



## Original Contribution

# Prediction of Coronary Heart Disease Risk using a Genetic Risk Score: The Atherosclerosis Risk in Communities Study

Alanna C. Morrison<sup>1</sup>, Lance A. Bare<sup>2</sup>, Lloyd E. Chambless<sup>3</sup>, Stephen G. Ellis<sup>4</sup>, Mary Malloy<sup>5</sup>, John P. Kane<sup>5</sup>, James S. Pankow<sup>6</sup>, James J. Devlin<sup>2</sup>, James T. Willerson<sup>7</sup>, and Eric Boerwinkle<sup>1,7</sup>

<sup>1</sup> Human Genetics Center and Division of Epidemiology, University of Texas Health Science Center at Houston, Houston, TX.

<sup>2</sup> Celera, Alameda, CA.

<sup>3</sup> Collaborative Studies Coordinating Center, University of North Carolina, Chapel Hill, NC.

<sup>4</sup> Department of Cardiovascular Medicine, The Cleveland Clinic Foundation, Cleveland, OH.

<sup>5</sup> Cardiovascular Research Institute, University of California, San Francisco, CA.

<sup>6</sup> Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.

<sup>7</sup> Texas Heart Institute, Houston, TX.

Received for publication May 4, 2006; accepted for publication January 16, 2007.

Recent studies have evaluated whether incorporating nontraditional risk factors improves coronary heart disease (CHD) prediction models. This 1986–2001 US study aggregated the contribution of multiple single nucleotide polymorphisms into a genetic risk score (GRS) and assessed whether the GRS plus traditional risk factors predict CHD better than traditional risk factors alone. The Atherosclerosis Risk in Communities (ARIC) cohort was followed for a median of 13 years for CHD events ( $n = 1,452$ ). Individuals were genotyped for 116 single nucleotide polymorphisms associated with CHD in multiple case-control studies. Single nucleotide polymorphisms nominally predicting incident CHD in the ARIC study were included in the GRS. The GRS was significantly associated with incident CHD in Blacks (hazard rate ratio = 1.20, 95% confidence interval: 1.11, 1.29) and Whites (hazard rate ratio = 1.10, 95% confidence interval: 1.06, 1.14) as well as in each tertile defined by the traditional cardiovascular risk score ( $p \leq 0.02$ ). When receiver operating characteristic curves based on traditional risk factors were recalculated after the GRS was added, the increase in the area under the receiver operating characteristic curve was statistically significant for Blacks and suggestive of improved CHD prediction for Whites. This study demonstrates the concept of aggregating information from multiple single nucleotide polymorphisms into a risk score and indicates that it can improve prediction of incident CHD in the ARIC study.

cardiovascular diseases; genetics; polymorphism, genetic; risk factors

Abbreviations: ACRS, ARIC Cardiovascular Risk Score; ARIC, Atherosclerosis Risk in Communities; AUC, area under the curve; CHD, coronary heart disease; GRS, genetic risk score; SNP, single nucleotide polymorphism.

Coronary heart disease (CHD) is the leading cause of mortality in the United States (1), and numerous studies have focused on identifying risk factors contributing to CHD. The Framingham Study developed a risk score that includes major risk factors, such as age, blood pressure, cigarette smoking, total cholesterol, high density lipoprotein

cholesterol, and diabetes status (2). A similar score was developed that demonstrates the ability of traditional risk factors to predict CHD in the Atherosclerosis Risk in Communities (ARIC) study (3).

Recently, studies have evaluated whether emerging risk factors, such as C-reactive protein, improve prediction of

Correspondence to Dr. Eric Boerwinkle, Human Genetics Center, University of Texas Health Science Center at Houston, 1200 Herman Pressler, Suite 453E, Houston, TX 77030 (e-mail: Eric.Boerwinkle@uth.tmc.edu).

CHD beyond traditional risk factors (3–5). Genetic polymorphisms contributing to CHD incidence are another type of emerging risk factor under investigation in studies generally focused on a priori selected candidate genes (6, 7). Advances in genomic technologies have made it possible to genotype and evaluate many single nucleotide polymorphisms (SNPs) throughout the human genome to identify novel CHD susceptibility genes (8). The reported contribution of any single gene variant to CHD risk has been small, underscoring the polygenic nature of the disease. This paper describes aggregation of the contribution of multiple SNPs into a single genetic risk score (GRS). The GRS comprises SNPs selected from both candidate genes and genes identified through large-scale genomic association studies of CHD. Identification of the definitive set of SNPs for inclusion in a GRS is evolving and was not the primary aim of this study. Rather, the SNPs described here are used to illustrate the concept of a GRS and to investigate whether addition of a GRS improves the prediction of incident CHD beyond that afforded by traditional risk factors. Replication of the predictive ability of the GRS is needed in other population-based samples.

## MATERIALS AND METHODS

### The ARIC study

The ARIC study is a prospective investigation of atherosclerosis and its clinical sequelae involving 15,792 US individuals aged 45–64 years at recruitment (1986–1989). A detailed description of the ARIC study design and methods is published elsewhere (9).

### Incident CHD

Incident CHD was determined by contacting participants annually and by surveying discharge lists from local hospi-

tals and death certificates for potential cardiovascular events (10). Details on quality assurance for ascertainment and classification of CHD events have been published elsewhere (10). CHD cases ( $n = 1,452$ ) were defined as participants with either definite or probable myocardial infarction, a silent myocardial infarction between examinations (indicated by electrocardiogram), a definite CHD death, or a coronary revascularization. Participants were excluded from analyses if they had a positive or unknown history of prevalent stroke, transient ischemic attack/stroke symptoms, or CHD at the initial visit; an ethnic background other than Blacks from Jackson, Mississippi, or Whites; restrictions on use of their DNA; or missing data for any of the traditional cardiovascular risk factors. The remaining 13,907 participants were followed for incident CHD for a median of 13 years between the baseline examination and December 31, 2001.

### Examination and laboratory measures

Cardiovascular risk factors considered in this study were measured at baseline and included age, systolic blood pressure, use of antihypertensive medication, total cholesterol, high density lipoprotein cholesterol, gender, diabetes, and smoking status. These risk factors comprise the ARIC Cardiovascular Risk Score (ACRS) established by Chambless et al. (3) and are shown in table 1 for male and female cases and controls in each racial group. Seated blood pressure was measured three times with a random-zero sphygmomanometer, and the last two measurements were averaged. Interviewers administered a questionnaire to assess use of antihypertensive medication. Plasma total cholesterol was measured by an enzymatic method (11). High density lipoprotein cholesterol was measured after dextran-magnesium precipitation of non-high density lipoproteins (12). Diabetes was defined by a fasting glucose level of  $\geq 126$  mg/dl, a nonfasting glucose level of  $\geq 200$  mg/dl, or a self-reported

**TABLE 1. Characteristics\* of the cardiovascular risk score for cases and controls, by gender and racial group, in the Atherosclerosis Risk in Communities study, United States, 1986–2001**

Characteristic	Cases				Controls			
	White females ( $n = 346$ )	White males ( $n = 793$ )	Black females ( $n = 147$ )	Black males ( $n = 166$ )	White females ( $n = 5,285$ )	White males ( $n = 3,887$ )	Black females ( $n = 2,100$ )	Black males ( $n = 1,183$ )
Age (years)	55.8 (5.4)	55.7 (5.4)	55.0 (5.7)	55.0 (5.8)	53.8 (5.7)	54.3 (5.7)	53.1 (5.7)	53.4 (5.9)
Systolic blood pressure (mmHg)	124.9 (19.2)	124.1 (17.3)	141.2 (27.1)	132.0 (20.2)	116.4 (17.4)	119.2 (15.4)	126.7 (20.1)	129.5 (21.1)
Total cholesterol (mg/dl)	236.4 (45.1)	220.3 (38.8)	232.3 (55.9)	220.3 (47.6)	216.7 (41.4)	208.0 (37.9)	215.3 (44.3)	209.8 (43.6)
High density lipoprotein cholesterol (mg/dl)	49.0 (13.5)	39.8 (10.9)	51.1 (14.3)	46.5 (13.2)	58.3 (17.2)	43.8 (12.6)	58.6 (17.4)	51.5 (17.4)
Diabetic	25.7	19.2	41.5	25.9	6.4	6.9	17.9	14.8
Antihypertensive medication use	36.1	24.1	68.0	48.2	17.3	16.2	41.4	29.3
Smoker	42.2	27.1	39.5	43.4	23.5	23.7	23.2	36.7

\* All values are expressed as mean (standard deviation) or percent.

diagnosis of diabetes or use of diabetic medication. Cigarette smoking status was classified as current or not current.

### Genotype determination

Two sources of SNPs were screened for their ability to predict incident CHD in the ARIC study. Multiple, large-scale genomic association studies of CHD identified 92 SNPs in 87 genes, and a literature review of CHD association studies resulted in the selection and genotyping of 24 SNPs in 21 a priori candidate genes (Web table 1; this information is described in a supplementary table referred to as “Web table 1” in the text and posted on the *Journal's* website (<http://aje.oupjournals.org/>)). Genotyping was carried out by allele-specific real-time polymerase chain reaction following polymerase chain reaction–based amplification of genomic DNA and an oligonucleotide ligation assay similar to a procedure previously described (13). Primers were designed and validated in-house, and the primer and probe sequences for the amplification reactions and any additional information about the genotyping methods are available from the authors upon request. Previous analyses of allele-specific real-time polymerase chain reaction assays showed that the accuracy of our genotyping is greater than 99 percent, as determined by internal comparisons of differentially designed assays for the same marker and comparisons for the same marker across different studies (14).

The 92 SNPs from the genomic association studies demonstrated a consistent association with CHD, either myocardial infarction or severe coronary artery disease, in one of three case-control studies. Subjects were self-described non-Hispanic Whites who were recruited as previously detailed by Shiffman et al. (15), had completed a questionnaire approved by the institutional review board, and gave informed consent to participate in genetic studies. The severe coronary artery disease study comprised 781 cases and 603 controls drawn from 5,800 angiography patients enrolled by the Cleveland Clinic Foundation. The severity of coronary artery disease was scored by summing the percentage of stenosis (determined by coronary angiography, lesions of <50 percent stenosis were scored as 0 percent) in 10 segments: the left main, and the proximal, medial, and distal segments of each of the left anterior descending, left circumflex, and right coronary arteries. Cases had the highest stenosis scores: at least 150 for males and 75 for females. Controls had stenosis scores of zero (unpublished data). In two association studies of myocardial infarction, cases with a history of myocardial infarction were compared with controls with no history of myocardial infarction. These myocardial infarction studies comprised 340 cases and 346 controls enrolled at the University of California, San Francisco (15) and 445 cases and 943 controls enrolled at the Cleveland Clinic Foundation (D. Shiffman, Celera, personal communication, 2006). Most of the approximately 14,000 SNPs investigated in these genomic association studies were missense SNPs, and others were in predicted splice sites or transcription factor binding sites. These approximately 14,000 assays comprise the 11,053 SNPs investigated by Shiffman et al. plus approximately 400 subsequently developed assays. Hence, the 92 SNPs genotyped in the ARIC

study have already undergone an extensive testing and replication process.

### Statistical analyses

Agreement of genotype frequencies with Hardy-Weinberg equilibrium expectations was tested by using a  $\chi^2$  goodness-of-fit test for the controls.

Traditional cardiovascular risk factors were evaluated in a Cox proportional hazards model of time to incident CHD (3). For incident CHD cases, the follow-up time interval was defined as the time between the initial clinical visit and the date of the first CHD event. For noncases, follow-up continued until December 31, 2001; the date of death; or the date of last contact if lost to follow-up. Each model was evaluated separately among Whites and Blacks. An ACRS was calculated for each subject by multiplying each model coefficient by the associated risk-variable value and then summing the products.

For each race, the 92 gene discovery SNPs and the 24 candidate gene SNPs were modeled as the additive effect of the risk-raising allele and were individually evaluated by Cox proportional hazards models adjusted for age and gender. The risk-raising allele is the allele that resulted in a hazard rate ratio greater than 1.0 for Whites. Within each race, SNPs associated with  $p$  values greater than 0.10 were pruned from further consideration for inclusion in the GRS. This pruning criterion was chosen so that SNPs conferring only modest levels of CHD risk could be aggregated into the GRS. SNPs associated with  $p$  values less than or equal to 0.10 entered the race-specific GRS with a coding value of 1 for the risk homozygote, 0 for the heterozygote, and  $-1$  for the nonrisk homozygote. The GRS for each individual was created by summing these values for each SNP in the GRS.

For each race, the GRS was included in a Cox proportional hazards model containing the traditional risk factors comprising the ACRS. A new combined risk score was calculated for each individual (ACRS + GRS) by multiplying each model coefficient by the associated risk variable and then summing the products. The area under the receiver operating characteristic curve equals the probability that a randomly drawn subject who had an incident event within a specified time would have a higher risk score than a subject who did not have an event in that time period. An area under the curve ( $AUC(\hat{t})$ ) for  $t$  equal to 13 years was calculated for both the ACRS and the ACRS + GRS (16). The  $AUC(\hat{t})$  is a consistent predictor of the area under the sensitivity( $t$ ) versus  $1 - \text{specificity}(t)$  curve for prediction of  $t$  years of risk. A comparison of the  $AUC(\hat{t})$  between the ACRS alone and the ACRS + GRS was performed via bootstrapping and evaluating the overlap of the 95 percent confidence interval of the point estimates (i.e., the median difference in  $AUC(\hat{t})$  among bootstrap samples).

An internal validation method using a bootstrap sampling procedure (16) was used to evaluate potential upward bias in the estimated hazard ratio of the GRS in the ARIC study. The seven-step procedure was implemented as follows:

1. Create a bootstrap sample by randomly sampling  $N$  participants with replacement from the original sample,

- where  $N$  is the total number of participants in the original sample.
2. Select a new panel of GRS SNPs from the bootstrap sample by using the same criteria used to select the panel of SNPs from the original sample.
  3. In the bootstrap sample, estimate the regression coefficient for the GRS by using a Cox proportional hazards model adjusted for age and sex.
  4. In the original sample, use the new panel of SNPs to create a GRS for each subject and estimate the regression coefficient (adjusted for age and sex) for the GRS.
  5. Subtract the regression coefficients calculated in step 4 from those in step 3 to obtain an estimate of the bias in the regression coefficient.
  6. After performing 2,000 iterations of steps 2–5, calculate the mean of the estimated bias in the regression coefficient.
  7. Subtract the mean bias in the regression coefficient as calculated in step 6 from the regression coefficient determined in the primary analysis to obtain an estimate for a bias-corrected hazard ratio.

## RESULTS

Of the 92 discovery SNPs and 24 candidate SNPs genotyped in the ARIC cohort, 110 and 112 were in accordance with Hardy-Weinberg expectations in Whites and Blacks, respectively. Of these SNPs, 11 in Whites and 11 in Blacks were selected for inclusion in the GRS based on their nominal association with incident CHD in a Cox proportional hazards model ( $p \leq 0.10$ ). The GRS for each race included 10 SNPs from the 92 gene discovery SNPs and one SNP from the 24 a priori candidate gene SNPs. These SNPs, the genes in which they reside, allele frequencies, and results of the Cox proportional hazards modeling are detailed in table 2. This same information for all 116 SNPs is presented in Web table 1. The two *ADAMTS1* SNPs, rs428785 and rs402007, are separated by 725 base pairs, and the genotypes were 99.7 percent concordant, demonstrating strong linkage disequilibrium. Therefore, only rs402007 was chosen to enter the GRS, resulting in a GRS for Whites composed of 10 SNPs. Only rs20455 in *KIF6* was common between races for the SNPs selected for inclusion in the GRS.

As expected, each of the traditional cardiovascular risk factor components of the ACRS was statistically significant when evaluated in a Cox proportional hazards model. Additionally, the components of the ACRS were significantly different between males and females except for the proportion of Whites using antihypertensive medication and smoking (data not shown). The GRS for Whites, based on the 10 SNPs described in table 2, was calculated for 8,361 individuals and ranged from  $-7$  to  $6$ . The GRS for Blacks, based on 11 SNPs (table 2), was calculated for 2,624 individuals and ranged from  $-9$  to  $5$ . The GRS was normally distributed for each race (figure 1). The range of individual risk scores indicates that no individual of either race had all risk-raising

alleles or all risk-lowering alleles. This finding is consistent with the hypothesis that, as a complex disease, CHD incidence is the result of a balance of the contribution of multiple genes with small effects.

The association of the GRS with CHD in each race was tested in a Cox proportional hazards model containing the traditional risk factors comprising the ACRS. For both Whites and Blacks, the GRS was a significant predictor of incident CHD after accounting for the effects of the traditional risk factors ( $p < 0.01$ ). The hazard ratios were 1.10 per unit (95 percent confidence interval: 1.06, 1.14) for the GRS for Whites and 1.20 per unit (95 percent confidence interval: 1.11, 1.29) for the GRS for Blacks. Thus, for a White homozygous for six at-risk alleles or a Black homozygous for five at-risk alleles, the hazard rate ratio would be 1.77 or 2.49, respectively. The GRS was also a significant predictor of incident CHD after accounting for the effects of the traditional risk factors ( $p < 0.02$ ) in gender-specific analyses within each race (data not shown). Use of an internal validation method based on a bootstrap sampling procedure indicated that in the 2,000 bootstrap samples, the SNPs in the original GRS panels for Whites and Blacks were the most frequently chosen for these new panels. The bootstrap estimate of a bias-corrected hazard ratio was 1.05 for the GRS for Whites and 1.09 for Blacks.

Next, the area under the receiver operating characteristic curves for the ACRS + GRS model, which included both genetic and traditional risk factors, was compared with the AUC for the ACRS model, which included only the traditional risk factors. For each race, including the GRS in the model used to calculate risk scores increased the AUC over that observed when the receiver operating characteristic curve was based on only traditional risk factors (figure 2). For Whites, the AUC was 0.766 for the ACRS + GRS model and 0.764 for the ACRS model. The increase in the AUC for Whites was 0.002 (95 percent confidence interval: 0.000, 0.006). For Blacks, the AUC was 0.769 for the ACRS + GRS model and 0.758 for the ACRS model. The increase in the AUC for Blacks was 0.011 (95 percent confidence interval: 0.002, 0.024). Thus, for Whites, the 95 percent confidence interval of the increase barely included zero (i.e., the null hypothesis), indicating that the increase in the AUC caused by including the GRS is marginally statistically significant at the 0.05 level. For Blacks, the 95 percent confidence interval of the increase did not include zero, suggesting that the increase in AUC caused by including the GRS is statistically significant.

To further investigate the ability of the GRS to predict incident CHD in the context of different levels of traditional risk for CHD, the association of the GRS with risk of CHD was evaluated within each tertile (e.g., 33.33 percentile) of the ACRS by a Cox proportional hazards model (table 3). Within each tertile of the ACRS, the GRS was a statistically significant predictor of incident CHD in both races ( $p \leq 0.02$ ). Additionally, the interaction between the GRS and the ACRS was not statistically significant for Whites ( $p = 0.62$ ) or Blacks ( $p = 0.12$ ), further supporting the idea that the GRS is predictive of CHD risk across the spectrum of traditional risk factors.

**TABLE 2. Genetic risk score SNPs,\* their genes, and their association with incident coronary heart disease, the Atherosclerosis Risk in Communities study, United States, 1986–2001**

Gene symbol	Gene name	SNP ID*	Chromosome location	Risk-raising allele (frequency)	Risk-lowering allele (frequency)	HRR*	95% CI*	<i>p</i> value
<i>Whites</i>								
<i>VAMP8</i>	Vesicle-associated membrane protein 8	rs1010	2p11.2	C (0.42)	T (0.58)	1.10	1.00, 1.19	0.04
<i>PALLD</i>	Palladin, cytoskeletal-associated protein	rs7439293	4q32.3	A (0.62)	G (0.38)	1.11	1.02, 1.21	0.02
<i>KIF6</i>	Kinesin family member 6	rs20455	6p21.2	G (0.36)	A (0.64)	1.09	1.00, 1.19	0.05
<i>MKI67</i>	Antigen identified by monoclonal antibody Ki-67	rs11016076	10q26.2	G (0.17)	C (0.83)	1.10	0.98, 1.24	0.09
<i>MYH15</i>	Myosin, heavy polypeptide 15	rs3900940	3q13.13	C (0.30)	T (0.70)	1.18	1.08, 1.29	<0.01
<i>LOC646377</i>	Similar to pyrin and the Hin domain family, member 1 alpha 1 isoform	rs2213948	1q23.1	A (0.16)	G (0.84)	1.12	1.00, 1.25	0.05
<i>HPS1</i>	Hermansky-Pudlak syndrome 1	rs2296436	10q24.2	T (0.92)	C (0.08)	1.18	1.00, 1.39	0.05
<i>SNX19</i>	Sorting nexin 19	rs2298566	11q24.3	C (0.75)	A (0.25)	1.12	1.01, 1.24	0.03
<i>ADAMTS1</i>	ADAM* metalloproteinase with thrombospondin type 1 motif, 1	rs428785	21q21.3	G (0.78)	C (0.22)	1.13	1.02, 1.26	0.02
<i>ADAMTS1</i>	ADAM metalloproteinase with thrombospondin type 1 motif, 1	rs402007	21q21.3	G (0.78)	C (0.22)	1.15	1.03, 1.28	0.01
<i>ADRB3</i> †	β3 adrenergic receptor	rs4994	8p12	C (0.08)	T (0.92)	1.26	1.12, 1.43	<0.01
<i>Blacks</i>								
<i>DMXL2</i>	Dmx-like 2	rs12102203	15q21.2	G (0.45)	A (0.55)	1.18	1.01, 1.39	0.04
<i>ZNF132</i>	Zinc finger protein 132	rs1122955	19q13.43	C (0.90)	T (0.10)	1.45	1.04, 2.03	0.03
<i>KIF6</i>	Kinesin family member 6	rs20455	6p21.2	G (0.79)	A (0.21)	1.23	0.99, 1.52	0.06
<i>F2</i>	Coagulation factor II (thrombin)	rs1799963	11p11.2	A (0.01)	G (0.99)	3.64	1.50, 8.82	<0.01
<i>OR2A25</i>	Olfactory receptor, family 2, subfamily A, member 25	rs2961135	7q35	G (0.32)	C (0.68)	1.22	1.03, 1.44	0.02
<i>KRT5</i>	Keratin 5	rs89962	12q13.13	T (0.07)	G (0.93)	1.29	0.97, 1.72	0.08
<i>CTNNA3</i>	Catenin (cadherin-associated protein), alpha 3	rs10822891	10q21.3	G (0.31)	A (0.69)	1.22	1.02, 1.45	0.03
<i>HAP1</i>	Huntingtin-associated protein 1	rs4796603	17q21.2	A (0.47)	T (0.53)	1.21	1.02, 1.43	0.03
<i>GIPR</i>	Gastric inhibitory polypeptide receptor	rs1800437	19q13.32	G (0.89)	C (0.11)	1.38	1.02, 1.87	0.04
<i>FSTL4</i>	Follistatin-like 4	rs3749817	5q31.1	A (0.08)	G (0.92)	1.28	0.97, 1.70	0.08
<i>THBS2</i> †	Thrombospondin 2	rs8089	6q27	G (0.21)	T (0.79)	1.26	1.04, 1.51	0.02

\* SNPs, single nucleotide polymorphisms; ID, identification; HRR, hazard rate ratio; CI, confidence interval; ADAM, a disintegrin and metalloproteinase.

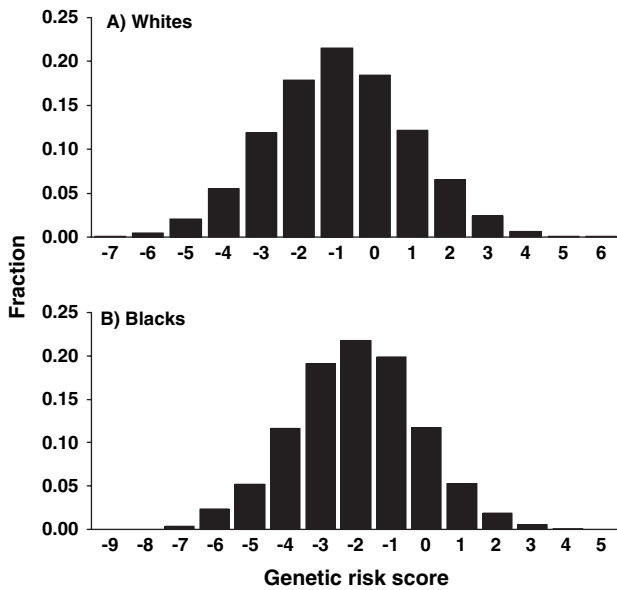
† SNPs from the list of a priori candidate genes.

## DISCUSSION

This study introduces the concept of a GRS that aggregates information from multiple genetic variants into a single score and indicates that a GRS can improve the ability to predict incident CHD beyond that afforded by traditional nongenetic risk factors. For both Blacks and Whites, the GRS was statistically significantly associated with CHD incidence in the ARIC cohort after adjustment for traditional risk factors as well as within all tertiles of the ACRS. Statistically significant risk prediction was further confirmed for Blacks, where inclusion of the GRS significantly in-

creased the area under the receiver operating characteristic curves beyond that provided by the conventional risk factors. For Whites, the increase in the AUC was marginally statistically significant. Improvement in the prediction of incident CHD by inclusion of the GRS in the model is on the magnitude observed for traditional risk factors, such as total cholesterol levels. Removing total cholesterol from the ACRS for Whites compared with the ACRS model resulted in a difference in the AUCs equal to 0.014 (95 percent confidence interval: 0.009, 0.020).

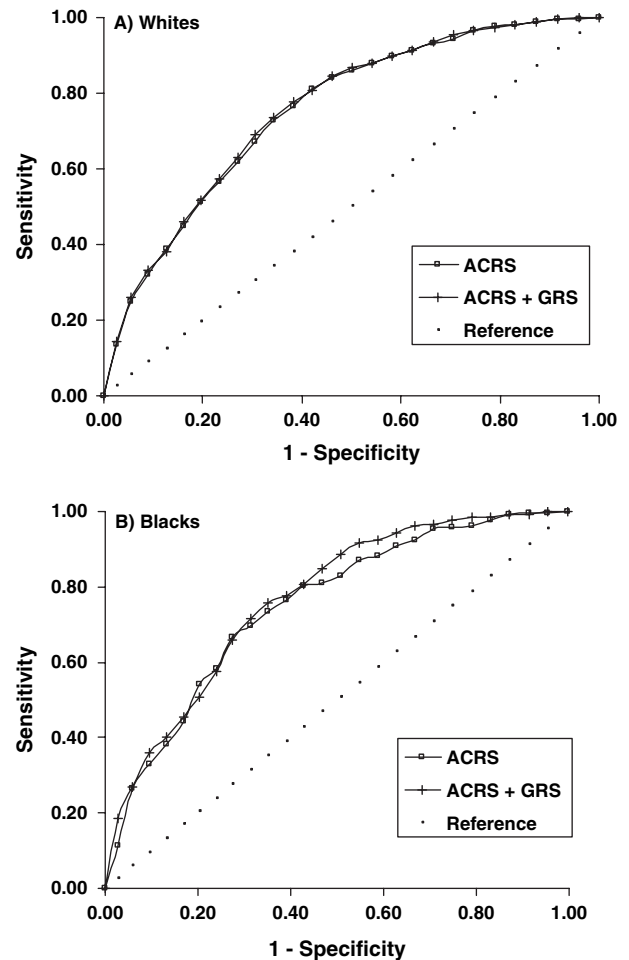
Several assumptions were made in calculating and investigating the GRS. First, gene discovery for CHD is an active



**FIGURE 1.** Distribution of the genetic risk score for predicting coronary heart disease in Whites (A) and Blacks (B), the Atherosclerosis Risk in Communities study, United States, 1986–2001.

area of research, and the set of SNPs selected for inclusion in the GRS is not definitive. Following up genetic linkage analyses (17, 18) and large-scale genomic association studies (8, 15) will undoubtedly identify new gene variants that should be incorporated into a GRS. Second, we generated and tested this GRS in the same sample set. Testing in independent sample sets is warranted to further investigate the prediction of CHD by this GRS. Third, the GRS risk score was calculated by using a linear weighting of  $-1$ ,  $0$ ,  $1$  for genotypes containing zero, one, or two risk alleles, respectively. This method did not attempt to incorporate specific genetic models. The additive model has been shown to perform well, even when the underlying model is unknown or wrongly specified (19). However, more detailed dominance patterns should be considered when a definitive set of SNPs is identified for the GRS. Additionally, this method did not attempt to weigh the SNPs by using the coefficients from the survival analysis or to include gene-gene or gene-environment interactions in calculating the GRS. The relation between the GRS and environmental measures, such as the traditional risk factors comprising the ACRS, has been addressed in other contexts (table 3). Overall, the intention was to demonstrate the concept of aggregating information from multiple SNPs into a risk score and to create a score that was simple to calculate and generalizable. Finally, different populations and population strata may be most appropriately evaluated with their own group-specific GRS. In this analysis, self-reported race was used as a stratifying variable and race-specific scores were calculated because of the significant differences in allele frequencies between races for the SNPs analyzed here.

Several other approaches to creating a score statistic for predicting disease have been reported. Horne et al. (20) introduced a regression method for calculating a risk score



**FIGURE 2.** Receiver operating characteristic curves using the Atherosclerosis Risk in Communities study cardiovascular risk score (ACRS) alone and incorporating the genetic risk score (GRS) for Whites (A) and Blacks (B), United States, 1986–2001.

that incorporated three genetic polymorphisms and other risk factors, and they found that the frequency of CHD was different at different regression score levels. In a simulation study, Ortlepp et al. (21) concluded that multiple SNPs are better than single SNPs and that as many as 200 SNPs may be necessary for “reasonable” genetic discrimination. Aston et al. (22) suggest that a score based on 90 SNPs in 78 genes can predict risk of breast cancer, but the identity of the SNPs and the algorithm for calculating the score remain proprietary. To our knowledge, there have been no previous applications of a GRS for predicting incident disease in a longitudinal context.

It is noteworthy that of the 21 a priori candidate genes considered, only the  $\beta 3$  adrenergic receptor (*ADRB3*) in Whites and thrombospondin-2 (*THBS2*) in Blacks predicted CHD in the ARIC study. The failure of the other 19 a priori candidate genes, despite their previously reported associations with CHD, is likely due to the modest level of risk associated with any single polymorphism. Calculations

**TABLE 3. Association of incident coronary heart disease with the genetic risk score within tertiles of the ACRS,\* Atherosclerosis Risk in Communities study, United States, 1986–2001**

ACRS tertile	Subjects (no. of coronary heart disease cases)	HRR†	95% CI*	<i>p</i> value
Whites				
1st	2,787 (69)	1.16	1.02, 1.31	0.02
2nd	2,771 (234)	1.11	1.03, 1.19	<0.01
3rd	2,803 (585)	1.09	1.04, 1.14	<0.01
Blacks				
1st	847 (17)	1.50	1.14, 1.97	<0.01
2nd	887 (51)	1.33	1.15, 1.54	<0.01
3rd		1.12	1.03, 1.23	0.01

\* ACRS, Atherosclerosis Risk in Communities study cardiovascular risk score; CI, confidence interval.

† Hazard rate ratios (HRRs) are for a one-unit increase in the genetic risk score.

reveal that the smallest detectable SNP-associated risk in this study at 80 percent power was a hazard rate ratio equal to 1.02 for Whites and 1.03 for Blacks. SNPs that truly confer only modest risk will not replicate in every study. This issue inspired us to investigate the concept of aggregating the polygenic contribution of multiple SNPs into a single GRS.

Of the variation identified from the large-scale genomic analyses, only *F2*, which encodes prothrombin, was from a probable candidate gene. The SNP rs1799963 encodes a G-to-A nucleotide substitution in the 3'-untranslated region of the gene. Table 2 shows that this SNP had the single largest hazard ratio, and it has previously demonstrated an association with elevated levels of plasma prothrombin and with increased risk of venous thrombosis (23). The second largest hazard ratio in table 2 was for an SNP in *ZNF132*. *ZNF132* is a transcription regulatory protein thought to be involved in cell proliferation, and rs1122955 encodes a G-to-D substitution at amino acid 203 in the polypeptide (24). The SNP in *KIF6* merits additional comment because it was the only SNP included in the GRS for both Whites and Blacks. *KIF6* is homologous to members of the kinesin gene family, playing a role in intracellular transport (25). In follow-up studies, *KIF6* gene variation is associated with measures of body size and insulin resistance (Morrison and Boerwinkle, unpublished observation). That the other genes are not significantly associated with CHD in both groups is of interest. Previous epidemiologic research has identified both similarities and differences in the influence of established cardiovascular risk factors between Blacks and Whites. In particular, in the ARIC study, low density lipoprotein cholesterol is similarly predictive of incident CHD in both groups, whereas hypertension is more predictive for Blacks than for Whites (26). In the current study, differences in genetic risk between Blacks and Whites may be attributable to differences in allele frequencies, linkage disequilibrium patterns, and sample sizes leading to differences in statistical power between the two groups.

That genetic information can be used for disease prediction is evident from the fact that both the American Heart Association and the Centers for Disease Control and Prevention recognize a positive family history as a risk factor for CHD and provide elaborate tools to help individuals assess their own family's history of disease (<http://www.s2mw.com/aha/fht/index.aspx>, [http://www.cdc.gov/genomics/public/famhist\\_alt.htm](http://www.cdc.gov/genomics/public/famhist_alt.htm)). Family history is a complex construct, entailing both genetic and common environmental components. Because an individual receives only half of each of his or her parent's genes, explicit evaluation of an individual's genetic constitution should provide a better assessment of genetic risk than family history. A GRS, like a positive family history, captures the summed effects of multiple genes. The results presented here argue for continued research in exploring the utility of a GRS for identifying those at most risk of CHD and for guiding treatment.

## ACKNOWLEDGMENTS

The Atherosclerosis Risk in Communities (ARIC) 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.

The authors thank the ARIC study staff for their important contributions.

Drs. Devlin and Bare are employees of Celera. Drs. Malloy and Kane receive support from a University of California Discovery Grant from the Industry-University Collaborative Research Program that is jointly funded by the University of California and the State of California, with matching funds from Celera.

## REFERENCES

1. American Heart Association. 2001 Heart and stroke statistical update. Dallas, TX: American Heart Association, 2001.
2. Wilson P, D'Agostino R, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; 97:1837–47.
3. Chambless L, Folsom A, Sharrett A, et al. Coronary heart disease prediction in the Atherosclerosis Risk in Communities (ARIC) Study. *J Clin Epidemiol* 2003;56:880–90.
4. Folsom AR, Chambless LE, Ballantyne CM, et al. An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the Atherosclerosis Risk in Communities study. *Arch Intern Med* 2006;166:1368–73.
5. Koenig W, Lowel H, Baumert J, et al. C-reactive protein modulates risk prediction based on the Framingham Score. Implications for future risk assessment: results from a large cohort study in southern Germany. *Circulation* 2004;109: 1349–53.
6. Stephens J, Humphries S. The molecular genetics of cardiovascular disease: clinical implications. *J Intern Med* 2003;253: 120–7.

7. Morrison A, Bray M, Folsom A, et al. *ADD1 460W* allele associated with cardiovascular disease in hypertensive individuals. *Hypertension* 2002;39:1053–7.
8. McCarthy J, Parker A, Salem R, et al. Large scale association analysis for identification of genes underlying premature coronary heart disease: cumulative perspective from analysis of 111 candidate genes. *J Med Genet* 2004;41:321–3.
9. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC Investigators. *Am J Epidemiol* 1989;129:687–702.
10. White A, Folsom A, Chambless L, 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;49:223–33.
11. Siedel J, Hagele EO, Ziegenhorn J, et al. Reagent for the enzymatic determination of serum total cholesterol with improved lipolytic efficiency. *Clin Chem* 1983;29:1075–85.
12. Warnick GR, Benderson JM, Albers JJ. Dextran sulfate-Mg<sup>2+</sup> precipitation procedure for quantification of high-density-lipoprotein cholesterol. *Clin Chem* 1982;28:1379–88.
13. Iannone M, Taylor J, Chen J, et al. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* 2000;39:131–40.
14. Li Y, Nowotny P, Holmans P, et al. Association of late-onset Alzheimer's disease with genetic variation in multiple members of the *GAPD* gene family. *Proc Natl Acad Sci U S A* 2004;101:15688–93.
15. Shiffman D, Ellis S, Rowland C, et al. Identification of four gene variants associated with myocardial infarction. *Am J Hum Genet* 2005;77:596–605.
16. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med* 2006;25:3474–86.
17. Barkley R, Chakravarti A, Cooper R, et al. Positional identification of hypertension susceptibility genes on chromosome 2. *Hypertension* 2004;43:477–82.
18. Horikawa Y, Oda N, Cox N, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000;26:163–75.
19. Horvath S, Xu X, Laird N. The family based association test method: strategies for studying general genotype-phenotype associations. *Eur J Hum Genet* 2001;9:301–6.
20. Horne B, Anderson J, Carlquist J, et al. Generating genetic risk scores from intermediate phenotypes for use in association studies of clinically significant endpoints. *Ann Hum Genet* 2005;69:176–86.
21. Ortlepp J, Lauscher J, Janssens U, et al. Analysis of several hundred genetic polymorphisms may improve assessment of the individual genetic burden for coronary artery disease. *Eur J Intern Med* 2002;13:485–92.
22. Aston C, Ralph D, Lalo D, et al. Oligogenic combinations associated with breast cancer risk in women under 53 years of age. *Hum Genet* 2005;116:208–21.
23. Poort S, Rosendaal F, Reitsma P, et al. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood* 1996;88:3698–703.
24. Tommerup N, Vissing H. Isolation and fine mapping of 16 novel human zinc-finger encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* 1995;27:259–64.
25. Miki H, Okada Y, Hirokawa N. Analysis of the kinesin superfamily: insights into structure and function. *Trends Cell Biol* 2005;15:467–76.
26. Jones D, Chambless L, Folsom R, et al. Risk factors for coronary heart disease in African Americans: The Atherosclerosis Risk in Communities Study, 1987–1997. *Arch Intern Med* 2002;162:2565–71.