

Five common gene variants identify elevated genetic risk for coronary heart disease

Lance A. Bare, PhD¹, Alanna C. Morrison, PhD², Charles M. Rowland, MS¹, Dov Shiffman, PhD¹, May M. Luke, MBA, PhD¹, Olga A. Iakoubova, MD, PhD¹, John P. Kane, MD, PhD³, Mary J. Malloy, MD³, Stephen G. Ellis, MD⁴, James S. Pankow, PhD⁵, James T. Willerson, MD⁶, James J. Devlin, PhD¹, and Eric Boerwinkle, PhD²

Purpose: Because multiple genetic variants influence risk for coronary heart disease, we combined multiple variants that had been associated with coronary heart disease in several studies into a genetic risk score and asked whether a high genetic risk score would be significantly associated with coronary heart disease after accounting for traditional risk factors. **Methods:** We considered five variants that were associated with coronary heart disease in two studies and confirmed in the Atherosclerosis Risk in Communities study: rs20455 (*KIF6*), rs3900940 (*MYH15*), rs7439293 (*PALLD*), rs2298566 (*SNX19*), and rs1010 (*VAMP8*). We calculated a genetic risk score for each Atherosclerosis Risk in Communities study participant and estimated the hazard ratio for incident coronary heart disease of a high genetic risk score (compared with not-high) in Cox models that adjusted for traditional risk factors during a median of 13 years of follow-up. **Results:** For white participants with a high genetic risk score (4% of the 9129 whites), compared with those without a high genetic risk score, the hazard ratio for incident coronary heart disease was 1.57 (95% confidence interval 1.21–2.04; $P = 0.001$). Internal validation using bootstrap samples estimated that a hazard ratio of 1.43 could be expected in external populations. **Conclusions:** After adjusting for traditional risk factors, those with a high genetic risk score had a 57% increased risk of incident coronary heart disease in the Atherosclerosis Risk in Communities study. **Genet Med 2007;9(10):682–689.**

Key Words: Atherosclerosis Risk in Communities study, coronary heart disease, genetic risk score, variant, single nucleotide polymorphism, complex disease

Coronary heart disease (CHD), the clinical manifestation of the development and progression of coronary atherosclerosis, is the most common cause of death in industrialized countries and is expected to be the most common cause of death worldwide by 2020.¹ CHD morbidity and mortality can be substantially reduced by preventive interventions, including lifestyle modifications, such as smoking cessation, and medical treatments, such as therapies that lower plasma cholesterol levels.²

From the ¹Celera, Alameda, California; ²Human Genetics Center, University of Texas at Houston Health Science Center, Houston, Texas; ³Cardiovascular Research Institute, University of California, San Francisco, California; ⁴Department of Cardiovascular Medicine, The Cleveland Clinic Foundation, Cleveland, Ohio; ⁵Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota; and ⁶Texas Heart Institute, Institute of Molecular Medicine, and University of Texas Health Science Center at Houston, Houston, Texas.

Lance A. Bare, PhD, Celera, 1401 Harbor Bay Parkway, Alameda, CA 94502. E-mail: Lance.Bare@celera.com.

Disclosure: L. A. Bare, C. M. Rowland, D. Shiffman, M. M. Luke, O. A. Iakoubova, and J. J. Devlin have employment and ownership interests in Celera.

Supplementary text and tables are available via the ArticlePlus feature at www.geneticsinmedicine.org. Please go to the October issue and click on the ArticlePlus link posted with the article in the Table of Contents to view this material.

Submitted for publication: June 8, 2007.

Accepted for publication: July 31, 2007.

DOI: 10.1097/GIM.0b013e318156fb62

For example, the National Cholesterol Education Program (NCEP) makes lifestyle modification and treatment recommendations based on low-density lipoprotein cholesterol (LDL-C) levels and a CHD risk assessment algorithm that considers the traditional risk factors age, sex, smoking, hypertension, family history of CHD, and plasma levels of high-density lipoprotein cholesterol (HDL-C).³ However, many patients who have CHD events are not considered to be at high risk for CHD by current risk assessment protocols.^{4,5}

Risk assessment algorithms could be improved by including genetic risk factors because CHD has a genetic component that is independent of traditional risk factors.^{6–8} However, genetic risk factors have not been incorporated into current risk assessment algorithms for two reasons. First, given the complex etiology and polygenic nature of CHD, multiple genetic variants probably contribute to the risk of CHD, and the risk associated with each individual genetic variant is likely to be modest. Consequently, knowledge of the genotype of a single genetic variant may not be sufficient to improve risk prediction in the population or to affect decisions regarding preventive measures. Second, genetic variants reported to be associated with CHD tend not to be reproducibly associated with CHD when they are studied in other populations.^{9,10} However, both of these obstacles to the application of genetics to CHD risk as-

assessment may be overcome by results from continuing genetic studies of CHD.¹¹

As more variants associated with CHD are identified, the modest risk associated with individual genetic variants can be aggregated into a genetic risk score (GRS), with the expectation that the risk associated with a high GRS would be sufficient to influence clinical decisions. A GRS for type 2 diabetes, which aggregated risk from multiple genetic variants, has been reported,¹² and Morrison et al. recently described the concept of using a GRS for CHD.¹³ The GRS for CHD reported by Morrison et al. was based on 10 single nucleotide polymorphisms (SNPs) that were found to be nominally associated ($P < 0.1$) with CHD in the Atherosclerosis Risk in Communities study (ARIC). We have now extended this initial concept study of a GRS for CHD in two ways. First, given that associations between genetic variants and CHD frequently cannot be confirmed in additional populations, we have tightened the inclusion criteria for the panel of SNPs used to calculate the GRS. The initially reported CHD GRS panel included SNPs if they were associated with CHD in ARIC and in at least one other association study, whereas the present study investigates a GRS based only on those SNPs that have the same risk allele associated with CHD ($P < 0.1$) in at least two association studies in addition to ARIC. Second, the initially reported GRS was treated as a continuous variable when evaluated for its association with CHD. However, in making clinical decisions, the high-risk group for a risk factor is frequently identified using a single cut point in a measurement. Therefore, we have used a cut point to identify those ARIC participants with a high GRS. We then asked whether a high GRS is associated with CHD in ARIC and whether the magnitude of the risk of a high GRS was similar to the magnitude of the risk of traditional risk factors.

MATERIALS AND METHODS

Study populations

A detailed description of the ARIC study design has been previously published.¹⁴ Briefly, ARIC is a prospective cohort study of 15,792 African American and white adults from four communities: Forsyth County, NC; Jackson, MS; suburban Minneapolis, MN; and Washington County, MD. Participants were 45–64 years old at enrollment and were followed prospectively for the occurrence of cardiovascular events. Baseline examinations, performed between 1987 and 1989, included a medical history interview, physical examination, and blood draw. DNA was prepared from blood drawn at baseline; genotyping was successful for more than 99.5% of the DNA samples. For the five SNPs in the GRS panel, the minimum genotyping success rate was 96.8%. The genetic analysis included the following traditional risk factors: age, systolic blood pressure, use of antihypertensive medication, total cholesterol, HDL-C, gender, diabetes, and smoking status. Traditional risk factors considered in this study were measured at the baseline visit. Medical events were identified by annual questionnaire, by hospital and death certificate surveillance, and by follow-up examinations conducted every 3 years. The following ARIC

Table 1
Baseline characteristics of participants in this genetic study of ARIC

	Total	White	African American
Number of participants	12,556 ^a	9,389	3,167
Age, yr	53.9 ± 5.7	54.1 ± 5.7	53.3 ± 5.8
Male	5,413 (43.1)	4,231 (45.1)	1,182 (37.3)
Current smoking	3,219 (25.6)	2,281 (24.3)	938 (29.6)
Diabetes	1,346 (10.7)	761 (8.1)	585 (18.5)
Hypertension	4,083 (32.5)	2,375 (25.3)	1,708 (53.9)
LDL cholesterol, mg/dL	136.8 ± 39.0	136.6 ± 37.7	137.2 ± 42.6
HDL cholesterol, mg/dL	52.2 ± 17.0	51.1 ± 16.8	55.3 ± 17.2
Total cholesterol, mg/dL	214.2 ± 41.6	214.2 ± 40.4	214.5 ± 44.8
Family history of CHD	2,306 (18.8)	1,969 (21.2)	337 (11.2)

Data are presented as number of participants (%) or mean ± SD. Conversion factor for converting mM to mg/dL for LDL, HDL, and total cholesterol is 0.0259.

^aNumber of ARIC participants remaining after applying the exclusions listed in the methods and after excluding participants missing genotype data for the five GRS SNPs.

participants were excluded from this incident CHD analysis: those who, at baseline, had a positive or unknown history of CHD, stroke, transient ischemic attack, or stroke symptoms; those who had an ethnic background other than white or African American; and those who had restrictions on the use of their DNA. The 13,907 participants remaining after these exclusions were followed for incident CHD for a median of 13 years after the baseline examination (Table 1). The ascertainment and classification of CHD events has been described previously¹⁵ and was adjudicated by the ARIC Morbidity and Mortality Classifications Committee. The CHD endpoint was a composite of definite or probable myocardial infarction (MI) ($n = 695$), silent MI between examinations (ascertained by electrocardiogram) ($n = 110$), definite fatal CHD death ($n = 136$), or coronary revascularization ($n = 511$). Only the first CHD event after enrollment ($n = 1452$) was counted as an incident event. Event-free participants were followed until the earliest of December 31, 2001, the date of last contact, or death. Appropriate institutional review boards approved the ARIC study, and all participants provided written informed consent.

The following two case-control studies of MI were used to confirm the risk allele for three of the five SNPs used to calculate the GRS. The first of these case-control studies comprised 475 cases and 619 controls that were recruited by the Cleveland Clinic Foundation Heart Center, Cleveland, OH (CCF): the CCF study. The second case-control study comprised 793 cases and 1000 controls that were recruited by the Genomic Resource in Arteriosclerosis at the University of California, San Francisco (UCSF): the UCSF study. All cases in these two studies had a history of MI and the controls did not. A description of the baseline characteristics and the recruitment criteria for the CCF and UCSF studies are described in Table 5 (available online only) and the “Case-Control Studies” section of the supporting information (available on-

line only). All subjects in these two studies were self-described, non-Hispanic whites who had completed an Institutional Review Board-approved questionnaire and given informed consent to participate in genetic studies. Genotypes in these case-control studies were determined using either a multiplex SNP assay or allele-specific real-time PCR as previously described.¹³

Statistical analyses

Statistical analyses were performed with SAS version 8.2 (SAS Institute Inc., Cary, NC) or STATA SE version 8.2 (Stata Corp., College Station, TX). In ARIC, deviation from Hardy-Weinberg equilibrium in noncases was tested using a χ^2 goodness-of-fit test separately in African Americans and whites as previously described.¹³ Individual SNPs were analyzed in race-specific Cox proportional hazards models of CHD that adjusted for age (continuous) and sex. These Cox models assumed an additive (on a log scale) genetic model: genotypes were coded 0, 1, or 2 for nonrisk homozygotes, heterozygotes, or risk homozygotes, respectively. All *P* values reported are two-sided; however, we did not consider any SNP to be associated with incident CHD in ARIC if the prespecified risk allele (based on association studies other than ARIC) was not also the risk allele in ARIC. The white ARIC participants provided 80% power to detect SNPs with hazard ratios ≥ 1.2 , 1.13, and 1.12, given an α of 0.05 and allele frequencies of 0.1, 0.3, and 0.5. False-positive report probabilities were calculated as described by Wacholder et al.¹⁶ using the following equation:

$$\text{FPRP} = \frac{\alpha(1 - \pi)}{\alpha(1 - \pi) + (1 - \beta)\pi}$$

where FPRP is the false-positive report probability, α is the nominal *P* value for the association, $1 - \beta$ is the statistical power to detect in ARIC an association with CHD for the SNP, and π is the prior probability that the SNP is associated with risk (that is, the probability of association without considering the ARIC data). Because prior probabilities used in calculating the false-positive report probabilities are subjective, we used a range of prior probabilities, as suggested by Wacholder et al.¹⁶ The range of prior probabilities for the SNP in *VAMP8* was based on previously published false discovery rates. The SNP in *VAMP8* was associated with early-onset MI in three case-control studies, with a false discovery rate of <0.1 in the third study.¹⁷ Because the false discovery rate was <0.1 , we used prior probabilities that ranged from 0.9 to 0.09. The *KIF6* SNP (rs20455) was associated with CHD in the placebo arms of two CHD prevention trials and the association remained significant after a Bonferroni correction for multiple testing.¹⁸ Because the significance threshold was 0.05, we used prior probabilities that ranged from 0.95 to 0.095. The other three of the five SNPs (in *PALLD*, *MYH15*, and *SNX19*) had been found to be associated with MI in the CCF and UCSF case-control studies reported here. For these three SNPs we used prior probabilities that ranged from 0.33 to 0.0033—10-fold above and below a point estimate (0.033) of the prior probability. The rationale for this point estimate and calculation details for the

false-positive report probabilities are described in the “False-Positive Report Probabilities” section of the supporting information (available online only). As a comparison, a point estimate of prior-probability for a randomly selected SNP might reasonably be 0.00001, assuming that there is one SNP in about a kilobase of genomic DNA (for a total of 3 million SNPs) and that about 30 SNPs are truly involved in CHD.

To calculate a GRS, the participant’s genotype for each SNP was assigned a value of 1 for a risk allele homozygote, 0 for a heterozygote, and -1 for a nonrisk allele homozygote. Then the GRS was calculated by summing the values for each of the five SNPs in the GRS panel.¹³

Confidence limits for rates of incident CHD were calculated using the method of Ulm.¹⁹ Because subjects were followed prospectively over time for occurrence of the CHD endpoint, we used a Cox proportional hazards model^{20,21} to estimate the relative hazard of experiencing a CHD event in subjects having a high GRS as compared with subjects having a low GRS. The Cox model also accounts for participants that are lost to follow-up. The Cox models adjusted for age, LDL-C level, and HDL-C level (all continuous variables) and also adjusted for sex, hypertension (systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or use of prescription medications for high blood pressure), diabetes (fasting glucose level ≥ 126 mg/dL (6.993 mM), nonfasting glucose level ≥ 200 mg/dL (11.1 mM), or self-reported history of either treatment for diabetes or physician diagnosis of diabetes, smoking status (current versus noncurrent), and family history of CHD (the baseline family history question asked whether the participant’s parents had had a heart attack—before age 65 for the mother or before age 55 for the father). To test for potential interactions, each risk factor was also tested in a fully adjusted Cox model that included a term for the GRS (high versus not-high) and a product term for the risk factor and the GRS. We used a bootstrap internal validation procedure to estimate to what extent the Cox coefficient estimated in ARIC is higher than the coefficient that would be expected in an external population^{22,23}; the algorithm is described in the “Internal Validation” section of the supporting information (available online only).

RESULTS

Selection of SNPs used to calculate the GRS

Morrison et al.¹³ previously described the concept of a GRS for CHD that was based on 10 SNPs that were nominally associated ($P < 0.1$) with risk of CHD among white participants of ARIC when tested in a Cox model that adjusted for age and sex. To investigate a GRS for CHD that was based on those of the 10 SNPs with the highest probability of being truly associated with CHD, we selected only those of the 10 SNPs whose risk allele in ARIC whites had also been associated with CHD in at least two association studies of white patients other than ARIC. These SNPs were in *VAMP8* (rs1010), *KIF6* (rs20455), *MYH15* (rs3900940), *PALLD* (rs7439293), and *SNX19* (rs2298566) (Table 2). The SNPs in *VAMP8* and *KIF6* have been reported to

Table 2
Five SNPs in the genetic risk score panel: Association with CHD in ARIC whites

Gene (SNP)	Model	Incident CHD (%)	Total	HR ^a	95% CI	P	FPRP
<i>MYH15</i> (rs3900940)							
1125Ala	CC	128 (14.46)	885	1.47	1.21–1.78	<0.001	
1125Thr/Ala	CT	404 (10.16)	3,977	1.09	0.96–1.24	0.182	
1125Thr	TT	477 (9.97)	4,784	1	Ref		
	Additive			1.17	1.07–1.28	0.001	0.001–0.131
<i>KIF6</i> (rs20455)							
719Arg	GG	144 (11.50)	1,252	1.22	1.02–1.48	0.033	
719Trp/Arg	GA	474 (10.86)	4,363	1.12	0.98–1.28	0.092	
719Trp	AA	382 (9.73)	3,926	1	Ref		
	Additive			1.11	1.02–1.21	0.021	0.001–0.091
<i>VAMP8</i> (rs1010)							
Chr2: 64624922 (3' UTR)	CC	194 (11.18)	1,736	1.19	1.00–1.42	0.051	
	CT	481 (10.59)	4,540	1.1	0.96–1.26	0.179	
	TT	310 (9.61)	3,226	1	Ref		
	Additive			1.09	1.00–1.19	0.046	0.003–0.189
<i>PALLD</i> (rs7439293)							
Chr4: 2089007 (Intronic)	AA	398 (10.78)	3,691	1.29	1.07–1.57	0.011	
	AG	462 (10.52)	4,393	1.22	1.01–1.48	0.043	
	GG	124 (8.89)	1,395	1	Ref		
	Additive			1.11	1.02–1.22	0.017	0.017–0.722
<i>SNX19</i> (rs2298566)							
878Arg	CC	597 (11.34)	5,263	1.15	0.90–1.48	0.266	
878Leu/Arg	AC	330 (9.42)	3,505	0.99	0.76–1.28	0.92	
878Leu	AA	61 (9.62)	634	1	Ref		
	Additive			1.12	1.01–1.24	0.029	0.029–0.816

SNP annotation based on build 126, (human genome map 36) of the NCBI SNP database.

^aHR per allele and *P* values were estimated from Cox proportional hazards model adjusted for age (continuous) and sex.

MYH15, myosin, heavy chain 15; *KIF6*, kinesin family member 6; *VAMP8*, vesicle-associated membrane protein 8; *PALLD*, palladin, cytoskeletal associated protein; *SNX19*, sorting nexin 19; HR, hazard ratio; 95% CI, 95% confidence interval, 3' UTR, 3' untranslated region; FPRP, false-positive report probability.¹⁶

be associated with CHD in other studies: the ARIC risk allele of rs1010 in *VAMP8* was associated with MI in three case-control studies¹⁷ and the ARIC risk allele of rs20455 in *KIF6* was associated with CHD in the placebo arms of two statin treatment studies.¹⁸ For the other three SNPs (rs3900940 in *MYH15*, rs7439293 in *PALLD*, and rs2298566 in *SNX19*), the same risk allele was also associated with MI ($P < 0.1$) in the two case-control studies reported here (Table 8, supporting information, available online only) and with CHD in ARIC. We did not include 5 of the original 10 SNPs reported in Morrison et al.¹³ in the new GRS panel. Four of the SNPs that were not included had been initially found to be associated with CHD in one of the three case-control studies described by Morrison et al.¹³; however, these four SNPs were either associated with CHD in only a single case-control study or the ARIC risk allele was not the risk allele in the case-control studies (Table 6, supporting

information, available online only). Finally we did not include rs4994 (in *ADRB3*) in the new GRS panel because the ARIC risk allele of this SNP was not associated with CHD in two published studies.

For the five SNPs selected for the panel used to calculate the GRS, we then estimated the probability that the SNPs had been falsely associated with CHD in ARIC. These estimates of the false-positive report probability¹⁶ for each SNP take into account the prior probability of association, which was estimated using the results from association studies conducted before testing the SNP in ARIC. Because these prior probabilities are based on subjective assumptions, we report false-positive report probability based on a range of assumptions as suggested by Wacholder et al.¹⁶ These false-positive report probabilities suggest that the associations between CHD in ARIC whites and the SNPs in *KIF6*, *MYH15*, and *VAMP8* are unlikely to be due

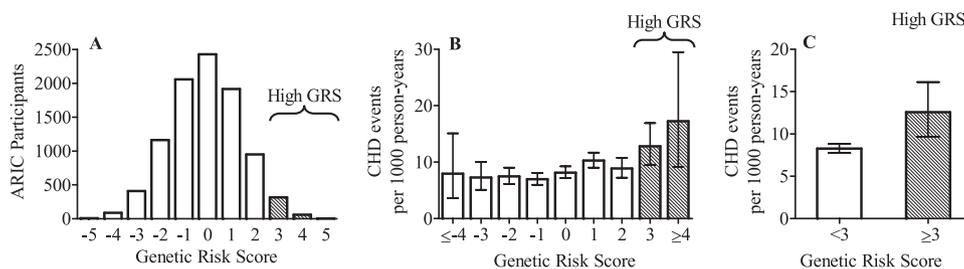


Fig. 1. The GRS in white participants of the ARIC study. A, Distribution of the GRS. B, The rate of incident coronary heart disease (CHD) per 1000 person-years for each GRS level. The -5 and -4 levels were combined because there were fewer than 10 events in the -5 level. Similarly, level 4 and level 5 were combined because there were fewer than 10 events in level 5. C, The rate of incident CHD per 1000 person-years in those with a GRS <3 and in those with a high GRS (≥3).

to chance (Table 2). This was also the case for the mid to low end of the false-positive report probability range for the SNPs in *PALLD* and *SNX19*; however, the high end of the range is consistent with chance association for these two SNPs.

Association of a high GRS with CHD

The genotypes of five SNPs (in *VAMP8*, *KIF6*, *MYH15*, *PALLD*, and *SNX19*) were used to calculate the GRS for each ARIC participant. A GRS was increased by 1 if the participant was homozygous for the risk allele, unchanged if heterozygous, and decreased by 1 if homozygous for the nonrisk allele. Therefore, individuals carrying all 10 possible risk alleles were assigned a GRS of 5 and those carrying no risk alleles were assigned a GRS of -5. The distribution of GRS in ARIC whites is shown in Figure 1A. A high GRS was defined as a GRS of 3 or higher because inspection of the relationship between GRS and rate of incident CHD revealed a marked increase in the CHD rate for those with a GRS of 3 or higher (Fig. 1B). The CHD rate in the high GRS group was 12.6 CHD events per 1000 person-years (95% confidence interval [CI] 9.6–16.1; Fig. 1C) whereas the CHD rate among those not in the high-GRS group (GRS of -5 to 2) was 8.3 (95% CI 7.7–8.8; Fig. 1C). The high-GRS group comprised 4% of the ARIC whites (Fig. 1A). A high GRS (3 or higher) was associated with incident CHD in ARIC whites (hazard ratio = 1.64; 95% CI 1.3–2.1; Table 3), and this risk estimate remained essentially unchanged after adjustment for traditional risk factors (hazard ratio = 1.57; 95% CI 1.2–2.0; Table 3). There was no significant interaction ($P \geq 0.09$) between the GRS and any of the traditional risk factors used for adjustment.

Because selection of the five SNPs used to calculate the GRS was partly based on data from the ARIC study, the hazard ratio of a high GRS might be higher in ARIC than would be expected in an external population. Therefore, we used a bootstrap internal validation procedure to investigate the extent to which the selection of the five SNPs might have resulted in an inflated hazard ratio estimate for a high GRS. For each bootstrap sample, we reselected a panel of SNPs used to calculate the GRS from the set of 51 SNPs that we had found to have the same risk allele for CHD in at least two studies other than ARIC. The risk alleles of these 51 SNPs are reported in Tables 7 and 8 of the supporting information (available online only), and the associations between these 51 SNPs and CHD in ARIC have been described by Morrison et al.¹³ After 2000 iterations of resampling, the five SNPs in the GRS panel were the most frequently chosen SNPs (data not shown). The results of the bootstrap resampling suggested that the hazard ratio for a high GRS might be 1.43 in an external population.

The NCEP guidelines define high-risk groups for traditional risk factors.³ To facilitate comparison of these high-risk groups to the group with a high GRS—specifically, comparison of the magnitude of their risk for CHD—for each traditional risk factor we compared the high-risk group with the rest of the participants and estimated hazard ratios for CHD in ARIC whites while adjusting for other risk factors. We found that the hazard ratio for a high GRS was of similar magnitude to the hazard ratios for most traditional risk factors that are considered in the NCEP guidelines (Table 4).

Table 3
Association of high GRS with incident CHD in ARIC

Subgroup	Incident CHD (%)	Total ^a	Unadjusted			Model 1 ^b			Model 2 ^c		
			HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Whites	934 (10.2)	9,129	1.64	1.27–2.12	<0.001	1.62	1.25–2.09	<0.001	1.57	1.21–2.04	0.001
African Americans	250 (8.3)	2,995	1.12	0.53–2.37	0.772	1.24	0.58–2.63	0.576	1.34	0.63–2.86	0.441

HR and P values for a high GRS (≥3) compared with not-high GRS. Four percent of ARIC whites had a high GRS, and 16.58% (61 of 368) of those with a high GRS had an event, compared with 9.96% (873 of 8761) of those with a not-high GRS that had an event.

^aNumber of participants in Model 2 analysis.

^bAdjusted for baseline age (continuous) and sex.

^cAdjusted as in Model 1 and also for diabetes, hypertension, current smoking status, HDL-C level (continuous), LDL-C level (continuous), and family history of CHD.

Table 4

Comparison of the hazard ratios of a high GRS and traditional risk factors

Risk factors	HR	95% CI
Diabetes	2.44	2.06–2.89
LDL-C \geq 130	1.66	1.44–1.91
Current smoking	1.58	1.37–1.82
High GRS	1.57	1.21–2.04
Women aged $>$ 55	1.53	1.19–1.96
Hypertension	1.51	1.31–1.73
HDL-C $<$ 40	1.50	1.31–1.72
Family history of CHD	1.45	1.25–1.67
Men aged $>$ 55	1.41	1.21–1.66

Each HR for incident CHD was adjusted for all other risk factors listed in this table. Cut points for traditional risk factors were based on the NCEP guidelines for CHD risk prediction,³ except for the age cut point for men, which was set to age $>$ 55 rather than age $>$ 45 because there were no men aged $<$ 45 enrolled in the ARIC study. All *P* values were $<$ 0.001.

HR, hazard ratio; 95% CI, 95% confidence interval.

Characterization of the GRS in the African American Participants of ARIC

We also tested the five SNP GRS, which had been defined based on the consistent association of SNPs in three studies of white populations, in the African American participants of ARIC. The allele frequencies for the SNPs in the GRS differed between African Americans and whites (Table 9 of the supporting information, available online only). In the African American ARIC participants, 2.3% had a high GRS (3 or higher), and the distribution of the GRS ranged from -4 to $+4$. No African Americans had a GRS of -5 or 5 because none were homozygous for all five SNPs. African Americans with a high GRS, compared with those without a high GRS, also had a higher risk of CHD after adjustment for traditional risk factors (hazard ratio = 1.34; 95% CI 0.63–2.86; Table 3), but this association was not statistically significant ($P = 0.441$).

DISCUSSION

We found that a high GRS—defined by five SNPs consistently associated with CHD in several studies—was associated with incident CHD after adjusting for traditional risk factors. Four percent of ARIC whites had a high GRS, and given the incidence of CHD events (an estimated 700,000 Americans will have a coronary event this year²⁴), the identification of 4% of the population who are at a high, but currently unrecognized, genetic risk of CHD could usefully guide preventive measures. To consider the GRS in the context of current CHD risk assessment practices, we compared it with each traditional risk factor considered by the NCEP guidelines for risk assessment and treatment recommendations: age, sex, LDL-C, HDL-C, hypertension, smoking, diabetes, and family history of CHD. Because these guidelines make recommendations based on whether patients are in the high-risk groups, we determined the hazard ratios for CHD of the high-risk groups for these

traditional risk factors. These hazard ratios were similar in magnitude to the hazard ratio of the high-GRS group (Table 4). Thus, for those individuals with a high GRS, the magnitude of the incremental risk associated with the GRS is similar to those associated with the traditional risk factors.

The internal validation by bootstrap resampling of ARIC whites gave some indication of what might be expected when this GRS is tested in other populations similar to ARIC. Bootstrap resampling suggested that a hazard ratio of 1.43 might be expected in an external population. However, the association of a high GRS with CHD needs to be tested in other large population-based studies to fully validate our observations in the ARIC cohort.

In the African American participants of ARIC the adjusted hazard ratio for a high GRS was 1.34, which is similar to the hazard ratio suggested by the bootstrap analysis for an external population; however, the association was not significant due in part to the lower power to detect associations of the GRS in the African American population. The power in the African American analysis is affected by the smaller number of participants, and the smaller percentage of African Americans with a high GRS. The smaller percentage of African American participants with a high GRS may be due to the differences in allele frequencies between African Americans and whites for the five SNPs used in the GRS panel.

In calculating the GRS, we used a simple algorithm described by Morrison et al.¹³ We assigned equal weight to each SNP because the evidence that could be used to assign different weights was not compelling: all five SNPs had similar hazard ratios for CHD in ARIC (Table 2). Similarly, because the apparent genetic model for SNPs can be different in different populations, we assigned equal weight to all risk alleles; thus, being a heterozygote for two SNPs would make the same contribution to the GRS as being a risk allele homozygote for one SNP and a nonrisk homozygote for a second SNP. This additive model for the risk alleles has reasonable power when the true genetic model is unknown.²⁵ The data from ARIC and the two case-control association studies described here was entirely consistent with the additive model for the SNPs in *MYH15*, *PALLD*, and *VAMP8*. The SNP in *KIF6* fits the additive model in ARIC, but fits the dominant model in other association studies.¹⁸ The data for the SNP in *SNX19* tended to suggest a recessive model. If supported by further investigation in other populations, the GRS algorithm could be modified to include different weights or specific genetic models, for example, a recessive model for the SNP in *SNX19*.

Because the SNPs that were tested in ARIC had different levels of prior evidence, we used a Bayesian approach that takes into consideration not only the observed *P* value, but also the power of the study to detect association, and the prior probability of the SNP to be associated with disease to estimate the likelihood that the association between the SNP and CHD in ARIC is a false-positive association¹⁶ (Table 2). The range of false-positive report probabilities we have estimated for the five SNPs that are used to calculate the GRS indicates that it is unlikely that they are false-positive findings in ARIC. Calculat-

ing the false-positive report probability required testing the association between the SNP and CHD in ARIC and estimating the prior probability that the SNP was associated with CHD. To test the association between each SNP and CHD in ARIC, we used an additive genetic model; that is, we assumed that the risk associated with the heterozygote would be intermediate (on the log scale) between the two homozygotes. We used an additive model because it is a reasonable approximation for both recessive and dominant models.²⁵ Using an additive model avoided the multiple hypothesis testing that would have resulted from testing each SNP in additive, dominant, and recessive models and avoided using the limited available data to make a definitive choice of a model other than the additive model.

The ranges of false-positive report probabilities for the five SNPs in the GRS panel also depended on estimates of the prior probabilities of association with CHD. The prior probability is the probability that an SNP was associated with CHD before it was tested in ARIC—the higher the prior probability, the lower the probability that the SNP is a false-positive. In estimating ranges of prior probability, we divided five SNPs into two groups. For the group of SNPs whose prior association with CHD had been described in published studies that adjusted for multiple hypothesis testing, we assumed a higher prior probability: False discovery rate analysis was used to account for multiple testing for the SNP in *VAMP8*,¹⁷ and the Bonferroni method was used for the SNP in *KIF6*.¹⁸ We assumed a lower prior probability of association for the SNPs in *MYH15*, *PALLD*, and *SNX19*; these were nominally associated with CHD in the two case-control studies of MI described in this report (Table 8, supporting information, available online only).

It remains to be determined how the five SNPs in the GRS panel affect the pathogenesis of CHD; however, it is interesting that all five genes could be involved in intracellular vesicle trafficking. *VAMP8* encodes vesicle-associated membrane protein 8, a member of the SNARE complex, that mediates membrane-membrane fusion and platelet degranulation.^{26,27} *SNX19* encodes sorting nexin 19, a member of the nexin family that is involved in intracellular trafficking of membrane vesicles.²⁸ The sorting nexin 19 protein product has been shown to interact with islet antigen-2, a transmembrane protein tyrosine phosphatase associated with dense core secretory vesicles and a major autoantigen in type 1 diabetes.²⁹ The functions of the other three genes are related to aspects of cytoskeletal organization and movement. *MYH15* encodes myosin heavy chain 15, a myosin heavy chain polypeptide that is evolutionarily distinct from skeletal and cardiac myosins.³⁰ *PALLD* (palladin, cytoskeletal associated protein) encodes a cytoskeleton protein involved in actin reorganization,³¹ and *KIF6* encodes kinesin family member 6, a member of the kinesin superfamily of proteins; kinesins are involved in the intracellular transport of membrane organelles, protein complexes, and mRNAs.³²

One potential limitation of this study is that the current GRS panel has not yet been externally validated in additional prospective studies, and additional well-validated SNPs might im-

prove the GRS. For example, carriers of an intergenic region on chromosome 9 (~75% of the white population) have recently been shown to be associated with CHD in multiple large studies,¹¹ and adding the chromosome 9 variant (rs10757274) to the GRS panel of SNPs increases the fraction of the population in the high GRS group (to ~20%) while slightly decreasing the hazard ratio of a high GRS (data not shown). For the current GRS we included all the SNPs that met our inclusion criteria, and we used a simple weighting scheme. Other risk scores with more sophisticated weighting schemes could be envisioned. However, we believe that data from additional studies would be needed to further optimize the GRS. A further limitation of this study is that it included a comparison to only the traditional risk factors considered by the NCEP guidelines; future studies could also compare the GRS with emerging risk factors.³³ Finally the population-attributable risk of a high GRS is lower than that of the more prevalent traditional risk factors.

In conclusion, a high GRS defined by five SNPs is associated with a 57% increased risk of CHD. Given the prevalence of CHD, identifying 4% of the population who are at greater risk could have a significant impact on public health.

ACKNOWLEDGMENTS

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. Celera funded genotyping and statistical analysis. The authors thank the staff and participants of the ARIC study and A. Arellano, B. Young, D. Lew, D. Leong, J. Catanese, M. Short, C. Tong, J. Louie, and D. Liu for their help in collecting and analyzing the genotype data, K. Lawson for data analysis, and J. Sninsky, T. White, B. Young, A. Begovich, A. Grupe, and K. Lau, for insightful comments.

References

- Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* 1997;349:1498–1504.
- Grundy SM, Cleeman JI, Merz CN, Brewer HB Jr, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation* 2004;110:227–239.
- 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) final report. *Circulation* 2002;106:3143–3421.
- Akosah KO, Schaper A, Cogbill C, Schoenfeld P. Preventing myocardial infarction in the young adult in the first place: how do the National Cholesterol Education Panel III guidelines perform? *J Am Coll Cardiol* 2003;41:1475–1479.
- Grundy SM. Metabolic syndrome: connecting and reconciling cardiovascular and diabetes worlds. *J Am Coll Cardiol* 2006;47:1093–1100.
- Marenberg ME, Risch N, Berkman LF, Floderus B, et al. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med* 1994;330:1041–1046.
- Shea S, Ottman R, Gabrieli C, Stein Z, et al. Family history as an independent risk factor for coronary artery disease. *J Am Coll Cardiol* 1984;4:793–801.
- Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* 2002;252:247–254.
- Casas JP, Cooper J, Miller GJ, Hingorani AD, et al. Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies. *Ann Hum Genet* 2006;70:145–169.
- Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA* 2007;297:1551–1561.

11. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007;316:1491–1493.
12. Weedon MN, McCarthy MI, Hitman G, Walker M, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med* 2006;3:e374.
13. Morrison AC, Bare LA, Chambless LE, Ellis SG, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* 2007;166:28–35.
14. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* 1989;129:687–702.
15. White AD, Folsom AR, Chambless LE, Sharret AR, 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;49:223–233.
16. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434–442.
17. Shiffman D, Rowland CM, Louie JZ, Luke MM, et al. Gene variants of VAMP8 and HNRPUL1 are associated with early-onset myocardial infarction. *Arterioscler Thromb Vasc Biol* 2006;26:1613–1618.
18. Iakoubova OA, Tong CH, Rowland CM, Kirchgessner TG, et al. Association of the Trp719Arg polymorphism in kinesin-like protein 6 with myocardial infarction and coronary heart disease in two prospective studies: the Cholesterol and Recurrent Events Study (CARE) and the West of Scotland Coronary Prevention Study (WOSCOPS). *J Am Coll Cardiol*. In press.
19. Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). *Am J Epidemiol* 1990;131:373–375.
20. Cox DR. Regression models and life-tables (with discussion). *J Royal Stat Soc B* 1972;34:187–220.
21. Kleinbaum D. Survival analysis: a self-learning text, 2nd ed. New York: Springer-Verlag, 2005:83–166.
22. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–387.
23. Harrell FE Jr. Regression modeling strategies. New York: Springer-Verlag, 2001:91.
24. Rosamond W, Flegal K, Friday G, Furie K, et al. Heart disease and stroke statistics—2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2007;115:e69–e171.
25. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006;7:781–791.
26. Polgar J, Chung SH, Reed GL. Vesicle-associated membrane protein 3 (VAMP-3) and VAMP-8 are present in human platelets and are required for granule secretion. *Blood* 2002;100:1081–1083.
27. Ren Q, Barber HK, Crawford GL, Karim ZA, et al. Endobrevin/VAMP-8 is the primary v-SNARE for the platelet release reaction. *Mol Biol Cell* 2006;18:24–33.
28. Worby CA, Dixon JE. Sorting out the cellular functions of sorting nexins. *Nat Rev Mol Cell Biol* 2002;3:919–931.
29. Hofmann MA, Yang Q, Harja E, Kedia P, et al. The RAGE Gly82Ser polymorphism is not associated with cardiovascular disease in the Framingham offspring study. *Atherosclerosis* 2005;182:301–305.
30. McGuigan K, Phillips PC, Postlethwait JH. Evolution of sarcomeric myosin heavy chain genes: evidence from fish. *Mol Biol Evol* 2004;21:1042–1056.
31. Mykkanen OM, Gronholm M, Ronty M, Lalowski M, et al. Characterization of human palladin, a microfilament-associated protein. *Mol Biol Cell* 2001;12:3060–3073.
32. Miki H, Setou M, Kaneshiro K, Hirokawa N. All kinesin superfamily protein, KIF, genes in mouse and human. *Proc Natl Acad Sci U S A* 2001;98:7004–7011.
33. Folsom AR, Chambless LE, Ballantyne CM, Coresh J, 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;166:1368–1373.