Original Contribution

Biogeographic Ancestry, Self-Identified Race, and Admixture-Phenotype Associations in the Heart SCORE Study

Indrani Halder*, Kevin E. Kip, Suresh R. Mulukutla, Aryan N. Aiyer, Oscar C. Marroquin, Gordon S. Huggins, and Steven E. Reis

* Correspondence to Dr. Indrani Halder, Heart and Vascular Institute, School of Medicine, University of Pittsburgh, 1704 Biosciences Tower, 200 Lothrop Street, Pittsburgh, PA 15213 (e-mail: halderi@upmc.edu).

Large epidemiologic studies examining differences in cardiovascular disease (CVD) risk factor profiles between European Americans and African Americans have exclusively used self-identified race (SIR) to classify individuals. Recent genetic epidemiology studies of some CVD risk factors have suggested that biogeographic ancestry (BGA) may be a better predictor of CVD risk than SIR. This hypothesis was investigated in 464 African Americans and 771 European Americans enrolled in the Heart Strategies Concentrating on Risk Evaluation (Heart SCORE) Study in March and April 2010. Individual West African and European BGA were ascertained by means of a panel of 1,595 genetic ancestry informative markers. Individual BGA varied significantly among African Americans and to a lesser extent among European Americans. In the total cohort, BGA was not found to be a better predictor of CVD risk factors than SIR. Both measures predicted differences in the presence of the metabolic syndrome, waist circumference, triglycerides, body mass index, very low density lipoprotein cholesterol, lipoprotein A, and systolic and diastolic blood pressure between European Americans and African Americans. These results suggest that for most nongenetic cardiovascular epidemiology studies, SIR is sufficient for predicting CVD risk factor differences between European Americans and African Americans. However, higher body mass index and diastolic blood pressure were significantly associated with West African BGA among African Americans, suggesting that BGA should be considered in genetic cardiovascular epidemiology studies carried out among African Americans.

African continental ancestry group; cardiovascular diseases; continental population groups; population genetics; race

Race-related differences in various cardiovascular and metabolic phenotypes are well documented. Many cardiovascular disease (CVD) risk factors such as hypertension, obesity, and C-reactive protein are disproportionately higher in African Americans compared with European Americans (1–8). In contrast, African Americans have more favorable lipid profiles (9–11). However, in all of these epidemiologic studies, "self-identified race" (SIR) is used to classify individuals into distinct categories, which fails to take into account the substantial heterogeneity in genetic heritage that exists in populations with mixed ancestry (12–18). Since most CVD risk factors have a complex genetic etiology, disregarding the genetic heritage of individuals may produce underestimation of race-related differences in CVD risk.

Starting in the late 1400s, the European colonial period in the New World brought together populations that had previously been geographically isolated (e.g., Europeans, West Africans, Native Americans). Gene flow among these groups resulted in admixed populations with variable distributions of chromosomes comprised of nonuniform segments from

different ancestral (i.e., geographic) populations (19–21). Studies of biogeographic ancestry (BGA) have demonstrated that the African-American gene pool in the United States reflects a mixture of European, West African, and, to a lesser extent, Native American ancestry (13–18). Individual admixture proportions vary substantially among self-identified African Americans from different regions of the United States, with a wide variation in BGA profiles not only in samples drawn from different geographic locations but also among samples drawn from a single geographic region (13, 14, 22). Compared with African Americans, BGA in self-identified European Americans has been studied less often but, when examined, has been shown to vary, albeit to a lesser extent than in African Americans (16–18, 23, 24). Taken together, such studies suggest that estimating BGA using ancestry informative markers (AIMs), generally markers with large allele frequency differences between ancestral populations, may be a better descriptor of ancestry than SIR.

Since it is likely that biologic mechanisms leading to increased CVD risk in African Americans are related to genes that may also be associated with African BGA, racial differences in CVD may be partly attributable to BGA variation. This theory is supported by studies that have demonstrated associations between cardiometabolic phenotypes (e.g., body mass index (BMI), lipid levels, blood pressure, and coronary artery calcification (CAC)) and BGA in African Americans (8, 14, 15, 25–29). However, no studies have examined how group affiliation as measured by BGA compares with SIR for predicting several CVD risk factors in a single cohort. We hypothesized that if CVD risk, as measured by the prevalence of different risk factors, varies as a function of BGA, it may be possible to use such genomic information to better characterize race-related differences in CVD. Therefore, we undertook the current study to investigate the BGA profiles of self-identified African-American and European-American subjects included in the Heart Strategies Concentrating on Risk Evaluation (Heart SCORE) Study and compare SIR with BGA in predicting race-related differences in a range of cardiometabolic outcomes. We also examined BGA-phenotype associations in the African Americans to replicate previous findings (8, 14, 15, 25–29).

MATERIALS AND METHODS

Study samples

Subjects included in this study are participants in the Heart SCORE project (30, 31), a prospective longitudinal cohort study of 2,000 adults (ages 45–74 years at study inclusion; 64% female; 53% European-American, 43% African-American, and 4% of "other" races) residing in the Greater Pittsburgh, Pennsylvania, metropolitan area. Participants have undergone extensive baseline data collection (demographic history, several traditional and emerging CVD risk factors, physical activity assessments) and are being followed annually. Exclusion criteria for this analysis included known comorbid conditions expected to reduce life expectancy to less than 4 years. The current study, conducted between July 2009 and April 2010, was based on the first 464 African-American and 771 European-American subjects who provided DNA samples

and were successfully genotyped with less than 5% missing data. All subjects provided written informed consent, and the study was approved by the Institutional Review Board of the University of Pittsburgh.

Phenotype data collection

At the baseline visit, participants' demographic information and medical history were collected; physical examinations were conducted to record vital signs and anthropometric measures of body fat distribution. Education was measured as the total number of years of schooling. Annual income was coded in 5 categories: <$10,000, $10,000–<$20,000, $20,000–<$40,000, $40,000–<$80,000, and ≥$80,000. Height and weight measures were used to calculate BMI (weight (kg)/height (m)^2). Smoking habits were coded as current smoking versus previous smoking/nonsmoking. Physical activity was measured using the Lipid Research Clinics physical activity questionnaire (32). Fasting blood samples were obtained for serologic and DNA analysis. Depressive symptoms were measured using the Center for Epidemiologic Studies Depression Scale (33).

Metabolic syndrome was defined using the National Cholesterol Education Program/Adult Treatment Panel III criteria (34) as the presence of 3 or more of the following: fasting serum glucose concentration ≥100 mg/dL or use of hypoglycemic medication; serum triglyceride concentration ≥150 mg/dL or use of lipid-lowering medication; serum high density lipoprotein (HDL) cholesterol concentration <40 mg/dL in men or <50 mg/dL in women or use of lipid-lowering medication; blood pressure ≥130/85 mm Hg or use of antihypertensive medication; and waist circumference >102 cm in men or >88 cm in women.

Lipoprotein levels were assayed in fasting venous blood as described previously (31). A vertical auto profile technique (Atherotech Diagnostics Lab, Birmingham, Alabama) was used to quantify concentrations of total, HDL, low density lipoprotein (LDL), very low density lipoprotein, and intermediate density lipoprotein cholesterol, triglycerides, and lipoprotein A. The vertical auto profile technique also classified LDL cholesterol levels into patterns A, A/B, and B based on increasing concentrations of LDL$_3$ cholesterol (35). HDL cholesterol values were log-transformed for normalization. Body density reflects the proportion of body fat present, with higher density indicating lower fat mass. Since it is a proportion, like BMI, it is expressed in units of density. Body density was computed by means of generalized caliper equations in men using chest, abdomen, and thigh skinfold measures (36) and in women using triceps, suprailiac, and thigh skinfold measures (37).

CAC score, measured using electron beam computed tomography, was available for 250 European Americans and 167 African Americans. CAC values were classified using Agatston scores as low (10–100), intermediate (101–400), or significant (>400) (38).

DNA isolation, SNP selection, and genotyping

Blood samples for genotyping were collected in 10 mmol/L ethylenediaminetetraacetic acid, and DNA was

isolated following standard protocols. A panel of 1,698 single nucleotide polymorphism (SNP) AIMs, which are included as part of the Illumina CARe iSelect cardiovascular array (Illumina Inc., San Diego, California), were genotyped in all samples. This array is a customized 50,000-SNP chip that assays multiple polymorphisms in 2,100 genes and was developed specifically for cardiometabolic phenotypes (39). The AIMs included in the chip were selected from published reports (23, 24) and are suitable for both European-American and African-American samples. Detailed information on the AIMs included in the panel is available on the IBC CardioChip website (http://www.bmic.upenn.edu/cvdsnp/), which lists the chromosomal position and functional status (coding, intronic, 3′- or 5′-untranslated region, etc.) of each AIM.

Genetic analyses

Hardy-Weinberg equilibrium was assessed using Genepop software (http://genepop.curtin.edu.au/), version 4 (40). Ancestral allele frequencies were ascertained in HapMap (www.hapmap.org) European (CEU: European Americans with primarily Western European ancestry) and West African (YRI: Yorubans from Nigeria) samples and are abbreviated henceforth as "EU" and "AF" to designate the 2 ancestral groups. Individual BGA estimates were computed using MLIAE (Maximum Likelihood Individual Admixture Estimator) software (17), which implements a maximum likelihood method for inferring individual admixture (41). Percentage of African biogeographic ancestry (AF-BGA) was used as a predictor in all subsequent analyses. The presence of significant admixture stratification was tested by means of the Individual Ancestry Correlation Test (16–18). This test is a split-half reliability test in which all items that measure the same construct (i.e., individual admixture) are randomly divided into 2 sets. Each half set is then used to infer individual ancestry. The split-half reliability estimate is a correlation between estimates obtained with each half set. High correlation between estimates obtained with the 2 panels indicates reliability of the admixture estimates and reinforces that the variability in admixture estimates seen in this sample is not due to chance effects. Thereby, this test confirms the presence of significant stratification in the sample. For this analysis, 50 independent tests were conducted using different partially overlapping combinations of markers in each set.

Statistical analyses

All statistical analyses were performed in SPSS, version 16 (SPSS, Inc., Chicago, Illinois). Differences in phenotype measures between races were compared by means of independent sample t tests or chi-square tests. The predictive values of SIR and AF-BGA were compared in 2 sets of models. SIR was used as a binary predictor in the first set of models. In the second set, African-American subjects were grouped into 4 categories based on quartiles of the AF-BGA distribution (corresponding to $\leq63\%$, 64%–72%, 73%–78%, and $\geq79\%$ AF-BGA) and were compared with all self-identified European-American subjects as a fifth group. Two additional models were examined for AF-BGA and are shown in Web

Table 1 (available on the *Journal*'s website (http://aje.oxfordjournals.org/)): 1) AF-BGA as a continuous variable and 2) African Americans grouped into 10 categories based on deciles of the AF-BGA distribution and compared with all European Americans. Covariates included in each model are described below. In secondary analyses, all models were examined separately in males and females, unless gender was used to define a metabolic syndrome criterion.

Age, smoking status, current physical activity level, education, and income were included as covariates in all analysis of variance, logistic regression, and linear regression models. For the metabolic syndrome, subjects were classified as "normal," "with metabolic syndrome," or "with a history of diabetes" using National Cholesterol Education Program criteria and were compared using analysis of variance, with gender as an additional covariate. Individual components of the metabolic syndrome were evaluated by means of logistic regression, with the following additional covariates: use of lipid-lowering medication (for waist circumference); BMI (for HDL cholesterol and triglycerides); and gender and BMI (for hypertension and glucose). Linear regression was used to examine systolic and diastolic blood pressure (with gender, BMI, and use of antihypertensive medication as additional covariates) and LDL cholesterol, very low density lipoprotein cholesterol, and lipoprotein A (with gender, BMI, and use of lipid-lowering medication as additional covariates). BMI was examined with gender and use of lipid-lowering medication as additional covariates. Body density was examined as a continuous variable with use of lipid-lowering medication as an additional covariate. In secondary analyses, the models were stratified by sex.

For CAC, a logistic regression model was used to compare low scorers with intermediate/significant scorers for association with SIR or AF-BGA after adjustment for age, gender, BMI, smoking status, alcohol use, depression symptoms (measured using the Center for Epidemiologic Studies Depression Scale), aspirin use, current level of physical activity, history of hypertension, diabetes, creatinine level, use of antihypertensive and lipid-lowering medications, HDL cholesterol, LDL_3 cholesterol, and LDL cholesterol pattern (A, B, or A/B), following a previous model (31).

Associations between AF-BGA and phenotype were examined separately in African Americans using similar analysis of variance, logistic regression, or linear regression models and the same covariates as in previous analyses, with AF-BGA as a continuous predictor. Waist circumference, HDL cholesterol, triglycerides, and fasting glucose were examined as continuous variables.

To account for multiple testing, a false discovery rate (FDR) method was implemented (42), and all FDR-adjusted P values are reported. FDR-adjusted P values less than 0.05 were considered significant.

RESULTS

Sample characteristics and individual BGA distribution in Heart SCORE

Demographic and clinical characteristics of subjects are shown in Table 1. As reported previously in Heart

Table 1.   Characteristics of European-American and African-American Subjects in the Heart SCORE Study, July 2009–April 2010

	African Americans (n = 464)		European Americans (n = 771)		P Value[a]
	%	Mean (SD)	%	Mean (SD)	
Age, years		58 (7.3)		59.6 (7.4)	<0.005
Female gender	71		62		<0.005
Current smoking	11		8		NS
Annual income					<0.005
<$10,000	6		3		
$10,000–$19,999.99	18		7		
$20,000–$39,999.99	32		26		
$40,000–$79,999.99	34		34		
≥$80,000	9		30		
Body mass index[b]		32.4 (6.6)		28.6 (5.3)	<0.005
Metabolic status					<0.005
Normal	71		65		
Metabolic syndrome	23		19		
Diabetes	6		16		
Systolic blood pressure, mm Hg		139.9 (19.9)		133.6 (18.6)	<0.005
Diastolic blood pressure, mm Hg		82.8 (9.8)		79.3 (10)	<0.005
Hypertension	72		59		<0.005
Total cholesterol, mg/dL		208.9 (43.2)		216.1 (41.5)	<0.005
Triglycerides, mg/dL		109.3 (62.1)		136.6 (89.4)	<0.005
Glucose, mg/dL		102 (34.4)		96.4 (17.9)	<0.01
Waist circumference, cm		99.8 (15.8)		93.8 (13.7)	<0.005
HDL cholesterol, mg/dL		59.2 (14.7)		56.7 (15.1)	0.005
LDL cholesterol, mg/dL		138.7 (37.4)		134.1 (34.3)	0.02
VLDL cholesterol, mg/dL		11.1 (7)		15.3 (11.3)	<0.005
Lipoprotein A, mg/dL		10.2 (5.4)		8.5 (4.7)	<0.005
Coronary artery calcification[c]		130.6 (316.4)		206.8 (539.4)	NS

Abbreviations: HDL, high density lipoprotein; Heart SCORE, Heart Strategies Concentrating on Risk Evaluation; LDL, low density lipoprotein; NS, not significant; SD, standard deviation; VLDL, very low density lipoprotein.

[a] False discovery rate-adjusted, 2-sided P value. Significance was set at $P < 0.05$.

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

[c] Coronary artery calcification values were classified using Agatston scores as low (10–100), intermediate (101–400), or significant (>400).

SCORE (30, 31), African Americans were significantly younger, had higher BMIs, had higher prevalences of hypertension and diabetes, and were more often current smokers than European Americans. African Americans were less likely to engage in moderate-to-strenuous physical activity or use lipid-lowering medication but were significantly more likely to use antihypertensive medication. The 250 European Americans and 167 African Americans for whom CAC measures were available had similar demographic and clinical characteristics as the total Heart SCORE sample.

All subjects were genotyped for 1,698 AIMs. Ancestral allele frequencies of 103 AIMs were not available in HapMap and were not used in analysis. Of the final 1,595 AIMs, 301 markers in African Americans and 56 markers in European Americans were not in Hardy-Weinberg equilibrium (FDR-adjusted $P < 0.05$). Since no single locus was examined independently, lack of agreement with Hardy-Weinberg equilibrium was not a criterion for excluding markers from BGA analyses.

In African Americans, AF-BGA ranged between 9% and 89% (mean = 68% (standard error, 0.004)) and showed substantial variation (Figure 1). Females ($n = 329$) had higher mean AF-BGA than males ($n = 153$) (69% vs. 65%; $P = 0.03$). The Individual Ancestry Correlation Test showed significant correlations in admixture estimates obtained with nonsyntenic marker panels ($R^2 = 93\%$; $P < 0.001$), indicating the presence of admixture stratification in this sample. Twenty-one subjects (11 females) had 9%–15% AF-BGA. These individuals did not differ significantly
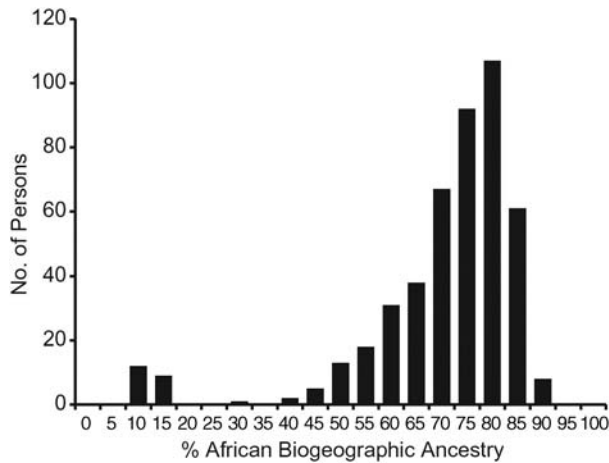
Figure 1. Distribution of West African biogeographic ancestry among African Americans in the Heart SCORE (Heart Strategies Concentrating on Risk Evaluation) Study, July 2009–April 2010. Biogeographic ancestry was estimated using 1,595 ancestry informative markers.

from others with regard to demographic or clinical measures. Mean AF-BGA for these 21 subjects was 11% in both genders. Mean AF-BGA after exclusion of these persons was 71% (standard error, 0.007). We conducted additional tests to investigate the BGA of these individuals; results are reported in the Web Appendix and in Web Figure 1.

Individual percentage of European biogeographic ancestry (EU-BGA) in European Americans ranged from 14% to 91% (mean EU-BGA = 87%), showed significantly less variation, and did not differ significantly between genders. Most European Americans (>95%) had very low levels (<15%) of non-European ancestry (Web Figure 2). No evidence of admixture stratification was detected in this group (Individual Ancestry Correlation Test: $P > 0.05$).

Predictive power of SIR- vs. BGA-based categorization

The purpose of these analyses was to compare whether SIR or BGA showed a significant association with a phenotype, not to test the strength of the association itself. There was no difference in predictive power between SIR and AF-BGA for 12 of the 15 outcomes studied in the total sample (Tables 2 and 3). Both categorization schemes yielded similar results (i.e., the presence or absence of significant association after covariate adjustment). Metabolic syndrome, waist circumference, triglycerides, BMI, very low density lipoprotein cholesterol, lipoprotein A, body composition, systolic blood pressure, and diastolic blood pressure were associated with both SIR and BGA, while total cholesterol, LDL cholesterol, and glucose were not associated with either. SIR but not BGA predicted HDL cholesterol, hypertension, and CAC ($P < 0.05$). As shown in Web Table 1, BGA and SIR demonstrated similar predictive abilities when AF-BGA deciles were used, which persisted even after removal of African Americans with <15%

AF-BGA and European Americans with <85% EU-BGA (Web Table 2). AF-BGA as a continuous variable significantly predicted all risk factors except for the presence of the metabolic syndrome, high glucose levels, and CAC (Web Table 1).

In secondary, gender-specific analyses, both SIR and BGA showed similar results (Table 4 and 5). However, AF-BGA was associated with the presence of the metabolic syndrome in males and diastolic blood pressure in females. In contrast, SIR alone was associated with triglycerides and CAC in males and systolic blood pressure in females.

Association of AF-BGA with BMI and diastolic blood pressure in African Americans

Within the group of self-identified African Americans, AF-BGA showed a positive association with both BMI ($\beta = 0.004$, $t = 2.214$, $P = 0.03$; Web Figure 3) and diastolic blood pressure ($\beta = 6.078$, $t = 2.18$, $P = 0.02$; Web Figure 4) and in both cases explained 1% of the variation in the phenotypes after other biologic and demographic covariates were accounted for. The association with BMI persisted after removal of persons with less than 15% AF-BGA from the analysis ($\beta = 0.61$, $t = 1.98$, $P = 0.046$). None of the other continuous or binary variables showed any association with AF-BGA (FDR-adjusted $P > 0.05$).

DISCUSSION

This study investigated whether BGA-based categorizations may be more accurate than SIR as predictors of cardiovascular phenotypes. Our results indicated strong coherence between SIR and BGA as predictors of CVD risk factors. Most phenotypes were either associated with both methods of classifying race or not associated with either method, except for high blood pressure, CAC, and higher HDL cholesterol, which were predicted by binary race alone. Although previous studies have compared the predictive power of BGA with SIR for a single phenotype (43, 44), to our knowledge, this is the first study that has comprehensively examined the predictive power of SIR versus BGA for a wide range of cardiometabolic phenotypes, and it shows SIR to be comparable to BGA in predicting racial differences in CVD risk factors. Limited evidence supporting the hypothesis was obtained in gender-specific analyses, where BGA was a better predictor of the metabolic syndrome in males and diastolic blood pressure in females. However, SIR was a better predictor of triglycerides and CAC in males and systolic blood pressure in females. Overall, these analyses did not show any strong trends favoring the use of BGA over SIR across a range of cardiometabolic outcomes in either gender.

The use of AIMs and BGA in many genetic studies raises the question of whether a nongenetic study is likely to gain from BGA measures as opposed to self-identified race. BGA measures characterize individuals along a continuum of ancestry, without identifying discrete subgroups. While this approach is a strength when studying a single population like African Americans, its utility in

Table 2. Comparison Between Self-Identified Race- and Biogeographic Ancestry-based Categories in Predicting the Risk of Metabolic Syndrome and Its Components in the Total Sample, Heart SCORE Study, July 2009–April 2010

	Self-Identified Race[a]			Biogeographic Ancestry[b]		
	OR[c]	95% CI	P Value[d]	OR	95% CI	P Value
Metabolic syndrome[e]			<0.0001			0.01
Waist circumference[f]	1.9	1.4, 2.4	<0.0005	1.2	1.1, 1.3	<0.0005
HDL cholesterol[g]	0.6	0.4, 0.8	0.002	0.9	0.8, 1.0	0.08
Glucose[h]	1	0.8, 1.4	0.7	1.1	0.9, 1.2	0.08
Triglycerides[i]	0.4	0.3, 0.6	<0.0005	0.7	0.6, 0.8	<0.0005
Hypertension[j]	1.5	1.1, 1.9	0.04	1.1	0.9, 1.2	0.08
Coronary artery calcification[k]	0.4	0.2, 0.9	0.04	0.7	0.6, 1.0	0.12

Abbreviations: CI, confidence interval; HDL, high density lipoprotein; Heart SCORE, Heart Strategies Concentrating on Risk Evaluation; OR, odds ratio.

[a] Self-identified African Americans compared with self-identified European Americans (reference group).

[b] All European Americans were coded as a single group (reference group) and compared with African Americans grouped into quartiles of biogeographic ancestry based on increasing percentage of African biogeographic ancestry.

[c] Odds ratio for the presence of metabolic syndrome criteria.

[d] False discovery rate-adjusted, 2-sided P value. Significance was set at $P < 0.05$.

[e] Metabolic syndrome was defined according to National Cholesterol Education Program criteria; subjects were classified as "normal" (reference group), "with metabolic syndrome" (group 1), or "with a history of diabetes" (group 2) and compared using analysis of variance (self-identified race: $F = 16.1$ (1 df); biogeographic ancestry: $F = 4.6$ (4 df)), with age, gender, smoking status, current physical activity level, education, and income as covariates. Individual components of the metabolic syndrome were evaluated by means of logistic regression, with age, smoking status, current physical activity level, education, and income as covariates in all models.

[f] Waist circumference >102 cm in men and >88 cm in women; use of lipid-lowering medication was used as an additional covariate.

[g] Serum HDL cholesterol concentration <40 mg/dL in men or <50 mg/dL in women or use of lipid-lowering medication; body mass index was used as an additional covariate.

[h] Fasting serum glucose concentration ≥100 mg/dL or use of hypoglycemic medication; body mass index and gender were used as additional covariates.

[i] Serum triglyceride concentration ≥150 mg/dL or use of lipid-lowering medication; body mass index and gender were used as additional covariates.

[j] Blood pressure ≥130/85 mm Hg or use of antihypertensive medication; body mass index and gender were used as additional covariates.

[k] Persons with low (10–100) Agatston scores were compared with those with intermediate (101–400) and significant (>400) scores.

understanding racial differences in disease risk in a multi-population model is not well documented or described. Our analysis had an important drawback, and thus the differences observed between SIR and BGA need to be interpreted with caution. A model which used an ancestry measure as a continuous variable (and not as groups based on ancestry proportions as we have done) could have been compared with a model with binary race only if all possible ancestral groups contributing to the 2 populations studied here had been used. Thus, if we had had ancestral allele frequency information on Native Americans and Asians (the 2 other large continental ancestral groups with potential contributions to the US population) in addition to the European and African ancestral groups, we could have used a global model of admixture and parsed each individual's ancestry on the basis of such a global model. In that case, we might have used a continuous measure of ancestry to compare BGA directly with SIR. For our study, which had almost 2 times as many European Americans as African Americans,

BGA estimates were skewed heavily in favor of European Americans, thereby biasing the associations with the continuous BGA measure, as seen by the significant P values obtained with all but 3 of the risk factors studied (see Web Table 1). BGA in European Americans is better described by a model which includes Native Americans and East Asians as ancestral groups (16, 17). We chose the 4-group model for categorizing African Americans, since using quartiles of African ancestry provided us with several large, approximately similar-sized groups. In addition, there is precedence in the literature for using quartiles of African ancestry to group individuals (43). We examined other categorization schemes, and our results were similar when using deciles of the AF-BGA distribution (Web Table 1), suggesting that any form of categorization based on BGA yields generally similar results.

A few African Americans had a very low AF-BGA and appeared to be outliers. As we show in the Web Appendix and Web Figure 1, the large numbers of AIMs used in this

Table 3.   Comparison Between Self-Identified Race- and Biogeographic Ancestry-based Categories in Predicting Levels of Continuously Measured Cardiovascular Disease Risk Factors in the Total Sample, Heart SCORE Study, July 2009–April 2010[a]

	Self-Identified Race				Biogeographic Ancestry			
	β	95% CI	t Statistic	P Value[b]	β	95% CI	t Statistic	P Value
Body mass index	3.3	2.6, 4.0	9.2	<0.0005	1.1	0.8, 1.3	8.4	<0.0005
Body density[c]	−0.007	−0.005, −0.007	−6	<0.0005	−0.002	−0.003, −0.001	−5	<0.0005
Total cholesterol, mg/dL	−3.5	−8.9, −1.9	−1.3	0.22	−1.4	−3.2, 0.5	−1.5	0.16
LDL cholesterol, mg/dL	−4.0	−8.6, −0.5	−1.7	0.1	−0.9	−2.5, −0.7	−1.1	0.28
VLDL cholesterol, mg/dL	−4.9	−6.2, −3.7	−7.7	<0.0005	−1.5	−1.9, −1.1	−6.8	<0.0005
Lipoprotein A, mg/dL	1.8	1.1, 2.4	5.5	<0.0005	0.6	0.3, 0.7	4.8	<0.0005
Systolic blood pressure, mm Hg	4.8	1.2, 5.6	4.1	<0.0005	1.6	0.4, 1.9	3.8	<0.0005
Diastolic blood pressure, mm Hg	2.5	1, 3.6	3.8	<0.0005	0.9	0.4, 1.3	3.9	<0.0005

Abbreviations: CI, confidence interval; Heart SCORE, Heart Strategies Concentrating on Risk Evaluation; LDL, low density lipoprotein; VLDL, very low density lipoprotein.

[a] Linear regression models were used to test for the association of race or biogeographic ancestry with each outcome variable. Age, smoking status, current physical activity level, education, and income were included as covariates in all models. For definitions of variables, see Tables 1 and 2.

[b] False discovery rate-adjusted, 2-sided P value. Significance was set at $P < 0.05$.

[c] Body density reflects the proportion of body fat present, with higher density indicating lower fat mass.

study provided highly precise BGA estimates and were thus used as reliable measures of genetic ancestry. We repeated our analysis after excluding these persons and observed similar coherence between BGA and SIR. Self-identified African Americans who did not differ from the remaining participants in terms of key demographic and clinical variables provided no a priori evidence for exclusion from the analysis. AF-BGA was associated with BMI and diastolic blood pressure in African Americans in the total sample and with BMI alone after exclusion of persons with low AF-BGA. These results confirm previous observations

(14, 15, 26–29, 45) but differ from other reports by failing to detect any association of BGA with CAC (27), lipids (29), or blood glucose (14). BGA explains approximately 1% of the variation in BMI and in diastolic blood pressure after demographic and clinical covariates are controlled for, indicating that some loci that are indicators of ancestry may also have phenotypic effects or be in strong linkage disequilibrium with disease-causing loci.

Inconsistency with previous results could be partly attributable to the use of few (<45) AIMs in previous studies (13–15, 18, 28, 29), which could have been subject to

Table 4.   Comparison Between Self-Identified Race- and Biogeographic Ancestry-based Categories in Predicting the Risk of Metabolic Syndrome and Its Components in Males and Females, Heart SCORE Study, July 2009–April 2010[a]

	Males						Females					
	Self-Identified Race			Biogeographic Ancestry			Self-Identified Race			Biogeographic Ancestry		
	OR[b]	95% CI	P Value[c]	OR	95% CI	P Value	OR	95% CI	P Value	OR	95% CI	P Value
Metabolic syndrome[d]	NS	NS	NS	NS	NS	0.001	NS	NS	<0.0001	NS	NS	0.005
Glucose	NS	NS	NS	NS	NS	NS	NS	NS		NS	NS	
Triglycerides	0.4	0.2, 0.6	0.004	NS	NS	NS	0.4	0.3, 0.6	<0.0005	0.7	0.6, 0.8	<0.0005
Hypertension	2.1	1.1, 3.8	0.02	1.3	1.0, 1.3	0.04	NS	NS	NS	NS	NS	NS
Coronary artery calcification	0.3	0.1, 0.8	0.03	NS	NS	NS	NS	NS	NS	NS	NS	NS

Abbreviations: CI, confidence interval; Heart SCORE, Heart Strategies Concentrating on Risk Evaluation; NS, not significant; OR, odds ratio.

[a] For definitions of variables, see Tables 1 and 2.

[b] Odds ratio for the presence of metabolic syndrome criteria.

[c] False discovery rate-adjusted, 2-sided P value. Significance was set at $P < 0.05$.

[d] Metabolic syndrome was defined according to National Cholesterol Education Program criteria; subjects were classified as "normal" (reference group), "with metabolic syndrome" (group 1), or "with a history of diabetes" (group 2) and compared using analysis of variance (males—self-identified race: NS; biogeographic ancestry: $F = 4.9$ (4 df); females—self-identified race: $F = 1.4$ (1 df); biogeographic ancestry: $F = 3.7$ (4 df)), with age, gender, smoking status, current physical activity level, education, and income as covariates. Individual components of the metabolic syndrome which did not use gender-specific cutoffs were evaluated by means of logistic regression, with age, smoking status, current physical activity level, education, and income as covariates in all models.

Table 5. Comparison Between Self-Identified Race- and Biogeographic Ancestry-based Categories in Predicting Levels of Continuously Measured Cardiovascular Disease Risk Factors in Males and Females, Heart SCORE Study, July 2009–April 2010[a]

	Males						Females					
	Self-Identified Race			Biogeographic Ancestry			Self-Identified Race			Biogeographic Ancestry		
	β	t Statistic	P Value[b]	β	t Statistic	P Value	β	t Statistic	P Value	β	t Statistic	P Value
Body mass index	1.8	3.2	0.002	0.5	2.5	0.02	3.9	8.5	<0.0005	1.2	7.9	<0.0005
Body density[c]	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Total cholesterol, mg/dL	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
LDL cholesterol, mg/dL	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
VLDL cholesterol, mg/dL	−3.9	−3.6	<0.0005	−1.1	−2.3	0.009	−5.4	−6.6	<0.0005	−1.7	−6.1	<0.0005
Lipoprotein A, mg/dL	NS	NS	NS	NS	NS	NS	2.6	6.1	<0.0005	0.8	5.2	<0.0005
Systolic blood pressure, mm Hg	8.5	4.3	<0.0005	3.1	4.4	<0.0005	3.4	2.3	0.03	NS	NS	NS
Diastolic blood pressure, mm Hg	4.8	4.1	<0.0005	1.7	3.9	<0.0005	NS	NS	NS	0.6	2.1	0.04

Abbreviations: Heart SCORE, Heart Strategies Concentrating on Risk Evaluation; LDL, low density lipoprotein; NS, not significant; VLDL, very low density lipoprotein.

[a] For definitions of variables, see Tables 1 and 2.

[b] False discovery rate-adjusted, 2-sided P value. Significance was set at $P < 0.05$.

[c] Body density reflects the proportion of body fat present, with higher density indicating lower fat mass.

BGA-measurement bias. Simulation studies have shown that using fewer markers leads to imprecise estimates with large confidence intervals and that more precise ancestry estimates are obtained with more markers (46, 47), although the exact number of AIMs required has not been established. Imprecise admixture estimates may in turn yield inconsistent or erroneous genotype-phenotype associations. For instance, an association between EU-BGA ascertained using 40 AIMs and bone mineral density (48) was negated when 170 AIMs were used to infer BGA (17). The large number of AIMs (>1,500) used in this study lends significant confidence to the precision of BGA estimates ascertained here and is a major strength of this study over previous reports. Our simulation studies with different numbers of AIMs further showed that BGA estimates also vary as a function of the specific AIM panel used. In addition, although many phenotypes were not associated with BGA, they still showed significant race-related differences, with and without adjustment for other covariates (Tables 1 and 2), and are thus prime candidates for admixture mapping in African Americans (49–51).

Model-based studies, such as the current one, have one disadvantage in that they forcibly partition an individual's genome into components that are based on certain assumptions about ancestry. Consequently, the 2-population ancestral model that we used may not be optimum for some persons included in this sample. For practical reasons related to sample access and genotyping costs, very few investigators have so far reported AIM allele frequencies in diverse European and African populations (13, 18, 48). The AIMs used here were part of an array (36) and were chosen from 2 previous reports (23, 24), neither of which documented allele frequencies in diverse ancestral European and African populations. It is possible that by using a single African ancestral group, we limited the spectrum of

ancestral African variation that contributed to the modern African-American gene pool (52, 53). Although the same argument is true for European ancestors, the significant allelic variation within Africa as compared with Europe (51, 53) probably exerts a stronger influence on BGA estimates in African Americans. Finally, including Native American ancestral allele frequencies, as reported previously (16, 17, 26), might have yielded a more complete representation of the genetic diversity in these cohorts. The approaches used for selecting SNP AIMs and inherent problems within those schemes are beyond the scope of this paper. However, previous reports (23, 24, 36) and our analysis of these AIMs in the HapMap data showed that these markers are able to distinguish between European and African ancestry with high efficiency, and are thus able to provide a measure of differentiation that is useful for continental BGA studies. It is possible that a model-free approach (e.g., principal components analysis) may provide alternative insights into the genetic structure of the cohort. Another limitation of this study is that SIR- and BGA-based models examined were not nested and thus could not be compared directly using statistical tests. We have thus reported odds ratios and β coefficients with confidence intervals and P values to facilitate a qualitative comparison of models.

In conclusion, the results of this study indicate that SIR is an acceptable surrogate for classifying individuals for nongenetic cardiovascular epidemiology studies. However, genetic cardiovascular epidemiology studies, especially those that focus on cohorts of recently admixed groups, may benefit from the application of admixture-based methods as demonstrated in previous studies (14, 15, 18). Since admixture profiles vary between populations, additional replication of these analyses in different and diverse cohorts is warranted prior to generalizing these results.

REFERENCES

1. Arnett DK, Tyroler HA, Burke G, et al. Hypertension and subclinical carotid artery atherosclerosis in blacks and whites. The Atherosclerosis Risk in Communities Study. *Arch Intern Med*. 1996;156(17):1983–1989.
2. Hutchinson RG, Watson RL, Davis CE, et al. Racial differences in risk factors for atherosclerosis. The ARIC Study. *Angiology*. 1997;48(4):279–290.
3. Harris MM, Stevens J, Thomas N, et al. Associations of fat distribution and obesity with hypertension in a bi-ethnic population. The ARIC Study. *Obes Res*. 2000;8(7):516–524.
4. Schreiner PJ, Heiss G, Tyroler HA, et al. Race and gender differences in the association of Lp(a) with carotid artery wall thickness. The Atherosclerosis Risk in Communities (ARIC) Study. *Arterioscler Thromb Vasc Biol*. 1996;16(3):471–478.
5. Metcalf PA, Sharrett AR, Folsom AR, et al. African American-white differences in lipids, lipoproteins, and apolipoproteins, by educational attainment, among middle-aged adults: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 1998;148(8):750–760.
6. Carlson CS, Aldred SF, Lee PK, et al. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am J Hum Genet*. 2005;77(1):64–77.
7. Lange LA, Carlson CS, Hindorff LA, et al. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA*. 2006;296(22):2703–2711.
8. Carroll JF, Fulda KG, Chiapa AL, et al. Impact of race/ethnicity on the relationship between visceral fat and inflammatory biomarkers. *Obesity (Silver Spring)*. 2009;17(7):1420–1427.
9. Brown SA, Hutchinson R, Morrisett J, et al. Plasma lipid, lipoprotein cholesterol, and apoprotein distributions in selected US communities. The Atherosclerosis Risk in Communities (ARIC) Study. *Arterioscler Thromb*. 1993;13(8):1139–1158.
10. Després JP, Couillard C, Gagnon J, et al. Race, visceral adipose tissue, plasma lipids, and lipoprotein lipase activity in men and women: the Health, Risk Factors, Exercise Training, and Genetics (HERITAGE) family study. *Arterioscler Thromb Vasc Biol*. 2000;20(8):1932–1938.
11. Morrison JA, Khoury P, Mellies M, et al. Lipid and lipoprotein distributions in black adults. The Cincinnati Lipid Research Clinic's Princeton School Study. *JAMA*. 1981;245(9):939–942.
12. Yaeger R, Avila-Bront A, Abdul K, et al. Comparing genetic ancestry and self-described race in African Americans born in the United States and in Africa. *Cancer Epidemiol Biomarkers Prev*. 2008;17(6):1329–1338.
13. Parra EJ, Marcini A, Akey J, et al. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet*. 1998;63(6):1839–1851.
14. Reiner AP, Ziv E, Lind DL, et al. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet*. 2005;76(3):463–477.
15. Fernández JR, Shriver MD, Beasley TM, et al. Association of African genetic admixture with resting metabolic rate and obesity among women. *Obes Res*. 2003;11(7):904–911.
16. Halder I, Shriver M, Thomas M, et al. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat*. 2008;29(5):648–658.
17. Halder I, Yang BZ, Kranzler HR, et al. Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mutat*. 2009;30(9):1299–1309.
18. Shriver MD, Parra EJ, Dios S, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet*. 2003;112(4):387–399.
19. Chakraborty R, Kamboh MI, Nwankwo M, et al. Caucasian genes in American blacks: new data. *Am J Hum Genet*. 1992;50(1):145–155.
20. Chakraborty R, Weiss KM. Frequencies of complex diseases in hybrid populations. *Am J Phys Anthropol*. 1986;70(4):489–503.
21. Darvasi A, Shifman S. The beauty of admixture. *Nat Genet*. 2005;37(2):118–119.
22. Parra EJ, Kittles RA, Argyropoulos G, et al. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol*. 2001;114(1):18–29.
23. Paschou P, Drineas P, Lewis J, et al. Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genet*. 2008;4(7):e1000114. (doi:10.1371/journal.pgen.1000114).
24. Price AL, Butler J, Patterson N, et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet*. 2008;4(1):e236. (doi:10.1371/journal.pgen.0030236).
25. Casazza K, Phadke RP, Fernandez JR, et al. Obesity attenuates the contribution of African admixture to the insulin secretory profile in peripubertal children: a longitudinal analysis. *Obesity (Silver Spring)*. 2009;17(7):1318–1325.
26. Casazza K, Willig AL, Gower BA, et al. The role of European genetic admixture in the etiology of the insulin resistance syndrome in children: are the effects mediated by fat accumulation? *J Pediatr*. 2010;157(1):50.e1–56.e1.
27. Wassel CL, Pankow JS, Peralta CA, et al. Genetic ancestry is associated with subclinical cardiovascular disease in African-Americans and Hispanics from the Multi-Ethnic Study of Atherosclerosis. *Circ Cardiovasc Genet*. 2009;2(6):629–636.
28. Wassel Fyr CL, Kanaya AM, Cummings SR, et al. Genetic admixture, adipocytokines, and adiposity in Black Americans: the Health, Aging, and Body Composition Study. *Hum Genet*. 2007;121(5):615–624.

29. Reiner AP, Carlson CS, Ziv E, et al. Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African-American participants in the CARDIA Study. *Hum Genet*. 2007;121(5):565–575.

30. Aiyer AN, Kip KE, Mulukutla SR, et al. Predictors of significant short-term increases in blood pressure in a community-based population. *Am J Med*. 2007;120(11):960–967.

31. Aiyer AN, Kip KE, Marroquin OC, et al. Racial differences in coronary artery calcification are not attributed to differences in lipoprotein particle sizes: the Heart Strategies Concentrating on Risk Evaluation (Heart SCORE) Study. *Am Heart J*. 2007;153(2):328–334.

32. Ainsworth BE, Jacobs DR Jr, Leon AS. Validity and reliability of self-reported physical activity status: the Lipid Research Clinics questionnaire. *Med Sci Sports Exerc*. 1993; 25(1):92–98.

33. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3):385–401.

34. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA*. 2001;285(19):2486–2497.

35. Chung BH, Wilkinson T, Geer JC, et al. Preparative and quantitative isolation of plasma lipoproteins: rapid, single discontinuous density gradient ultracentrifugation in a vertical rotor. *J Lipid Res*. 1980;21(3):284–291.

36. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. *Br J Nutr*. 1978;40(3):497–504.

37. Jackson AS, Pollock ML, Ward A. Generalized equations for predicting body density of women. *Med Sci Sports Exerc*. 1980,12(3):175–181.

38. Agatston AS, Janowitz WR, Hildner FJ, et al. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol*. 1990;15(4):827–832.

39. Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50 K SNP array for large-scale genomic association studies. *PLoS One*. 2008;3(10):e3583. (doi:10.1371/journal.pone.0003583).

40. Rousset F, Raymond M. Testing heterozygote excess and deficiency. *Genetics*. 1995;140(4):1413–1419.

41. Hanis CL, Chakraborty R, Ferrell RE, et al. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. *Am J Phys Anthropol*. 1986;70(4):433–441.

42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57(1):289–300.

43. Peralta CA, Risch N, Lin F, et al. The association of African ancestry and elevated creatinine in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Am J Nephrol*. 2010;31(3):202–208.

44. Casazza K, Thomas O, Dulin-Keita A, et al. Adiposity and genetic admixture, but not race/ethnicity, influence bone mineral content in peripubertal children. *J Bone Miner Metab*. 2010;28(4):424–432.

45. Tang H, Jorgenson E, Gadde M, et al. Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. *Hum Genet*. 2006;119(6): 624–633.

46. Pfaff CL, Parra EJ, Bonilla C, et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet*. 2001;68(1):198–207.

47. Tang H, Peng J, Wang P, et al. Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol*. 2005;28(4):289–301.

48. Bonilla C, Shriver MD, Parra EJ, et al. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York City. *Hum Genet*. 2004;115(1):57–68.

49. McKeigue PM. Multipoint admixture mapping. *Genet Epidemiol*. 2000;19(4):464–467.

50. McKeigue PM, Carpenter JR, Parra EJ, et al. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet*. 2000;64(2): 171–186.

51. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet*. 2004;74(5):979–1000.

52. Campbell CD, Ogburn EL, Lunetta KL, et al. Demonstrating stratification in a European American population. *Nat Genet*. 2005;37(8):868–872.

53. Cavalli-Sforza LL, Piazza A. Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur J Hum Genet*. 1993;1(1):3–18.