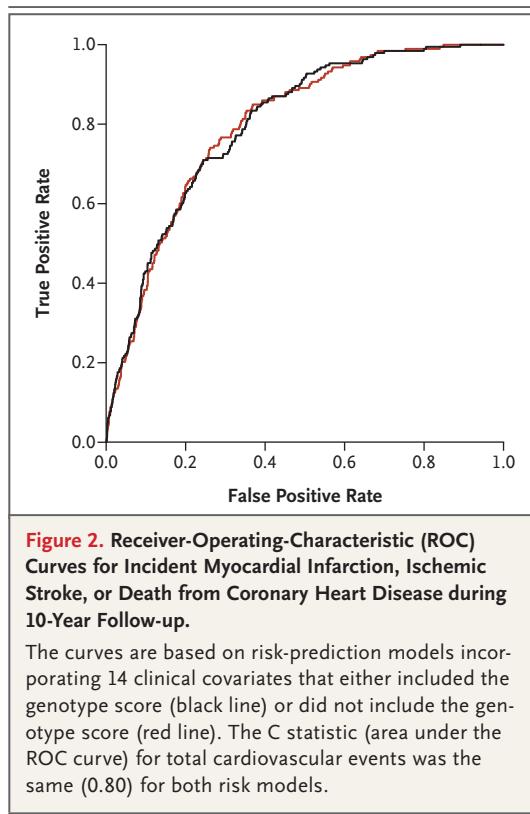
the net reclassification index²³ to account for the correct movement in categories (higher risk for subjects in whom cardiovascular disease subsequently developed and lower risk for subjects free of incident cardiovascular disease), there was a significant improvement in risk classification according to a model that incorporated the genotype score, as compared with one that did not ($P=0.01$) (Table 2 of the Supplementary Appendix). We also estimated a measure of improvement in model performance that did not depend on the choice of categories, called the integrated discrimination index.²³ A model with the genotype score showed a greater integrated discrimination index than did a model without the genotype score ($P=0.02$).

To further characterize the relationship between genotype and cardiovascular events, we performed several additional exploratory analyses. These included analyses comparing the incremental predictive value of the SNPs combined in the genotype score with that of the SNPs modeled as a set, analyses of the incremental predictive value of the genotype score for individual components of the cardiovascular event end point, analyses of whether specific SNPs that comprised the score appeared to drive the association with cardiovascular disease, and analyses of the separate predictive value of the SNPs related to LDL cholesterol levels and those related to HDL cholesterol levels. All the additional analyses are discussed in the Supplementary Appendix.

DISCUSSION

In a sample of approximately 5400 subjects from a community-based cohort, we observed that SNPs from nine loci were associated with the baseline level of LDL or HDL cholesterol; four of these associations represented a first replication. During more than 10 years of follow-up, we found that a panel of nine lipid-associated polymorphisms also predicted risk for a first cardiovascular event and provided independent information beyond that obtained from baseline lipid levels and other established risk factors. Nevertheless, the genotype score did not improve the accuracy of the clinical risk prediction, as assessed by the C statistic. Among subjects in the intermediate Adult Treatment Panel III risk category, 26% were reclassified into a higher or lower risk category when the genotype score was added to the other covariates.

These results highlight several important prin-



principles. First, for continuous, normally distributed quantitative traits like LDL and HDL cholesterol levels, multiple common alleles influence trait variation, with each allele conferring a modest effect.²⁵ Second, although each SNP exerts a modest effect, a combination of SNPs, in aggregate, can have a substantial influence on lipid levels. Third, lipid-associated SNPs may provide incremental information about cardiovascular risk beyond lipid levels.

Why might a panel of SNPs associated with levels of LDL or HDL cholesterol help predict cardiovascular outcomes even after accounting for baseline lipid levels? It has been proposed that a single measurement of blood lipids may inadequately reflect the lifetime exposure to altered lipid levels.^{12,26} Lipid measurements may be affected by dietary, diurnal, and random factors.²⁷ By contrast, genotypes are invariant over time, can be precisely measured on a single assay, and can exert an effect on lipid levels over a lifetime. An alternative possibility is that SNPs may influence atherogenesis through mechanisms unrelated to either LDL or HDL cholesterol.

When added to conventional risk factors, the current SNP panel did not improve the ability to

determine which patients will have cardiovascular disease and which will remain disease-free, although it did modestly improve risk prediction, as measured by clinical risk reclassification. It remains possible that with the addition of more SNPs related to LDL or HDL cholesterol or SNPs directly related to the risk of myocardial infarction (e.g., the chromosome 9p21 SNPs recently associated with coronary artery disease and myocardial infarction²⁸⁻³⁰), a panel of SNPs could prove useful in risk discrimination or targeting therapies.

Several limitations of our study deserve mention. Since this analysis is among the first in the genetic-association literature to report on a panel of lipid-modulating SNPs,³¹ we may have overestimated the effect size per unfavorable allele. Additional studies should clarify the true effect size per unfavorable allele.

The specific SNP set in this panel may need refinement. We selected a set of loci and SNPs on the basis of previous studies and a strong biologic rationale. However, we did not test other loci meeting these criteria. At each of the selected loci, the particular SNPs that we tested may not actually be causal but rather correlated to the causal SNP.

In addition, due to the rapid progress of sequencing and genomewide association approaches, a flood of new association discoveries is expected in the near future.³²⁻³⁴ It will be important to systematically test the contribution of such newly identified loci to future risk-prediction models.

Our study sample was entirely of European ancestry, and some of the SNPs we studied are not polymorphic in other ethnic groups (e.g., rs11591147 at *PCSK9* is monomorphic in persons of African ancestry).¹² This factor may limit the generalizability of our findings. However, several of the SNPs (including the *LPL* and *LIPC* variants) have been shown to relate to levels of LDL or HDL cholesterol in a similar fashion in both whites and nonwhites.³⁵

Moreover, in the interest of simplicity, we assigned an equal weight in our genotype score to each unfavorable allele, regardless of the strength of the effect on cholesterol. However, the effect on risk may be more accurately estimated by weighting the contribution of an allele to the score on the basis of its contribution to levels of LDL or HDL cholesterol and on the relationship of such levels to cardiovascular events. In addition, the effect of each SNP on the risk of cardiovascular disease may not relate to its effect on cholesterol in a linear fashion.

Finally, reclassification analysis may be affected in important ways by the choice of risk categories. For example, end points predicted in Adult Treatment Panel III differed from those in our study, so alternative risk-categorization schemes may be appropriate. To overcome this limitation, we also assessed the integrated discrimination index, a metric that is not dependent on the choice of categories.

In summary, in our analysis, nine SNPs from nine genes were found to be associated with levels of either LDL or HDL cholesterol. This panel of polymorphisms was independently associated with a risk of first myocardial infarction, ischemic stroke, or death from coronary heart disease. The SNP panel did not improve risk discrimination but did modestly improve clinical risk reclassification for individual subjects beyond standard clinical factors.

Supported by a Doris Duke Charitable Foundation Clinical Scientist Development Award, the Fannie E. Rippel Foundation, the Donovan Family Foundation, and a grant (K23-HL083102)

from the National Institutes of Health (NIH), all to Dr. Kathiresan; by the Swedish Medical Society, the Swedish Heart and Lung Foundation, the Ernhold Lundströms Research Foundation, the Mossfelt Foundation, and the Region Skane, all to Dr. Melander; by the Swedish Medical Research Council, the Medical Faculty of Lund University, Malmö University Hospital, the Albert Pålsson Research Foundation, and the Crafoord Foundation, all to Drs. Melander and Orho-Melander; by the Sigrid Juselius Foundation, the Finnish Diabetes Research Foundation, the Folkhalsan Research Foundation, the Clinical Research Institute of Helsinki University Central Hospital, and a Linné grant from the Swedish Research Council, all to Dr. Groop; by a Burroughs Wellcome Fund Clinical Scholar in Translational Research award and a Distinguished Clinical Scholar award from the Doris Duke Charitable Foundation, both to Dr. Altschuler; and by a Doris Duke Charitable Foundation Clinical Scientist Development Award, a Burroughs Wellcome Fund Career Award for Medical Scientists, and a grant (K23-HL080025) from the NIH, all to Dr. Newton-Cheh.

Dr. Kathiresan reports receiving consulting fees from Novartis and Pfizer; Dr. Hirschhorn, having an equity interest in and receiving consulting fees from Correlagen Diagnostics; Dr. Groop, receiving consulting fees from Tethys, lecture fees from Lilly and GlaxoSmithKline, and grant support from Novartis; and Dr. Altschuler, receiving consulting fees from Merck and MPM Capital and grant support from Novartis and having an equity interest in MPM Capital. No other potential conflict of interest relevant to this article was reported.

REFERENCES

- Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J III. Factors of risk in the development of coronary heart disease — six year follow-up experience: the Framingham Study. *Ann Intern Med* 1961; 55:33-50.
- Miller NE, Miller GJ. High-density lipoprotein and atherosclerosis. *Lancet* 1975; 1:1033.
- Namboodiri KK, Kaplan EB, Heuch I, et al. The Collaborative Lipid Research Clinics Family Study: biological and cultural determinants of familial resemblance for plasma lipids and lipoproteins. *Genet Epidemiol* 1985;2:227-54.
- Benn M, Nordestgaard BG, Jensen JS, Grande P, Sillesen H, Tybjaerg-Hansen A. Polymorphism in APOB associated with increased low-density lipoprotein levels in both genders in the general population. *J Clin Endocrinol Metab* 2005;90:5797-803.
- Boright AP, Connelly PW, Brunt JH, Morgan K, Hegele RA. Association and linkage of LDLR gene variation with variation in plasma low density lipoprotein cholesterol. *J Hum Genet* 1998;43:153-9.
- Kotowski IK, Pertsemliadis A, Luke A, et al. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet* 2006;78:410-22.
- Frikke-Schmidt R, Nordestgaard BG, Jensen GB, Tybjaerg-Hansen A. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest* 2004;114:1343-53.
- Boekholdt SM, Sacks FM, Jukema JW, et al. Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects. *Circulation* 2005;111:278-87.
- Guerra R, Wang J, Grundy SM, Cohen JC. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A* 1997;94:4532-7.
- Wittrup HH, Andersen RV, Tybjaerg-Hansen A, Jensen GB, Nordestgaard BG. Combined analysis of six lipoprotein lipase genetic variants on triglycerides, high-density lipoprotein, and ischemic heart disease: cross-sectional, prospective, and case-control studies from the Copenhagen City Heart Study. *J Clin Endocrinol Metab* 2006;91:1438-45.
- Sing CF, Davignon J. Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 1985;37:268-85.
- Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006;354:1264-72.
- Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331-6.
- Berglund G, Elmståhl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study: design and feasibility. *J Intern Med* 1993;233:45-51.
- Rip J, Nierman MC, Ross CJ, et al. Lipoprotein lipase S447X: a naturally occurring gain-of-function mutation. *Arterioscler Thromb Vasc Biol* 2006;26:1236-45.
- Altschuler D, Groop L, Hughes T. Whole genome scan for type 2 diabetes in a Scandinavian cohort. Cambridge, MA: Broad Institute, 2007. (Accessed February 12, 2008, at <http://www.broad.mit.edu/diabetes/scandinavs/index.html>.)
- Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.
- Jerntorp P, Berglund G. Stroke registry in Malmö, Sweden. *Stroke* 1992;23:357-61.
- Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81:515-26.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics* 2005; 21:3940-1.
- Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of

- High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA* 2001;285:2486-97.
22. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928-35.
23. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-72.
24. Pero RW, Olsson A, Berglund G, Janzon L, Larsson SA, Elmstahl S. The Malmö biological bank. *J Intern Med* 1993;233:63-7.
25. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinburgh* 1918;52:399-433.
26. Brown MS, Goldstein JL. Biomedicine: lowering LDL — not only how low, but how long? *Science* 2006;311:1721-3.
27. Cooper GR, Myers GL, Smith SJ, Schlant RC. Blood lipid measurements: variations and practical utility. *JAMA* 1992;267:1652-60. [Erratum, *JAMA* 1997;278:478.]
28. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007;316:1491-3.
29. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007;316:1488-91.
30. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007;357:443-53.
31. Humphries SE, Cooper JA, Talmud PJ, Miller GJ. Candidate gene genotypes, along with conventional risk factor assessment, improve estimation of coronary heart disease risk in healthy UK men. *Clin Chem* 2007;53:8-16.
32. Romeo S, Pennacchio LA, Fu Y, et al. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 2007;39:513-6.
33. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, or triglycerides in humans. *Nat Genet* 2008;40:189-97.
34. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;40:161-9.
35. Nettleton JA, Steffen LM, Ballantyne CM, Boerwinkle E, Folsom AR. Associations between HDL-cholesterol and polymorphisms in hepatic lipase and lipoprotein lipase genes are modified by dietary fat intake in African American and White adults. *Atherosclerosis* 2007;194(2):e131-e140.

Copyright © 2008 Massachusetts Medical Society.

FULL TEXT OF ALL JOURNAL ARTICLES ON THE WORLD WIDE WEB

Access to the complete text of the *Journal* on the Internet is free to all subscribers. To use this Web site, subscribers should go to the *Journal's* home page (www.nejm.org) and register by entering their names and subscriber numbers as they appear on their mailing labels. After this one-time registration, subscribers can use their passwords to log on for electronic access to the entire *Journal* from any computer that is connected to the Internet. Features include a library of all issues since January 1993 and abstracts since January 1975, a full-text search capacity, and a personal archive for saving articles and search results of interest. All articles can be printed in a format that is virtually identical to that of the typeset pages. Beginning 6 months after publication, the full text of all Original Articles and Special Articles is available free to nonsubscribers who have completed a brief registration.

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008;358:1240-9.

Polymorphisms Associated with Cholesterol and Risk of Cardiovascular Events

Sekar Kathiresan, M.D.,* Olle Melander, M.D., Ph.D.,† Dragi Anevski, Ph.D., Candace Guiducci, B.S., Noël P. Burt, B.S., Charlotta Roos, M.Sc., Joel N. Hirschhorn, M.D., Ph.D., Göran Berglund, M.D., Ph.D., Bo Hedblad, M.D., Ph.D., Leif Groop, M.D., Ph.D., David M. Altshuler, M.D., Ph.D., Christopher Newton Cheh, M.D., M.P.H.,† Marju Orho-Melander, Ph.D.*

*† denotes equal contribution

ONLINE SUPPLEMENTARY APPENDIX

RISK RECLASSIFICATION BY GENOTYPE SCORE

We evaluated the ability of genotype score to reclassify risk across National Cholesterol Education Program Adult Treatment Panel III risk categories using two separate approaches suggested by Cook et al. and Pencina et al. (see references in main text). Individuals in the Adult Treatment Panel III intermediate risk category constituted 9% of the study sample (370 of 4232 subjects), 26% (95) were reclassified into a higher or lower risk category in the risk model with genotype score compared to one without (**Supplementary Table 1**). When we used the ‘net reclassification index’ to account for the correct movement in categories – upward for individuals who subsequently developed cardiovascular disease and downward for individuals free of incident cardiovascular disease, there was a significant improvement in risk classification by a model that incorporated genotype score compared to one that did not ($P=0.01$, **Supplementary Table 2**).

Supplementary Table 1. Comparison of Observed and Predicted Risks for Cardiovascular Disease With and Without Genotype Score*

Model Without Genotype Score 10-Year Risk (%)	Model With Genotype Score 10-Year Risk (10%)			% Reclassified
	0 to ≤10%	>10 to ≤20%	>20%	
0 to ≤10%				
Total, n	3671	68	0	-
% [†]	98.0	1.8	0	1.8
Observed 10-year risk (%) [‡]	2.9	17.6	0	-
>10 to ≤20%				
Total, n	61	275	34	-
%	16.5	74.3	9.2	25.7
Observed 10-year risk (%)	8.2	14.5	14.7	-
>20%				
Total, n	0	23	100	-
%	0	18.7	81.3	18.7
Observed 10-year risk (%)	0	0	26.0	-

*Comparison is for the fully-adjusted model without and with genotype score (Models 3a and 3b as shown in Supplementary Table 1)

[†]Percent classified in each stratum in the model with genotype score

[‡]Observed proportion of individuals who developed cardiovascular disease at 10 years of follow-up

Supplementary Table 2. Reclassification Among Individuals Who Experience a Cardiovascular Disease Event and Those Who Do Not Experience a Cardiovascular Disease Event on Follow-Up

Model Without Genotype Score Frequency (Row per cent)	Model With Genotype Score			Total
	0 to \leq 10%	>10 to \leq 20%	>20%	
Participants who experience a cardiovascular disease event				
0 to \leq 10%	105 (89.7)	12 (10.3)	0 (0.0)	117
>10 to \leq 20%	5 (10.0)	40 (80.0)	5 (10.0)	50
>20%	0 (0.0)	0 (0.0)	26 (100.0)	26
Total	110	52	31	193
Participants who do not experience a cardiovascular disease event				
0 to \leq 10%	3566 (98.5)	56 (1.5)	0 (0.0)	3622
>10 to \leq 20%	56 (17.5)	235 (73.4)	29 (9.1)	320
>20%	0 (0.0)	23 (23.7)	74 (76.3)	97
Total	3622	314	103	4039

INCREMENTAL PREDICTIVE VALUE OF SNPS IN MULTIVARIABLE MODELS

To explore alternate approaches to modeling the impact of lipid-modulating SNPs on cardiovascular risk, we fitted three pairs of nested multivariable Cox proportional hazards regression models with or without entering all nine validated lipid-modulating SNPs. The SNPs were incorporated into models either as a set or as a continuous genotype score (see the main text for a description of the calculation of the genotype score).

We used a likelihood-ratio test to determine the P value for the addition of the SNPs, either as a set or as a genotype score, to a model without them. In each case, the $-2 \log$ likelihood from the larger model (clinical covariates and SNPs) was subtracted from that of the smaller model (clinical covariates only).

The first pair of models compared age and sex as covariates (Model 1a) with age, sex, and the SNPs (Model 1b). The second pair compared age, sex, LDL cholesterol, HDL cholesterol, and log triglycerides as covariates (Model 2a) with these clinical covariates plus the SNPs (Model 2b). The third pair compared a full set of 14 available covariates [age, sex, family history of myocardial infarction, LDL cholesterol, HDL cholesterol, triglycerides, systolic blood pressure, diastolic blood pressure, diabetes mellitus, body mass index, cigarette smoking, C-reactive protein, lipid lowering therapy, and anti-hypertensive treatment (Model 3a)] with a full set of risk factors plus the SNPs (Model 3b). In all of these analyses, the lipid-related SNPs were significantly associated with cardiovascular events (see **Supplementary Table 4**).

Supplementary Table 3. Prediction of First Cardiovascular Event in Models With and Without 9 Validated Lipid Polymorphisms

Model	P for Difference Between 'a' and 'b' Models			
	SNPs modeled as a set*		SNPs modeled as genotype score [†]	
	Change in χ^2	P	Change in χ^2	P
Model 1a: Age, sex	30.1	0.0004	19.5	0.00001
Model 1b: Age, sex, SNPs				
Model 2a: Age, sex, LDL cholesterol, HDL cholesterol, triglycerides	22.3	0.002	13.0	0.0003
Model 2b: Age, sex, LDL cholesterol, HDL cholesterol, triglycerides, SNPs				
Model 3a: Age, sex, parental or sibling history of myocardial infarction, LDL cholesterol, HDL cholesterol, log triglycerides, systolic blood pressure, diastolic blood pressure, body mass index, diabetes mellitus, cigarette smoking, log C-reactive protein, lipid lowering therapy, and anti-hypertensive treatment	23.6	0.005	13.6	0.0002
Model 3b: Age, sex, parental or sibling history of myocardial infarction, LDL cholesterol, HDL cholesterol, log triglycerides, systolic blood pressure, diastolic blood pressure, body mass index, diabetes mellitus, cigarette smoking, log C-reactive protein, lipid lowering therapy, anti-hypertensive treatment, and SNPs				

*The difference between models 1a and 1b; 2a and 2b; and 3a and 3b with the addition of 9 SNPs in aggregate was evaluated using the difference of -2log likelihoods, referenced to a chi square distribution with 9 degrees of freedom.

[†]The difference between model pairs with the addition of a single composite genotype score was determined as a chi square distribution with 1 degree of freedom.

SECONDARY ANALYSES

We considered individually the nonfatal myocardial infarction or death due to coronary heart disease and ischemic stroke outcomes. Genotype score was associated with incident myocardial infarction or death due to coronary heart disease (multivariable-adjusted hazard ratio 1.17, 95% confidence interval 1.07 – 1.29, $P=0.001$). Genotype score was also associated with incident ischemic stroke (multivariable-adjusted hazard ratio 1.13, 95% confidence interval 1.00 – 1.27, $P=0.047$).

To explore whether specific SNPs that comprised the score appeared to drive the association with cardiovascular disease, we used backward elimination (retention threshold $P<0.05$) in a model that included the full set of clinical covariates and 9 SNPs (Model 3). Five SNPs were retained as predictors of incident cardiovascular disease events: *LDLR* rs1529729 ($P=0.02$), *PCSK9* rs11591147 ($P=0.04$), *LIPC* rs1800588 ($P=0.005$), *LPL* rs328 ($P=0.049$), and *APOE* rs4420638 ($P=0.04$). A score composed of just these 5 SNPs was also associated with incident cardiovascular disease (multivariable-adjusted $P=0.0001$).

We evaluated the relative contributions of SNPs related to LDL cholesterol and those related to HDL cholesterol by constructing separate LDL and HDL cholesterol genotype scores. Both the LDL cholesterol genotype score (composed of 5 SNPs, score range from 0-10) and the HDL cholesterol genotype score (composed of 4 SNPs, score range from 0-8) were independently related to incident cardiovascular disease events in fully-adjusted models (HR 1.11, 95% CI 1.01-1.23, $P=0.03$ for LDL cholesterol genotype score; HR 1.22, 95% CI 1.08-1.37, $P=0.001$ for HDL cholesterol genotype score).

We tested for but did not observe a statistical interaction between genotype score and each of the following variables: age, gender, cigarette smoking, family history of myocardial infarction, and body mass index (each interaction $P > 0.05$).