American Journal of Epidemiology
Vol. 171, No. 1
DOI: 10.1093/aje/kwp377
Advance Access publication:
December 2, 2009

Original Contribution

Genetic Variants Identified in a European Genome-Wide Association Study That Were Found to Predict Incident Coronary Heart Disease in the Atherosclerosis Risk in Communities Study

Jan Bressler, Aaron R. Folsom, David J. Couper, Kelly A. Volcik, and Eric Boerwinkle*

* Correspondence to Dr. Eric Boerwinkle, Human Genetics Center, University of Texas Health Science Center at Houston, P.O. Box 20334, Houston, TX 77225-0334 (e-mail: eric.boerwinkle@uth.tmc.edu).

Initially submitted April 13, 2009; accepted for publication September 18, 2009.

In 2007, the Wellcome Trust Case Control Consortium (WTCCC) performed a genome-wide association study in 2,000 British coronary heart disease (CHD) cases and 3,000 controls after genotyping 469,557 single nucleotide polymorphisms (SNPs). Seven variants associated with CHD were initially identified, and 5 SNPs were later found in replication studies. In the current study, the authors aimed to determine whether the 12 SNPs reported by the WTCCC predicted incident CHD through 2004 in a biracial, prospective cohort study (Atherosclerosis Risk in Communities) comprising 15,792 persons aged 45–64 years who had been selected by probability sampling from 4 different US communities in 1987–1989. Cox proportional hazards models with adjustment for age and gender were used to estimate CHD hazard rate ratios (HRRs) over a 17-year period (1,362 cases in whites and 397 cases in African Americans) under an additive genetic model. The results showed that 3 SNPs in whites (rs599839, rs1333049, and rs501120; HRRs were 1.10 ($P = 0.044$), 1.14 ($P < 0.001$), and 1.14 ($P = 0.030$), respectively) and 1 SNP in African Americans (rs7250581; HRR = 1.60, $P = 0.05$) were significantly associated with incident CHD. This study demonstrates that genetic variants revealed in a case-control genome-wide association study enriched for early disease onset may play a role in the genetic etiology of CHD in the general population.

coronary disease; genetic variation; genetics; genomics; heart diseases; myocardial infarction; polymorphism, single nucleotide

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CHD, coronary heart disease; HDL, high density lipoprotein; LDL, low density lipoprotein; SDF-1, stromal cell-derived factor 1; SNP, single nucleotide polymorphism; WTCCC, Wellcome Trust Case Control Consortium.

Editor's note: *An invited commentary on this article appears on page 24.*

Genetic studies designed to identify sequence variants implicated in complex disorders have recently focused on genome-wide association studies of the relations between large numbers of single nucleotide polymorphisms (SNPs) measured simultaneously and disease risk. The development of high-density genotyping arrays and the availability of the HapMap database (http://hapmap.ncbi.nlm.nih.gov/), which describes common patterns of sequence variation in 4 different populations (1), have facilitated this approach and enabled the successful detection of SNPs that contribute to the etiology of such common diseases as type 2 diabetes (2–7), obesity (8–11), coronary heart disease (CHD) (2, 12, 13), and Alzheimer's disease (14, 15). In many of these studies, SNPs in genes previously reported to be associated with the phenotype of interest, such as transcription factor 7-like 2 and type 2 diabetes (2–5, 16), as well as SNPs in genes or pathways not previously known to be involved in disease pathogenesis, were identified. The replicated association of the fat mass- and obesity-associated gene with obesity is one such example and demonstrates the utility of this approach (9–11).

The Wellcome Trust Case Control Consortium (WTCCC) was formed in Great Britain to carry out an experiment in

which approximately 2,000 cases for each of 7 complex diseases and 3,000 shared controls were genotyped using the Affymetrix GeneChip 500K Mapping Array Set (2). Seven SNPs showing either strong ($P < 5 \times 10^{-7}$) or moderate ($P = 10^{-5}$–10^{-7}) association with CHD were identified in a sample enriched for premature myocardial infarction or coronary revascularization occurring before the 66th birthday. A second report was subsequently published by the same consortium that presented evidence for replication in the German Myocardial Infarction Family Study for 3 genetic variants, including 2 of the 7 most likely susceptibility loci for CHD from the first report (17). Four additional loci were then identified when a combined analysis of the data from the original WTCCC study and the German Myocardial Infarction Family Study was undertaken. All of the SNPs are located on separate chromosomes or chromosomal regions, so they are not in linkage disequilibrium with each other; information concerning putative candidate genes in these regions and their function is provided in the Appendix Table. Our aim in the current study was to determine whether any of the 12 SNPs found to contribute to coronary artery disease susceptibility in the WTCCC study were associated with incident CHD in a large, biracial, population-based cohort.

MATERIALS AND METHODS

Study population

The Atherosclerosis Risk in Communities (ARIC) Study is a prospective longitudinal investigation of the development of atherosclerosis involving 15,792 persons aged 45–64 years who were selected by probability sampling from 4 different communities in the United States. At the time of recruitment in 1987–1989, the participants were residents of Forsyth County, North Carolina; Jackson, Mississippi (African Americans only); the northwestern suburbs of Minneapolis, Minnesota; or Washington County, Maryland. Participants in the ARIC Study were excluded from analysis if they were African Americans from the Minnesota or Maryland field center ($n = 48$), because of the small numbers recruited from these sites, or if they were neither African-American nor white ($n = 43$). Additional exclusion criteria were: missing genotype data for all sequence variants ($n = 540$); a positive or unknown history of CHD, prevalent stroke, or a history of transient ischemic attack at the initial clinic visit ($n = 1{,}446$); and participant refusal for the use of DNA ($n = 40$). When these persons were excluded, the study sample consisted of 13,675 participants.

The initial clinical examination, referred to herein as visit 1, was carried out in 1987–1989 and included an interview designed to elicit information about the presence of cardiovascular disease risk factors, socioeconomic status, and family medical history. Incidence of CHD was determined over a period of 17 years by telephoning participants annually and by surveying discharge lists from local hospitals and death certificates from state vital statistics offices for potential cardiovascular events (18). Incident CHD cases were defined as those involving a definite or probable myocardial infarction, a silent myocardial infarction between examinations determined by electrocardiography, a definite CHD death, or a coronary revascularization procedure. Participants were followed for a mean of 13.1 years. A total of 1,759 CHD cases were identified through December 31, 2004.

All of the persons enrolled in the ARIC Study provided written informed consent, and the study design and methods were approved by institutional review boards at the collaborating medical centers. A detailed description of the ARIC Study has been published previously (19).

Baseline clinical and laboratory measurements

The prevalence of diabetes at visit 1 was defined as a fasting glucose level $\geq$126 mg/dL, a nonfasting glucose level $\geq$200 mg/dL, and/or self-reported physician diagnosis of or treatment for diabetes. Body mass index (weight (kg)/height $(m)^2$) was calculated from height and weight measurements obtained at the baseline examination. Seated blood pressure was measured 3 times using a random-zero sphygmomanometer, and the last 2 measurements were averaged. Hypertension was defined as systolic blood pressure $\geq$140 mm Hg or diastolic blood pressure $\geq$90 mm Hg or current use of antihypertensive medication. Cigarette smoking status was analyzed by comparing current smokers with persons who had formerly or never smoked. Plasma total cholesterol level was measured by an enzymatic method (20), and the portion of low density lipoprotein (LDL) cholesterol was calculated (21). The level of high density lipoprotein (HDL) cholesterol was measured after dextran-magnesium precipitation of non-HDL cholesterol (22).

Genotype determination

Genotyping of 12 SNPs previously identified either in the genome-wide association study of coronary artery disease carried out in the United Kingdom (2) or in the German Myocardial Infarction Family Study (17) was performed using the TaqMan assay (Applied Biosystems, Foster City, California). The oligonucleotide sequences for polymerase chain reaction primers and TaqMan probes are available upon request from the authors. Allele detection was performed using the ABI Prism 7700 Sequence Detection System (Applied Biosystems). The genotype call rate, or the percentage of samples to which a genotype was assigned, was determined prior to exclusion of individuals from the analysis and ranged from 93.8% for rs2943634 to 94.8% for rs688034. After the application of all exclusion criteria, the proportion of missing genotype data in the final study sample did not exceed 2.4% for any of the genetic variants. We also assessed the genotyping success rate by analyzing the concordance between genotypes for pairs of blind duplicates included with the DNA samples from the study participants. Kappa coefficients (23), an index of the percentage of agreement between measurements, corrected for agreement occurring by chance, were calculated for each SNP and ranged from 0.96 to 0.99.

Statistical analysis

Statistical analysis was carried out using the Stata statistical software program, version 9.0 (Stata Corporation,

Table 1. Clinical Characteristics of the Study Population by Race and Incident Coronary Heart Disease Case Status, Atherosclerosis Risk in Communities Study, 1987–1989

	African Americans					Whites				
	Incident CHD Cases (n = 397)		Noncases (n = 3,206)		P Value[a]	Incident CHD Cases (n = 1,362)		Noncases (n = 8,710)		P Value[a]
	No. or Mean	%	No. or Mean	%		No. or Mean	%	No. or Mean	%	
Age, years	54.8 (5.77)[b]		53.1 (5.78)		<0.001	55.5 (5.48)		53.9 (5.71)		<0.001
Body mass index[c]	29.9 (5.71)		29.6 (6.15)		0.351	27.9 (4.67)		26.8 (4.84)		<0.001
Cholesterol level, mmol/L										
HDL cholesterol	1.27 (0.37)		1.45 (0.46)		<0.001	1.11 (0.32)		1.35 (0.44)		<0.001
LDL cholesterol	3.89 (1.23)		3.50 (1.08)		<0.001	3.88 (0.98)		3.48 (0.96)		<0.001
Total cholesterol	5.85 (1.32)		5.50 (1.14)		<0.001	5.80 (1.07)		5.50 (1.04)		<0.001
Male gender	197	49.6	1,131	35.3	<0.001	931	63.4	3,643	41.8	<0.001
Diabetes mellitus	125	32.0	523	16.6	<0.001	267	19.7	557	6.4	<0.001
Hypertension	285	71.8	1,662	52.1	<0.001	520	38.3	2,038	23.5	<0.001
Current smoking	160	40.5	893	27.9	<0.001	424	31.1	2,041	23.4	<0.001

Abbreviations: CHD, coronary heart disease; HDL, high density lipoprotein; LDL, low density lipoprotein.

[a] The significance of the difference between group mean values was determined by t test; categorical variables were evaluated by Pearson's chi-squared test.

[b] Numbers in parentheses, standard deviation.

[c] Weight (kg)/height $(m)^2$.

College Station, Texas). The hypothesis that observed genotypes were in Hardy-Weinberg equilibrium was tested in noncases using a χ^2 goodness-of-fit test. The proportions, mean values, and standard deviations were calculated for established CHD risk factors for both the incident CHD cases and the comparison group of persons who did not meet the case definition. Cox proportional hazards models were used to estimate hazard rate ratios for CHD. For analyses of CHD cases, follow-up time intervals were defined as the time between visit 1 and the date of the first CHD event. For noncases, follow-up continued through one of the following dates: December 31, 2004, the date of death, or the date of last contact if the participant was lost to follow-up.

Covariates were assessed for statistical significance in the models using the Wald χ^2 statistic. The results of all statistical analyses are presented separately by self-reported racial group. A 2-sided P value less than 0.05 was considered statistically significant, with no correction applied for multiple comparisons, since there were strong prior odds of association based on the results of the WTCCC case-control studies (2, 17). Power analyses were conducted using the Cox regression module of the Power Analysis and Sample Size program (24).

RESULTS

Proportions, means, and standard deviations for the established CHD risk factors are shown in Table 1. The mean values for all clinical characteristics at visit 1 differed significantly between participants who developed CHD and noncases for both whites and African Americans, with the exception of mean body mass index for African-American participants. For both racial groups, CHD cases included a higher frequency of males, persons affected by diabetes and hypertension, and smokers in comparison with noncases.

Genotype frequencies for the polymorphisms identified in the WTCCC case-control study differed significantly in noncases between whites and African Americans (data not shown), so subsequent statistical analyses were performed separately by race. The allele and genotype frequencies for all genetic variants were in accordance with Hardy-Weinberg equilibrium expectations for whites. One of the 12 SNPs did not meet Hardy-Weinberg expectations at a P value of 0.05 for African Americans in the study sample (rs7250581, $P = 0.01$); this may be attributable to the low minor allele frequency observed for this genetic variant. Table 2 shows genotype frequencies for the WTCCC SNPs in the ARIC cohort after stratification by race. There were significant differences between incident CHD cases and noncases for 3 of the SNPS in whites (rs1333049, rs501120, and rs8055236).

Table 3 shows results from Cox proportional hazards models used to estimate hazard rate ratios for incident CHD, stratified by both SNP and racial group. An additive genetic model was chosen for these analyses for consistency with the WTCCC reports (2, 17), and all of the genotypes for the individual SNPs were coded with respect to the risk allele for coronary artery disease determined by the WTCCC so that direction of effect could be easily compared. After adjustment for age and gender (model 1), rs7250581 was a nominally significant predictor of incident CHD for African Americans (hazard rate ratio = 1.60, $P = 0.05$), while 3 of the SNPS (rs599839, rs1333049, and rs501120) were associated with incident CHD for whites (hazard rate ratios were 1.10 ($P = 0.044$), 1.14 ($P < 0.001$), and 1.14 ($P = 0.030$), respectively). When

Table 2. Genotype Frequencies for Genetic Variants Identified in a WTCCC Study[a] of Coronary Artery Disease, by Race, Atherosclerosis Risk in Communities Study, 1987–1989

dbSNP[b] ID No.	WTCCC Risk/Minor Allele	Chromosome	African Americans						Whites					
			CHD Cases		Noncases		q^c	*P* Value[d]	CHD Cases		Noncases		q^c	*P* Value[d]
			No.	%	No.	%			No.	%	No.	%		
rs599839	A/G	1p13					0.73	0.774					0.23	0.204
AA			33	8.5	236	7.6			821	61.7	5,062	59.2		
AG			151	38.9	1,194	38.4			444	33.4	3,017	35.3		
GG			204	52.6	1,676	54.0			65	4.9	469	5.5		
rs17465637	C/A	1q14					0.75	0.881					0.28	0.595
CC			28	7.2	204	6.5			712	53.5	4,462	52.0		
AC			142	36.5	1,152	36.9			519	39.0	3,457	40.3		
AA			219	56.3	1,769	56.6			99	7.5	656	7.7		
rs17672135	T/C	1q43					0.10	0.766					0.12	0.112
TT			308	80.2	2,550	81.4			1,051	78.2	6,609	76.8		
CT			71	18.5	552	17.6			266	19.8	1,871	21.7		
CC			5	1.3	31	1.0			27	2.0	128	1.5		
rs2943634	C/A	2q36					0.59	0.883					0.34	0.361
CC			62	16.1	535	17.1			598	44.9	3,746	44.1		
AC			187	48.6	1,500	48.0			596	44.8	3,764	44.3		
AA			136	35.3	1,091	34.9			137	10.3	988	11.6		
rs383830	T/A	5q21					0.37	0.167					0.20	0.671
TT			176	45.5	1,265	40.5			873	65.2	5,513	64.0		
AT			160	41.3	1,412	45.2			410	30.6	2,742	31.8		
AA			51	13.2	449	14.3			56	4.2	360	4.2		
rs6922269	A/A	6q25					0.55	0.340					0.27	0.491
GG			68	17.4	637	20.3			717	53.4	4,569	53.2		
AG			206	52.7	1,555	49.6			517	38.5	3,403	39.6		
AA			117	29.9	944	30.1			108	8.1	621	7.2		
rs1333049	C/C	9p21					0.23	0.384					0.47	0.007
GG			239	62.2	1,849	59.3			321	23.9	2,389	27.8		
CG			120	31.3	1,083	34.8			687	51.1	4,276	49.7		
CC			25	6.5	184	5.9			336	25.0	1,939	22.5		
rs501120	T/C	10q11					0.41	0.606					0.13	0.042
TT			133	34.4	1,064	34.5			1,050	78.1	6,465	75.2		
CT			196	50.8	1,505	48.8			277	20.6	1,977	23.0		
CC			57	14.8	513	16.7			17	1.3	157	1.8		
rs17228212	C/C	15q22					0.12	0.631					0.28	0.696
TT			304	77.2	2,414	77.4			684	50.7	4,446	51.7		
CT			86	21.8	653	21.0			546	40.5	3,443	40.1		
CC			4	1.0	50	1.6			118	8.8	704	8.2		
rs8055236	G/T	16q23					0.54	0.985					0.19	0.002
GG			87	22.4	684	22.0			873	65.0	5,674	66.1		
GT			189	48.6	1,522	49.0			399	29.7	2,630	30.6		
TT			113	29.0	901	29.0			71	5.3	286	3.3		
rs7250581	G/A	19q12					0.03	0.172					0.18	0.243
GG			374	95.9	2,922	93.6			920	68.7	5,717	66.6		
AG			16	4.1	190	6.1			380	28.4	2,564	29.9		
AA			0	0.0	8	0.3			39	2.9	301	3.5		
rs688034	T/T	22q12					0.10	0.521					0.33	0.365
CC			317	81.9	2,547	81.3			624	46.5	3,891	45.1		
CT			64	16.5	554	17.7			561	41.8	3,778	43.8		
TT			6	1.6	31	1.0			157	11.7	952	11.1		

Abbreviations: CHD, coronary heart disease; ID, identification; rs, reference SNP; SNP, single nucleotide polymorphism; WTCCC, Wellcome Trust Case Control Consortium.

[a] WTCCC, 2007 (2).

[b] The National Center for Biotechnology Information's SNP database (http://www.ncbi.nlm.nih.gov/SNP/).

[c] Frequency among noncases of the minor allele identified in the WTCCC study (2).

[d] Significance of the difference in genotype frequencies between incident CHD cases and noncases, evaluated by Pearson's chi-squared test.

Table 3. Hazard Rate Ratios for Incident Coronary Heart Disease From Cox Regression Models, by Race, Atherosclerosis Risk in Communities Study, 1987–2004

dbSNP[a] ID No.	African Americans						Whites					
	Model 1[b]			Model 2[c]			Model 1[b]			Model 2[c]		
	HRR	95% CI	*P* Value[d]	HRR	95% CI	*P* Value[d]	HRR	95% CI	*P* Value[d]	HRR	95% CI	*P* Value[d]
rs599839	1.06	0.91, 1.24	0.425	1.07	0.91, 1.25	0.397	1.10	1.00, 1.21	0.044	1.02	0.93, 1.12	0.684
rs17465637	1.02	0.87, 1.19	0.845	1.01	0.85, 1.19	0.942	1.04	0.95, 1.13	0.421	1.04	0.95, 1.13	0.439
rs17672135	0.90	0.72, 1.14	0.395	0.89	0.70, 1.14	0.361	1.06	0.94, 1.19	0.333	1.05	0.93, 1.19	0.405
rs2943634	0.98	0.85, 1.13	0.797	1.00	0.86, 1.15	0.954	1.05	0.97, 1.14	0.243	1.02	0.94, 1.11	0.553
rs383830	1.11	0.96, 1.28	0.166	1.06	0.91, 1.23	0.431	1.05	0.95, 1.15	0.344	1.03	0.94, 1.14	0.493
rs6922269	1.04	0.90, 1.19	0.587	1.05	0.90, 1.21	0.537	1.03	0.95, 1.12	0.487	1.01	0.93, 1.11	0.734
rs1333049	0.93	0.79, 1.11	0.426	0.92	0.77, 1.10	0.372	1.14	1.06, 1.24	<0.001	1.17	1.08, 1.26	<0.001
rs501120	1.07	0.92, 1.24	0.380	1.08	0.92, 1.25	0.342	1.14	1.01, 1.28	0.030	1.18	1.05, 1.33	0.006
rs17228212	0.96	0.78, 1.20	0.746	0.96	0.77, 1.19	0.723	1.05	0.96, 1.14	0.283	1.03	0.94, 1.12	0.512
rs8055236	1.00	0.87, 1.15	0.992	1.02	0.88, 1.18	0.790	0.92	0.84, 1.01	0.096	0.94	0.85, 1.03	0.201
rs7250581	1.60	0.99, 2.59	0.054	1.39	0.84, 2.31	0.204	1.07	0.97, 1.19	0.167	1.05	0.95, 1.16	0.361
rs688034	0.99	0.78, 1.25	0.921	0.98	0.78, 1.25	0.904	0.96	0.89, 1.04	0.351	0.94	0.86, 1.02	0.131

Abbreviations: CI, confidence interval; HRR, hazard rate ratio; ID, identification; rs, reference SNP; SNP, single nucleotide polymorphism.

[a] The National Center for Biotechnology Information's SNP database (http://www.ncbi.nlm.nih.gov/SNP/).

[b] Results were adjusted for age and gender.

[c] Results were adjusted for age, gender, body mass index, smoking, diabetes case status, hypertension case status, high density lipoprotein cholesterol, and low density lipoprotein cholesterol.

[d] Analysis was performed using an additive genetic model, coded with respect to the risk allele identified by the Wellcome Trust Case Control Consortium (2).

further adjustments were made for a group of CHD risk factors including body mass index, smoking, diabetes status, hypertension status, and HDL and LDL cholesterol levels (model 2), the associations with rs7250581 for African Americans and rs599839 for whites were no longer detected. Since rs599839 was reported to be significantly associated with serum LDL cholesterol concentration in 2 genome-wide association studies conducted in European Caucasian populations (25, 26), a third model incorporating only age, gender, and LDL cholesterol was fitted, and the association with incident CHD was eliminated (hazard rate ratio = 1.03, 95% confidence interval: 0.94, 1.13; $P = 0.556$). In addition, if the results of the comparison of genotype frequencies between cases and noncases for each genetic variant are also considered in the evaluation of the effect of adjustment for covariates (Table 2), only rs1333049 and rs501120 are consistently associated with coronary artery disease case status in whites in all analyses. For rs7250581 in African Americans, the further addition of either body mass index, smoking, diabetes status, or hypertension status to the model adjusted for age and gender resulted in the absence of a significant association with incident CHD, while the inclusion of either LDL or HDL cholesterol did not affect the observed relation (data not shown).

DISCUSSION

Coronary artery disease and its clinical sequelae, including myocardial infarction, together constitute the single greatest cause of death worldwide (27, 28). CHD has a complex etiology, with multiple genes and environmental factors believed to be involved in its pathogenesis. Although the heritability of CHD is estimated to be 40%–60% (29), with a risk to first-degree relatives that is 5–7 times higher than that for members of the general population (30), the reproducible identification of genetic variants underlying this increased susceptibility has been difficult.

A genome-wide association study conducted by the WTCCC compared the frequency of 469,557 SNPs in approximately 2,000 cases and 3,000 shared controls for 7 different complex diseases, including CHD (2). The CHD cases in the WTCCC study included only persons with premature cardiovascular events occurring before the age of 66 years, to maximize the likelihood of detecting a genetic component contributing to disease causation. Seven novel variants were identified, with the strongest association being found for an SNP (rs1333049) located in the same region of chromosome 9p21 that was independently reported to harbor polymorphisms conferring increased risk of myocardial infarction in 5 different Caucasian populations (12, 13). However, the ascertainment scheme chosen by the WTCCC may also have resulted in the choice of CHD cases that were not representative of the population from which they were drawn; thus, SNPs identified using this type of study design may not be associated with CHD outside of the selected subgroup. Our aim in the current investigation was to determine whether any of the 7 SNPs associated with CHD in the initial case-control study (2) and the 5 additional

Table 4. Cox Regression Hazard Rate Ratios for Incident Coronary Heart Disease That Were Detectable at 80% Power, by Race, Atherosclerosis Risk in Communities Study, 1987–2004

Cox Regression Model	dbSNP[a] ID No.	African Americans		Whites	
		β	HRR	β	HRR
Model 1[b]	rs599839	0.2216	1.248	0.1272	1.136
Model 2[c]	rs599839	0.2220	1.249	0.1280	1.137
Model 1	rs17465637	0.2274	1.255	0.1201	1.128
Model 2	rs17465637	0.2278	1.256	0.1202	1.128
Model 1	rs17672135	0.3322	1.394	0.1632	1.177
Model 2	rs17672135	0.3326	1.395	0.1632	1.177
Model 1	rs2943634	0.2014	1.223	0.1133	1.120
Model 2	rs2943634	0.2016	1.223	0.1133	1.120
Model 1	rs383830	0.2027	1.225	0.1335	1.143
Model 2	rs383830	0.2030	1.225	0.1335	1.143
Model 1	rs6922269	0.2007	1.222	0.1208	1.128
Model 2	rs6922269	0.2010	1.223	0.1208	1.128
Model 1	rs1333049	0.2320	1.261	0.1074	1.113
Model 2	rs1333049	0.2323	1.261	0.1075	1.113
Model 1	rs501120	0.2036	1.226	0.1589	1.172
Model 2	rs501120	0.2039	1.226	0.1590	1.172
Model 1	rs17228212	0.3040	1.355	0.1184	1.126
Model 2	rs17228212	0.3042	1.356	0.1185	1.126
Model 1	rs8055236	0.1977	1.219	0.1371	1.147
Model 2	rs8055236	0.1979	1.219	0.1371	1.147
Model 1	rs7250581	0.5567	1.745	0.1385	1.149
Model 2	rs7250581	0.5581	1.747	0.1386	1.149
Model 1	rs688034	0.3322	1.394	0.1135	1.120
Model 2	rs688034	0.3326	1.395	0.1135	1.120

Abbreviations: HRR, hazard rate ratio; ID, identification; rs, reference SNP; SNP, single nucleotide polymorphism.

[a] The National Center for Biotechnology Information's SNP database (http://www.ncbi.nlm.nih.gov/SNP/).

[b] Results were adjusted for age and gender.

[c] Results were adjusted for age, gender, body mass index, smoking, diabetes case status, hypertension case status, high density lipoprotein cholesterol, and low density lipoprotein cholesterol.

SNPs identified in replication studies conducted by the WTCCC (17) predicted incident CHD in the large, biracial, population-based ARIC cohort.

Two of the sequence variants described by the WTCCC in case-control studies (rs1333049 and rs501120) were significantly associated with incident CHD for white participants in the ARIC Study. Results from the Cox proportional hazards models used to estimate the risk of CHD showed that the 2 intergenic SNPs were independently associated with CHD even after adjustment for multiple risk factors. One of these SNPs (rs1333049) maps to a region of chromosome 9p21 previously reported to be associated with risk of myocardial infarction among whites in the ARIC cohort (12). The 58-kilobase genomic interval defining the risk allele is located near the cyclin-dependent kinase inhibitor 2A, cyclin-dependent kinase inhibitor 2B, and cyclin-dependent kinase inhibitor antisense RNA loci, while there are no annotated genes or micro-RNAs within the region itself (Appendix Table).

The rs501120 variant is located approximately 127 kilobases downstream of the gene encoding stromal cell-derived factor 1 (*SDF-1*) on chromosome 10q11. SDF-1 is a member of the family of chemoattractant cytokines known as chemokines and is the ligand for cell-surface chemokine receptor 4. *SDF-1* has been implicated in a wide variety of biologic processes, including trafficking of hematopoietic stem and progenitor cells (31), migration of lymphocytes and monocytes (32), and the recruitment of endothelial progenitor cells derived from bone marrow to ischemic tissues in animal models (33, 34). *SDF-1* has also been reported to be involved in the induction of platelet aggregation, with high expression in smooth muscle cells, endothelial cells, and macrophages in human atherosclerotic plaques (35). Although the latter function suggests that the rs501120 T allele could confer susceptibility to CHD by playing a role in the regulation of thrombus formation after plaque rupture, this would require rs501120 to exert its effect at a considerable distance from the *SDF-1* coding region (Appendix Table).

One additional SNP (rs599839) was associated with incident CHD in whites when the results were adjusted only for age and gender, but this relation was abolished after either LDL or HDL cholesterol was added as an additional covariate ($P = 0.556$ and $P = 0.109$, respectively). Since the rs599839 polymorphism was recently reported to be significantly associated with serum LDL cholesterol concentration in 2 genome-wide association studies conducted in European Caucasian populations (25, 26), these results suggest that the influence of this SNP on CHD risk in the ARIC cohort may operate through its effect on plasma lipid levels. The noncoding rs599839 variant is located near 3 coding genes on chromosome 1p13, including sortilin, which is involved in the receptor-mediated binding of lipoprotein lipase on the surface of adipocytes (36). A second SNP that is in linkage disequilibrium with rs599839 (rs646776; $r^2 = 0.89$ in HapMap Utah residents with Northern and Western European ancestry) has been shown to be correlated with the expression of sortilin in human liver cells (25), providing a possible link to known events in lipoprotein metabolism (Appendix Table).

When incident CHD was examined in African Americans, one of the 12 SNPs (rs7250581) showed a nominally significant relation after adjustment for age and gender that was abrogated after the addition of multiple covariates known to be associated with cardiovascular disease. However, since the observed genotype frequencies for this variant did not meet Hardy-Weinberg equilibrium expectations, the possibility of a false-positive association must also be considered. As Table 4 shows, there was 80% power to detect a hazard rate ratio of 1.2–1.3 for 8 of the polymorphisms studied in African Americans, and a hazard rate ratio of 1.1–1.2 could be detected for all of the genetic variants in whites. Therefore, the failure to identify significant associations between most of the SNPs described in a population of Northern European origin in the WTCCC case-control study and

incident CHD in African Americans could be due to differences in linkage disequilibrium in the genomic regions where the respective genetic markers reside; the SNPs identified by the WTCCC may be correlated with a true causative variant in whites but not in populations of African origin. Another possibility is that there may be variation in additional, as-yet-unknown genetic or environmental factors that might modify the effect of the risk variants, although this would need to be examined in a larger population of sufficient size to allow detection of such gene-environment or gene-gene interactions (37). Alternatively, the SNPs identified by the WTCCC were associated with prevalent CHD case status, so for both white and African-American participants, at least some of the polymorphisms analyzed here might play a role in CHD only after it has become well-established rather than in earlier events in disease pathogenesis.

The odds ratios of 1.2 found for the 2 SNPs showing a significant association with CHD in whites under an additive genetic model were in accordance with the modest effects reported for most common sequence variants influencing complex disease in genome-wide association studies (38). However, homozygous carriers of the susceptibility alleles for the rs1333049 and rs501120 variants comprised 22.5% and 75.2% of white persons in the ARIC cohort, respectively, so the public health impact of these polymorphisms could be substantial, and further investigation in other community-based cohorts is warranted. Neither the WTCCC case-control studies nor this study was designed to test alternative hypotheses of disease causation (such as the contribution of copy number variation or of rare variants with large effects) that might help to further explain the observation that family history of CHD is a strong predictor of disease risk.

ACKNOWLEDGMENTS

Author affiliations: Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas (Jan Bressler, Kelly A. Volcik, Eric Boerwinkle); Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota (Aaron R. Folsom); Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina (David J. Couper); and Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas (Eric Boerwinkle).

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022.

Conflict of interest: none declared.

REFERENCES

1. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. International HapMap Consortium. *Nature*. 2007;449(7164):851–861.
2. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
3. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. *Science*. 2007;316(5829):1331–1336.
4. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316(5829):1341–1345.
5. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881–885.
6. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet.* 2007;39(6):770–775.
7. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316(5829):1336–1341.
8. Liu YJ, Liu XG, Wang L, et al. Genome-wide association scans identified *CTNNBL1* as a novel gene for obesity. *Hum Mol Genet.* 2008;17(12):1803–1813.
9. Scuteri A, Sanna S, Chen WM, et al. Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits [electronic article]. *PLoS Genet.* 2007;3(7):e115.
10. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007;316(5826):889–894.
11. Dina C, Meyre D, Gallina S, et al. Variation in *FTO* contributes to childhood obesity and severe adult obesity. *Nat Genet.* 2007;39(6):724–726.
12. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316(5830):1488–1491.
13. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316(5830):1491–1493.
14. Reiman EM, Webster JA, Myers AJ, et al. *GAB2* alleles modify Alzheimer's risk in *APOE* ε4 carriers. *Neuron*. 2007;54(5):713–720.
15. Grupe A, Abraham R, Li Y, et al. Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum Mol Genet.* 2007;16(8):865–873.
16. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet.* 2006;38(3):320–323.
17. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357(5):443–453.
18. White AD, Folsom AR, Chambless LE, et al. 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(2):223–233.
19. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol.* 1989;129(4):687–702.

20. Siedel J, Hägele EO, Ziegenhorn J, et al. Reagent for the enzymatic determination of serum total cholesterol with improved lipolytic efficiency. *Clin Chem.* 1983;29(6):1075–1080.
21. 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(6):499–502.
22. Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg^{2+} precipitation procedure for quantitation of high-density-lipoprotein cholesterol. *Clin Chem.* 1982;28(6):1379–1388.
23. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
24. Hintze J. *Power Analysis and Sample Size* (*PASS*) [software]. Kaysville, UT: NCSS; 2006.
25. 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(2):189–197.
26. Sandhu MS, Waterworth DM, Debenham SL, et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet*. 2008;371(9611):483–491.
27. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet*. 1997;349(9063):1436–1442.
28. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030 [electronic article]. *PLoS Med.* 2006;3(11):e442.
29. Motulsky A, Brunzell J. Genetics of coronary artery disease. In: King R, Rotter J, Motulsky A, eds. *The Genetic Basis of Common Disease*. New York, NY: Oxford University Press; 2002:105–125.
30. Slack J, Evans KA. The increased risk of death from ischaemic heart disease in first degree relatives of 121 men and 96 women with ischaemic heart disease. *J Med Genet.* 1966;3(4):239–257.
31. Lapidot T, Dar A, Kollet O. How do stem cells find their way home? *Blood*. 2005;106(6):1901–1910.
32. Bleul CC, Fuhlbrigge RC, Casasnovas JM, et al. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J Exp Med*. 1996;184(3):1101–1109.
33. Asahara T, Murohara T, Sullivan A, et al. Isolation of putative progenitor endothelial cells for angiogenesis. *Science*. 1997;275(5302):964–967.
34. Yamaguchi J, Kusano KF, Masuo O, et al. Stromal cell-derived factor-1 effects on ex vivo expanded endothelial progenitor cell recruitment for ischemic neovascularization. *Circulation*. 2003;107(9):1322–1328.
35. Abi-Younes S, Sauty A, Mach F, et al. The stromal cell-derived factor-1 chemokine is a potent platelet agonist highly expressed in atherosclerotic plaques. *Circ Res.* 2000;86(2):131–138.
36. Nielsen MS, Jacobsen C, Olivecrona G, et al. Sortilin/neurotensin receptor-3 binds and mediates degradation of lipoprotein lipase. *J Biol Chem.* 1999;274(13):8832–8836.
37. Burton PR, Hansell AL, Fortier I, et al. Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol*. 2009;38(1):263–273.
38. Altshuler D, Daly M. Guilt beyond a reasonable doubt. *Nat Genet.* 2007;39(7):813–815.
39. Vincent JB, Skaug J, Scherer SW. The human homologue of flamingo, *EGFL2*, encodes a brain-expressed large cadherin-like protein with epidermal growth factor-like domains, and maps to chromosome 1p13.3–p21.1. *DNA Res.* 2000;7(3):233–235.
40. Shima Y, Kengaku M, Hirano T, et al. Regulation of dendritic maintenance and growth by a mammalian 7-pass transmembrane cadherin. *Dev Cell.* 2004;7(2):205–216.
41. Hsieh WJ, Hsieh SC, Chen CC, et al. Human DDA3 is an oncoprotein down-regulated by p53 and DNA damage. *Biochem Biophys Res Commun*. 2008;369(2):567–572.
42. Morris NJ, Ross SA, Lane WS, et al. Sortilin is the major 110-kDa protein in GLUT4 vesicles from adipocytes. *J Biol Chem.* 1998;273(6):3582–3587.
43. Shi J, Kandror KV. Sortilin is essential and sufficient for the formation of Glut4 storage vesicles in 3T3-L1 adipocytes. *Dev Cell.* 2005;9(1):99–108.
44. Bosserhoff AK, Buettner R. Expression, function and clinical relevance of MIA (melanoma inhibitory activity). *Histol Histopathol*. 2002;17(1):289–300.
45. Arndt S, Bosserhoff AK. TANGO is a tumor suppressor of malignant melanoma. *Int J Cancer.* 2006;119(12):2812–2820.
46. Leader B, Leder P. Formin-2, a novel formin homology protein of the cappuccino subfamily, is highly expressed in the developing and adult central nervous system. *Mech Dev.* 2000;93(1-2):221–231.
47. Leader B, Lim H, Carabatsos MJ, et al. Formin-2, polyploidy, hypofertility and positioning of the meiotic spindle in mouse oocytes. *Nat Cell Biol.* 2002;4(12):921–928.
48. Prasannan P, Pike S, Peng K, et al. Human mitochondrial C1-tetrahydrofolate synthase: gene structure, tissue distribution of the mRNA, and immunolocalization in Chinese hamster ovary calls. *J Biol Chem.* 2003;278(44):43178–43187.
49. Walkup AS, Appling DR. Enzymatic characterization of human mitochondrial C1-tetrahydrofolate synthase. *Arch Biochem Biophys*. 2005;442(2):196–205.
50. McCully KS. Homocysteine and vascular disease. *Nat Med.* 1996;2(4):386–389.
51. Sherr CJ. The Pezcoller Lecture: cancer cell cycles revisited. *Cancer Res.* 2000;60(14):3689–3695.
52. Hannon GJ, Beach D. $p15^{INK4B}$ is a potential effector of TGF-β-induced cell cycle arrest [letter]. *Nature*. 1994;371(6494):257–261.
53. Pasmant E, Laurendeau I, Héron D, et al. Characterization of a germ-line deletion, including the entire *INK4/ARF* locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with *ARF*. *Cancer Res.* 2007;67(8):3963–3969.
54. Yokote K, Kobayashi K, Saito Y. The role of Smad3-dependent TGF-β signal in vascular response to injury. *Trends Cardiovasc Med*. 2006;16(7):240–245.
55. Kalinina N, Agrotis A, Antropova Y, et al. Smad expression in human atherosclerotic lesions: evidence for impaired TGF-β/Smad signaling in smooth muscle cells of fibrofatty lesions. *Arterioscler Thromb Vasc Biol.* 2004;24(8):1391–1396.
56. Edlin RS, Tsai S, Yamanouchi D, et al. Characterization of primary and restenotic atherosclerotic plaque from the superficial femoral artery: potential role of Smad3 in regulation of SMC proliferation. *J Vasc Surg*. 2009;49(5):1289–1295.
57. Hug C, Wang J, Ahmad NS, et al. T-cadherin is a receptor for hexameric and high-molecular-weight forms of Acrp30/adiponectin. *Proc Natl Acad Sci U S A*. 2004;101(28):10308–10313.
58. Ling H, Waterworth DM, Stirnadel HA, et al. Genome-wide linkage and association analyses to identify genes influencing

adiponectin levels: the GEMS Study. *Obesity (Silver Spring)*. 2009;17(4):737–744.
59. Org E, Eyheramendy S, Juhanson P, et al. Genome-wide scan identifies *CDH13* as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Hum Mol Genet.* 2009;18(12):2288–2296.
60. Ivanov D, Philippova M, Tkachuk V, et al. Cell adhesion molecule T-cadherin regulates vascular cell adhesion, phenotype and motility. *Exp Cell Res.* 2004;293(2):207–218.
61. Jarrous N, Eder PS, Wesolowski D, et al. Rpp14 and Rpp29, two protein subunits of human ribonuclease P. *RNA*. 1999; 5(2):153–157.
62. van Eenennaam H, Pruijn GJ, van Venrooij WJ. hPop4: a new protein subunit of the human RNase MRP and RNase P ribonucleoprotein complexes. *Nucleic Acids Res.* 1999;27(12): 2465–2472.
63. Nishioka M, Kohno T, Takahashi M, et al. Identification of a 428-kb homozygously deleted region disrupting the *SEZ6L* gene at 22q12.1 in a lung cancer cell line. *Oncogene*. 2000; 19(54):6251–6260.
64. Gorlov IP, Meyer P, Liloglou T, et al. Seizure 6-like (*SEZ6L*) gene and risk for lung cancer. *Cancer Res.* 2007;67(17): 8406–8411.

Appendix Table. Single Nucleotide Polymorphisms Associated With Coronary Heart Disease That Were Identified in a Study by the Wellcome Trust Case Control Consortium[a]

dbSNP[b] ID No.	Chromosome	Nearest Gene(s)	OMIM[c] No.	Function	Reference No. (Current Study)	Genomic Distance (Base Pairs)[d]
rs599839	1p21	*CELSR2*	604265	Knockdown in rat pups resulted in abnormal dendritic arborization	39, 40	29,515
	1p13.3	*PSRC1*		Down-regulated by p53 and DNA damage	41	3′-UTR
	1p21.3–p13.1	*SORT1*	602458	Sorting protein in rat adipocyte vesicles transporting SLC2A4 to plasma membrane in response to insulin; binds LPL in adipocytes	36, 42, 43	118,407
rs17465637	1q41	*MIA3*		Extracellular protein broadly expressed in human tissues; tumor suppressor in melanoma cell lines	44, 45	Intronic
rs17672135	1q43	*FMN2*	606373	Maternal effect gene required for progression through meiosis I in mice	46, 47	Intronic
rs2943634	2q36					
rs383830	5q21.1	*FAM174A*		Hypothetical protein		77,848
rs6922269	6q25.1	*MTHFD1L*	611427	MTHFD1L catalyzes tetrahydrofolate synthesis in mitochondria; involved in de novo synthesis of purines and regeneration of methionine from homocysteine; homocysteine identified as a risk factor for vascular disease	48–50	Intronic
rs1333049	9p21	*CDKN2A/CDKN2B*	600160	*CDKN2A* encodes 2 proteins: p16(INK4), a cyclin-dependent kinase inhibitor, and p14(ARF), involved in p53 regulation; CDKN2B is an effector of TGFB-mediated cell cycle arrest	51, 52	150,455
			600431			116,181
		CDKN2BAS		Noncoding RNA	53	130,703
rs501120	10q11.1	*CXCL12*	600835	Involved in platelet aggregation and expressed in macrophages in human atherosclerotic plaques	35	126,685
rs17228212	15q22.33	*SMAD3*	603109	Transcriptional modulator activated by TGFB; *Smad3* (−/−) mice show enhanced intimal hyperplasia after vascular injury; *SMAD3* expression detected in human atherosclerotic lesions and restenotic plaques	54–56	Intronic
rs8055236	16q24.2–q24.3	*CDH13*	601364	Receptor for adiponectin; associated with adiponectin levels and blood pressure; expressed in endothelial and smooth muscle cells, atherosclerotic lesions, and restensotic plaques; inhibits attachment of human aortic smooth muscle cells and endothelial cells in culture	57–60	Intronic
rs7250581	19q12	*POP4*	606114	Subunit of human ribonuclease P; associated with RMRP, which is involved in pre-rRNA processing	61, 62	32,815
rs688034	22q12.1	*SEZ6L*	607021	Associated with lung cancer	63, 64	Intronic

Abbreviations: *CDH13*, cadherin 13, H-cadherin (heart); *CDKN2A*, cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4); *CDKN2B*, cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4); *CDKN2BAS*, CDKN2B antisense RNA (non-protein-coding); *CELSR2*, cadherin, EGF LAG 7-pass G-type receptor 2 (flamingo homolog, *Drosophila*); *CXCL12*, chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1); *FAM174A*, family with sequence similarity 174, member A; *FMN2*, formin 2; ID, identification; LPL, lipoprotein lipase; *MIA3*, melanoma inhibitory activity family, member 3; *MTHFD1L*, methylenetetrahydrofolate dehydrogenase (NADP+-dependent) 1-like; OMIM, Online Mendelian Inheritance in Man; p53, transformation-related protein 53; *POP4*, processing of precursor 4, ribonuclease P/MRP subunit (*Saccharomyces cerevisiae*); *PSRC1*, proline/serine-rich coiled-coil 1; RMRP, RNA component of mitochondrial RNA processing endoribonuclease; rs, reference SNP; *SEZ6L*, seizure-related 6 homolog (mouse)-like; SLC2A4, solute carrier family 2 (facilitated glucose transporter), member 4; *SMAD3*, SMAD family member 3; SNP, single nucleotide polymorphism; *SORT1*, sortilin 1; TGFB, transforming growth factor β1; UTR, untranslated region.

[a] Wellcome Trust Case Control Consortium, 2007 (2).

[b] The National Center for Biotechnology Information's SNP database (http://www.ncbi.nlm.nih.gov/SNP/).

[c] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Maryland) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Maryland) (http://www.ncbi.nlm.nih.gov/omim/).

[d] Genomic distance from the SNP (calculated using HapMap Data Release 27, February 2009 (1)).