Admixture Mapping of Obesity-related Traits in African Americans: The Atherosclerosis Risk in Communities (ARIC) Study*

Ching-Yu Cheng[1–3], David Reich[4,5], Josef Coresh[1,6], Eric Boerwinkle[7], Nick Patterson[5], Man Li[1], Kari E. North[8], Arti Tandon[5], Joan E. Bailey-Wilson[2], James G. Wilson[9] and W.H. Linda Kao[1]

Obesity is an important cause of morbidity and mortality worldwide. In the United States, the prevalence of obesity is higher in African Americans than whites, even after adjustment for socioeconomic status (SES). This leads to the hypothesis that differences in genetic background may contribute to racial/ethnic differences in obesity-related traits. We tested this hypothesis by conducting a genome-wide admixture mapping scan using 1,350 ancestry-informative single-nucleotide polymorphisms (SNPs) in 3,531 self-identified blacks from the Atherosclerosis Risk in Communities (ARIC) study. We used these markers to estimate the overall proportions of European ancestry (PEAs) for each individual and then scanned for the association between PEA and obesity-related traits (both continuous and dichotomous) at each locus. The median (interquartile range) PEA was 0.151 (0.115). PEA was inversely correlated with continuous BMI, weight, and subscapular skinfold thickness, even after adjusting for socioeconomic factors. In contrast, PEA was positively correlated with BMI-adjusted waist circumference. Using admixture mapping on dichotomized traits, we identified a locus on 2p23.3 to be suggestively associated with BMI (locus-specific lod = 4.11) and weight (locus-specific lod = 4.07). After adjusting for global PEA, each additional copy of a European ancestral allele at the 2p23.3 peak was associated with a BMI decrease of ~0.92 kg/m^2 ($P = 2.9 \times 10^{-5}$). Further mapping in this region on chromosome 2 may be able to uncover causative variants underlying obesity, which may offer insights into the control of energy homeostasis.

Obesity (2010) **18,** 563–572. doi:10.1038/oby.2009.282

INTRODUCTION

Obesity is an important cause of morbidity and mortality worldwide, increasing the risk of hypertension, diabetes, and cardiovascular diseases. Environmental factors such as socioeconomic status (SES), physical inactivity, and excess caloric consumption can affect an individual's risk of obesity (1). Although environmental factors are important, there is considerable evidence that genetic factors also play a significant role in the pathogenesis of obesity. Family studies have shown that heritability of BMI ranges from 25% in adoption studies to 70% in twin studies, with an overall estimate of ~40% (2,3).

Although obesity, as defined by a BMI ≥30 kg/m^2, is highly prevalent in the United States, its prevalence varies among racial/ethnic groups. Based on the 2003–04 National Health and Nutrition Examination Study, African Americans are ~1.5 times more likely to be obese than their white counterparts (45.0% of African Americans and 30.6% of white adults were obese) (4). Studies have indicated that even in homogeneous SES groups, the prevalence of obesity is still higher in African Americans (4,5). The disproportionate level of obesity in African Americans probably cannot be fully explained by SES or environmental factors alone, which prompted our hypothesis that the differences in genetic background may partially account for differences in obesity risk across racial/ethnic populations.

Admixture mapping is an efficient method to scan the genome in recently admixed populations, such as African Americans, for genomic regions which may harbor variants that not only

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [2]Inherited Disease Research Branch, National Human Genome Research Institute, Baltimore, Maryland, USA; [3]Department of Opthalmology, National Yang Ming University School of Medicine and Taipei Veterans General Hospital, Taipei, Taiwan; [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; [5]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Boston, Massachusetts, USA; [6]Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, Maryland, USA; [7]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA; [8]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; [9]The V.A. Medical Center and the University of Mississippi Medical Center, Jackson, Mississippi, USA. Correspondence: W.H. Linda Kao (wkao@jhsph.edu)

*This article was originally published as Integrative Physiology.

Received 21 September 2008; accepted 17 July 2009; published online 20 August 2009. doi:10.1038/oby.2009.282

differ in frequency but can also partially explain differences in phenotypes between genetically diverse populations, such as Europeans and Africans (6–11). In African Americans (~6 generations since first mixture, on average), genomic segments with contiguous European or African ancestry have not had much time to break up by recombination (9–11). Therefore admixture mapping only requires genotyping genetic markers every couple of million base pairs, which is 100–1,000 times less density than is required by genome-wide linkage disequilibrium association studies (12).

Since the identification of appropriate markers and the development of analytical methods (10) appropriate for admixture mapping studies in African Americans, studies have reported significant associations with genetic loci for multiple sclerosis (13), prostate cancer (14), hypertension (15), inflammatory markers (16), white blood cell count (17), and end-stage renal disease (18). Although the regions localized by admixture mapping are generally broad, a recent study has been successful in precisely identifying genetic variants that affect the circulating levels of interleukin-6 soluble receptor and interleukin-6 (16), and another study succeeded in following-up an initial broad admixture mapping by a fine-mapping study that found the specific variants responsible for risk of prostate cancer (19).

Although previous studies have shown that degree of African admixture is positively correlated with BMI in African Americans (20,21), suggesting that genetic admixture may have a modest to strong effect on obesity susceptibility, there has not been a comprehensive genome-wide scan to search for regions that may account for this observed association. In the present study, we carried out a genome-wide admixture mapping scan using 1,350 ancestry-informative markers in 3,531 self-identified blacks from the Atherosclerosis Risk in Communities (ARIC) study.

METHODS AND PROCEDURES

Study populations

Subjects of the present study were from the 4,266 African-American participants of the ARIC study. The ARIC study is a prospective epidemiologic study that examines clinical and subclinical atherosclerotic disease in a cohort of 15,792 persons, aged 45–64 years at their baseline examination. Participants were selected by probability sampling from four US communities: Forsyth County, NC (12% African American); Jackson, MS (100% African American); the northwest suburbs of Minneapolis, MN (<1% African American); and Washington County, MD (<1% African American). The sampling procedure and methods used in ARIC have been described in detail elsewhere (22). The baseline examination took place from 1987 to 1989. Data from the baseline examination were used for the present analysis. Classification of ethnicity was based on self-report. All procedures were conducted according to institutionally approved protocols for study of human subjects and written informed consent was obtained from all subjects.

The final sample size for the present study was 3,531 African Americans after applying the following exclusion criteria: (i) African-American subjects who lived in Minneapolis, MN, or Washington County, MD ($n = 55$), (ii) did not consent to genetic studies or did not have DNA samples available ($n = 255$), and (iii) samples that were not genotyped successfully or that failed to pass quality control ($n = 425$, see section "Elimination of poorly performing and problematic samples").

Obesity-related phenotypes

The ARIC study has extensive anthropometric data as well as behavioral and environmental data relevant to obesity. Anthropometric measurements were performed with the participants in fasting state with an empty bladder and wearing light-weight, nonconstricting underwear and no shoes. Height was measured with a wall-mounted ruler, and weight was measured using a balance scale that was zeroed daily. BMI was calculated as weight (in kg)/height (in meters) squared. Waist and hip circumferences were measured using an anthropometric tape applied at the level of umbilicus and at the maximal protrusion of the gluteal muscles, respectively. The reliability of circumference measurements was high (intraobserver reliability of coefficient, $R > 0.91$) (23). Subscapular skinfold thickness and triceps skinfold thickness were measured using a Lange caliper on standardized right-side location. The reliability of intertechnician measurements of triceps skinfolds and subscapular skinfolds was high ($R > 0.91$) (23).

An additional phenotype was "waist circumference adjusted for BMI," which was obtained by using the residual from a linear regression of waist circumference as the dependent variable and BMI as independent. Waist circumference adjusted for BMI has been shown highly correlated with abdominal fat and was used as a surrogate for visceral fat (24).

SES and other covariates

Measurements of other potential factors relevant to obesity are described as follows. Information on SES, including personal education, occupation, and family income, was collected during the baseline interview. For the purpose of this analysis, education level was categorized as three groups: high school not completed; high school graduate or vocational school completed; and some college, college completed or some graduate or professional school. Occupations were assigned according to the criteria of the 1980 US Census and categorized as six groups: managerial and professional specialty; technical, sales and administrative support; service; farming, forestry and fishing or precision production, craft and repair; operators, fabricators, and laborers; and homemakers (25). Total combined family income was categorized as <$5,000–7,999, $8,000–24,999, and ≥$25,000. Approximately 6% of study participants did not provide their income information and thus they are coded as a separate category.

Physical activity during leisure time was assessed by a modified version of the questionnaire developed by Baecke *et al.* (26). Diet was characterized using a modified version of the 61-item food frequency questionnaire developed by Willett *et al.* (27), and the dietary fat was summarized with the use of the Keys score (28).

Diabetes was defined as the presence of any one of the following at the baseline examination: (i) fasting glucose ≥7.0 mmol/l (126 mg/dl), (ii) nonfasting glucose ≥11.1 mmol/l (200 mg/dl), (iii) current use of diabetic medication, or (iv) a positive response to the question "Has a doctor ever told you that you had diabetes (sugar in the blood)?"

Genotyping for the admixture SNP panel

A total of 1,536 single-nucleotide polymorphisms (SNPs) were included in our admixture panel. This panel was constructed by using the panel of ancestry-informative markers previously published by Smith *et al.* (11), and then improving this panel by mining new ancestry-informative markers from the data sets of Hinds *et al.* (29) and the phase 2 International Haplotype Map (30), and validating them to confirm that they were indeed ancestry informative. These SNPs were then prioritized as most informative about West African vs. European ancestry, according to their predicted usefulness for determining ancestry (**Supplementary Table S1** online). Genotyping was performed by the Center for Inherited Disease Research (Johns Hopkins University, Baltimore), using the Illumina BeadLab platform (31). The ARIC study has a rigorous quality control program, including blind duplicates. Many genotypes in duplicates were obtained using the Illumina BeadLab technologies in ARIC African-American participants, and CEPH and Yoruban samples. The mismatch rate among 218,461 duplicate genotypes was 0.1%.

Frequency estimates from ancestral populations

To estimate frequency for each of the SNPs in Africans and Europeans, we used genotype data from 175 West African and 259 European samples (11,14,32). These samples provided a Bayesian prior distribution for the reference and variant alleles in the two ancestral populations.

Elimination of poorly performing and problematic samples

After genotyping, samples were eliminated based on the following criteria: (i) samples with low (<94%) call rate (n = 372), (ii) samples showing gender mismatch between self-reported data and genetically estimated gender based on 50 markers on the X chromosome (n = 20), and (iii) duplicate samples (defined as >75% match in the genotypes between two samples; n = 14).

We used built-in data checking programs in the ANCESTRYMAP (10) software to exclude samples with an apparent excess or deficiency of heterozygous genotypes compared with the expectation from the individuals' global European ancestry. An apparent excess of heterozygous genotypes (defined as the Z-score >10) usually indicates the individuals have parents with divergent ancestries (e.g., one parent who is entirely of European ancestry) and such individuals nearly always have estimated European ancestry close to 0.5 (10). We removed such individuals (n = 15) from our study because they contradict the assumptions of our admixture mapping software. We did not identify any individuals with an apparent deficiency of heterozygous genotypes (defined as the Z-score less than −6) (10). Individuals who appeared to be genetic outlier (i.e., an estimated proportion of European ancestry (PEA) >0.85, see section "Estimating genetic ancestry") were further excluded (n = 4) from the subsequent admixture mapping scan.

SNP quality control

SNPs were dropped if there were atypical clustering patterns (n = 46) or ill-defined clusters (n = 73). An additional SNP was dropped because of relatively low genotyping success rate (95%). This left us with 1,416 SNPs (all with genotyping call rate >97%). We then used a series of criteria as described previously to further eliminate SNPs from the analysis (13). First, we eliminated SNPs (n = 15) if they did not meet the requirement for Hardy–Weinberg equilibrium ($P > 0.01$) in both ancestral West African and European populations. We then examined the SNP frequency and confirmed that the frequencies of all SNPs in the African-American participants were appropriately intermediate between the frequencies in the West African and European ancestral populations (10). Lastly, we eliminated SNPs (n = 51) that were in linkage disequilibrium with each other in the two ancestral populations, as this can produce false positive signals of association (10). After imposing these requirements, 1,350 SNPs were left for analysis.

Estimating genetic ancestry

For each individual, we estimated a global ancestry, as indicated by PEA using the ANCESTRYMAP software (10). To account for uncertainty in the unknown parameters (i.e., average proportion of alleles inherited from population and number of generations since admixture) that emerge from the hidden Markov model analysis, ANCESTRYMAP uses a Markov chain Monte Carlo approach. All Markov chain Monte Carlo runs used 100 burn-in and 200 follow-on iterations, as recommended (10). We also carried out principal component analysis on the genotypes using the EIGENSTRAT (33,34) software to infer continuous axes of genetic variation. The axes of variation are defined as the top eigenvectors of a covariance matrix among individuals.

Assessing correlations between global ancestry estimate (PEA) and obesity-related phenotypes

Statistical analyses were carried out using Stata 9.2 (Stata, College Station, TX). Obesity-related quantitative traits were initially adjusted for age, sex, and study sites using multivariate linear regression models. In separate models, the traits were also adjusted for SES, including education level, occupation, and income. Residuals of the traits from the regression models were then obtained for each individual. The correlations between the residuals and PEA were calculated using Spearman's rank correlation coefficients (r_s). Additionally, we assessed the correlations only in individuals without diabetes, to avoid problems related to comorbidity and treatment. Other covariates were also examined in the analyses, including the dietary Keys score (35) and physical activity during leisure time, which are known to be associated with obesity.

Admixture mapping scan

We used ANCESTRYMAP (10) to search for association with genomic regions that have an increased proportion of either European or African ancestry. For the purpose of this analysis, study participants were ranked by the residuals estimated from the linear regression models with obesity traits as dependant variables, and age, sex, study site, and SES as independent variables. The top 30% of participants with the highest residuals were defined as cases and the bottom 30% as controls for each trait. Because ANCESTRYMAP uses Bayesian statistics, a prior distribution of risk models is required (10). We tested 12 prespecified risk models for each trait to assess overall evidence of association by averaging all models. The first six models used 0.5-, 0.66-, 0.8-, 1.2-, 1.5-, and 2.0-fold increased risk due to inheritance of one copy of a European ancestral allele for cases, with a control risk of 1. The next six models used the same risk set as the first for cases, but the control risks were set to the reciprocal of the case risks. This set of models reflects the hypothesis that European ancestral alleles are less likely to confer risks and also tests for the alternative possibilities (10).

Two statistics are produced by the ANCESTRYMAP software. First, it calculates a "locus–genome statistic" in cases only by comparing the likelihood of any locus being a disease locus (average PEA at the locus) vs. it being not related to disease (average PEA across the genome) (10). A locus-specific lod ($\log_{10}$ of this likelihood) score >4 is considered as suggestive significant and >5 as significant. To account for multiple hypothesis testing, a genome-wide lod score, which assesses whether there is a risk locus anywhere in the genome, was calculated by averaging the locus-specific lod scores across all loci in the genome. We interpret a genome-wide lod score >1 as suggestive significant and >2 as significant at a genome-wide level. Second, ANCESTRYMAP produces a "case–control statistic," which compares mean estimates of PEA in cases vs. controls at every locus in the genome, and ensures that any deviation in ancestry from the genome-wide average is seen only in cases, but not in controls (10). Under null hypothesis, this statistic is distributed as a standard normal distribution. To account for multiple comparisons in the case–control analysis, we presented Bonferroni-adjusted P values that had been corrected for 1,000 hypotheses (markers) tested. Both the locus–genome and the case–control statistics assess the association between marker and phenotype that is above and beyond the association between overall ancestry and phenotype (10).

Testing for associations of the local ancestry at the admixture signal to BMI

After the initial admixture scans which were based on a dichotomous phenotype (i.e., cases and controls), we included all samples in an analysis of continuous traits to check whether the results were consistent. To do this, we used the ANCESTRYMAP software to obtain local estimates of ancestry at the site of the admixture signal (10). To determine whether there is evidence of residual association with the local estimates of ancestry after adjustment for global ancestry (PEA), we performed linear regression analysis. First, we applied a normal-quantile transformation for BMI (the trait showing evidence of association, see "Results" section). We ranked the BMI values and obtained the percentile of each value. The percentile was matched to a standard normal distribution with corresponding Z-score, which was then used for the regression analysis. Next, we determined the association and its significance by regressing local estimates of ancestry on the transformed BMI. Finally, we assessed whether the local ancestry

estimated at the admixture peak was associated with the transformed BMI even after accounting for global ancestry. This approach enabled us to include all samples into an analysis of the continuous BMI phenotype, instead of just using defined case and control samples according to the highest and lowest levels.

RESULTS

Basic demographic and phenotypic characteristics by study site for the 3,531 subjects included in the analysis are summarized in Table 1. ARIC participants from Jackson had higher BMI, were less likely to have attended college and reported lower incomes, physical activity and dietary fat intake than those from Forsyth County (all $P < 0.05$). The eliminated individuals ($n = 735$, see section "Study populations") did not significantly (all $P > 0.05$) different from those individuals included in the analysis in BMI, weight, waist circumference, hip circumference, and triceps skinfold thickness (data not shown), except that they had slightly higher waist-to-hip ratio (0.93 ± 0.07, $P = 0.002$) and thinner subscapular skinfold thickness (31.0 ± 13.4 mm, $P = 0.006$) than those included.

The distribution of estimated PEA in the African-American participants appeared right-skewed (Figure 1). Overall, the median (interquartile range) PEA was 0.151 (0.115) in all participants, 0.150 (0.111) in Jackson participants, and 0.172 (0.136) in Forsyth County participants (P for the difference between the two sites = 0.004, by Wilcoxon rank-sum test). Although the PEA was higher in participants from Forsyth County, there were highly overlapping distributions of ancestry between the two ARIC study sites based on principal component analysis for detection of substructure (Figure 2). We found a very high correlation between individuals' first eigenvector (axis of variation) and the estimated PEA ($r_s = 0.99$, $P < 0.0001$); thus the

Table 1 Baseline characteristic of the study participants by study site

	Jackson, MS ($n = 3,126$)	Forsyth County, NC ($n = 405$)	Combined ($n = 3,531$)
Age	53.4 ± 5.8	54.5 ± 5.9[a]	53.5 ± 5.8
Sex (% female)	62.6	57.8	62.1
BMI (kg/m^2) ≥30 (%)	41.5	32.9[a]	40.5
BMI (kg/m^2)	29.8 ± 6.2	28.6 ± 6.2[a]	29.6 ± 6.2
Weight (lb)	184.7 ± 38.3	176.3 ± 39.0[a]	183.8 ± 38.5
Waist circumference (cm)	99.1 ± 15.3	99.1 ± 14.8	99.1 ± 15.2
Hip circumference (cm)	108.0 ± 12.2	106.0 ± 12.0[a]	107.8 ± 12.2
Waist-to-hip ratio	0.92 ± 0.08	0.93 ± 0.07[a]	0.92 ± 0.08
Subscapular skinfold thickness (mm)	33.2 ± 14.4	27.7 ± 11.4[a]	32.5 ± 14.2
Triceps skinfold thickness (mm)	27.6 ± 12.5	25.3 ± 11.8[a]	27.3 ± 12.5
College or above (%)	30.1	35.2[a]	30.7
Family annual income >$25,000 (%)	28.0	39.5[a]	29.4
Executive, managerial, or professional occupations (%)	20.7	21.0	20.7
Index of physical activity during leisure time	2.0 ± 0.6	2.3 ± 0.5[a]	2.1 ± 0.6
Dietary Keys score	42.5 ± 8.7	41.4 ± 9.6[a]	42.4 ± 8.8
Diabetes (%)	19.9	17.0	19.5

Data are shown as mean ± s.d. for continuous variables.
[a]$P < 0.05$, between two study sites.

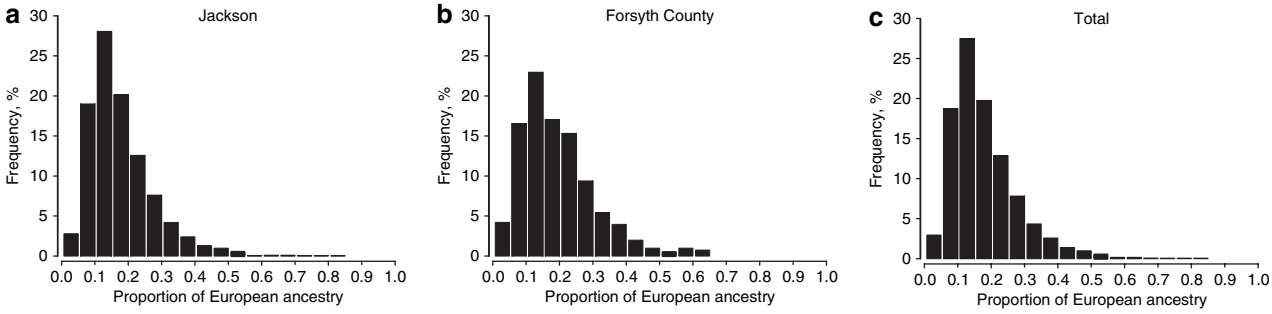

Figure 1 Histograms of estimated proportions of European ancestry (PEAs) in African Americans in the Atherosclerosis Risk in Communities study. The PEA was determined using the ANCETRYMAP software for (**a**) 3,126 individuals in Jackson, MS, (**b**) 405 in Forsyth County, NC, and in (**c**) all 3,531 subjects. Four individuals from Jackson who had an estimated PEA >0.85 (0.951, 0.949, 0.942, and 0.937, respectively) were excluded.

Table 2 Correlations between obesity-related traits and proportions of European ancestry

	Adjustment for age, sex, and study site		Adjustment for age, sex, study site, and SES[a]		Adjustment for age, sex, study site, and SES[a] in nondiabetic subjects	
Traits	r_s	*P* value	r_s	*P* value	r_s	*P* value
BMI	−0.090	8.0×10^{-8}	−0.069	4.3×10^{-5}	−0.075	8.7×10^{-5}
Weight	−0.079	2.7×10^{-6}	−0.062	2.3×10^{-4}	−0.067	2.4×10^{-4}
Waist circumference	−0.055	0.001	−0.028	0.096	−0.037	0.051
Hip circumference	−0.054	0.001	−0.044	0.009	−0.050	0.009
Waist-to-hip ratio	−0.032	0.060	0.005	0.749	0.002	0.932
Waist circumference adjusted for BMI	0.042	0.013	0.060	3.6×10^{-4}	0.054	0.005
Subscapular skinfold thickness	−0.063	1.7×10^{-4}	−0.055	0.001	−0.055	0.004
Triceps skinfold thickness	−0.036	0.035	−0.033	0.053	−0.042	0.029

Correlations between adjusted traits and proportions of European ancestry were assessed using Spearman's rank correlation coefficient (r_s).
[a]Socioeconomic status (SES), including personal education, occupation, and family income.

first axis of variation may well reflect continental origins and ancestry effects.

In the ARIC participants, we observed a significant trend ($P < 0.0001$) toward decreased PEA from 0.164 (0.128) in the nonobese individuals (defined as BMI <25), to 0.156 (0.119) in overweight individuals (BMI 25–29.9), and 0.143 (0.103) in the obese individuals (BMI ≥30). The correlations between PEA and each of the eight continuous obesity-related traits are shown in **Table 2**. When correcting for age, sex, and study site, we found that all traits, except waist-to-hip ratio, were significantly correlated with PEA. Additionally adjusting for SES weakened most of the associations, but did not abolish them all. That is, BMI ($P = 4.3 \times 10^{-5}$), weight ($P = 2.3 \times 10^{-4}$), hip circumference ($P = 0.009$), and subscapular skinfold thickness ($P = 0.001$) were still significantly, inversely correlated with PEA. In contrast, waist circumference adjusted for BMI was significantly, positively correlated with PEA ($P = 3.6 \times 10^{-4}$). The evidence of significant correlations for these five traits persisted and the strength of the correlations was similar, even after additionally correcting for other covariates, including dietary fat and physical activity during leisure time (data not shown). To minimize the potential impact of diabetes on obesity-related traits, we further examined these correlations among only nondiabetic individuals ($n = 2{,}780$) and found the strengths of correlations with BMI, weight, and hip circumference to be stronger (i.e., higher absolute value of r_s). For waist circumference adjusted for BMI, the correlation appeared to be weaker in nondiabetic individuals.

Although there were differences between the two study sites for some of the obesity-related traits, the site-specific correlations between the traits and PEA were similar for all traits except for waist-to-hip ratio, which was not significantly associated with PEA in either site (see **Supplementary Table S2** online). Therefore, in the present study, we pooled samples from two sites to increase study power for the above correlation analysis and the following admixture scans.

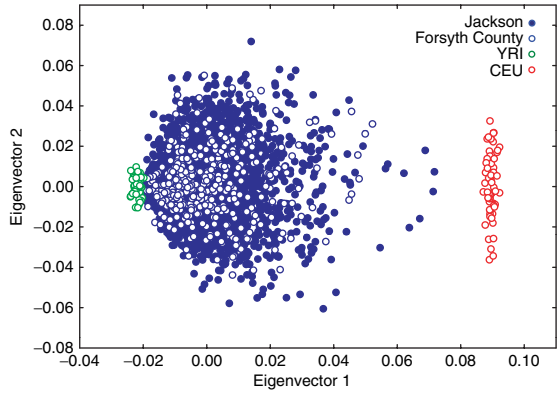

Figure 2 The top two axes of variation (eigenvector) from principal component analysis in the Atherosclerosis Risk in Communities (ARIC) African Americans and the HapMap YRI and CEU samples. The first axis of variation may reflect continental origins. This analysis found highly overlapping distributions of ancestry between the two ARIC study sites—Jackson and Forsyth County—although the PEAs were significantly different ($P = 0.004$).

Admixture mapping scans were performed on each of the eight obesity-related traits (adjusted for age, sex, study site, and SES). Using the top 30% of participants with the highest values as cases and the bottom 30% as controls, we found genome-wide suggestive significant evidence of associations with BMI (**Table 3**). The strongest association for BMI was at 27.3 Mb on chromosome 2 (2p23.3) between rs13025681 and rs7593448. The peak locus-specific lod was 4.11 (**Figure 3a**), which meets our priori defined thresholds for suggestive significance (lod = 4). Averaging 10 to the power of the lod scores across all loci in the genome and taking the $\log_{10}$ of this average produced a genome-wide association score of 1.14, again meeting our threshold of 1 for suggestive significance accounting for multiple comparisons. The most extreme case–control

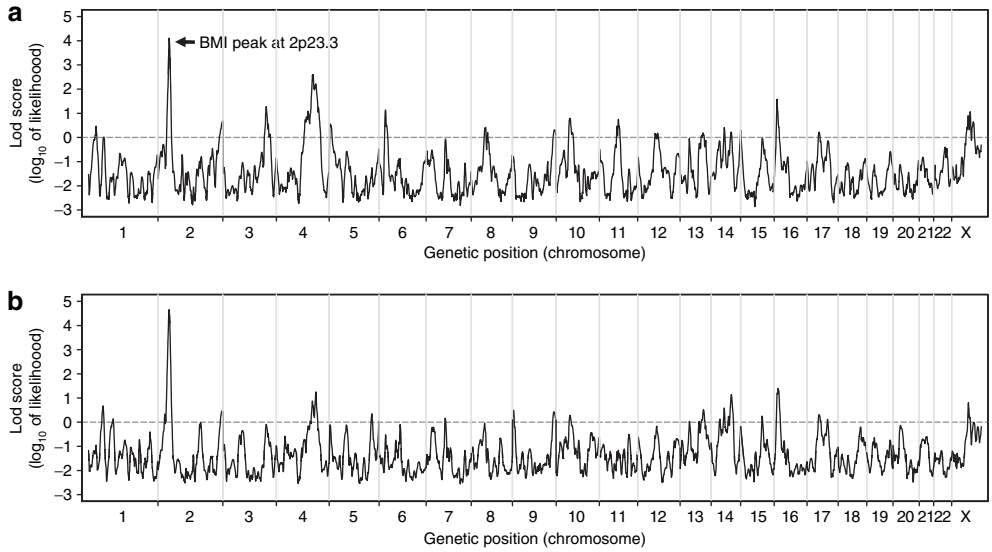

Figure 3 Genome-wide admixture mapping scans for loci underlying BMI. (**a**) A suggestive significant signal (locus-specific lod = 4.11) was detected at 2p23.3 for BMI in all subjects. Averaging the lod scores across all loci in the genome, we obtained a genome-wide lod score of 1.14, reaching the threshold of 1 for suggestiveness. We observed other peaks for BMI at 4q31.21 (locus-specific lod = 2.50) and at 16p13.3 (locus-specific lod = 1.58), both of which did not reach the thresholds (lod = 4) for suggestiveness. (**b**) The admixture association at 2p23.3 was stronger in nondiabetic subjects, with the peak locus-specific lod rising to 4.65 and genome-wide score increasing to 1.73.

Table 3 Summary of locus–genome and case–control statistics from the admixture mapping scans of BMI[a]

			Scores at chromosome 2 peak	
	No. of cases/controls	Genome-wide lod score[b]	Case-only locus-specific lod[c]	Case–control *Z*-statistic
All African-American subjects	1,053/1,054	1.14	4.11	–5.09
Nondiabetic African-American subjects only	914/915	1.73	4.65	–5.41
All African-American subjects, best-fit model of 0.79-fold risk because of European ancestry with the control risk of 1.25	1,053/1,054	2.17	5.20	–5.09
Nondiabetic African-American subjects only, best-fit model of 0.79-fold risk because of European ancestry with the control risk of 1.30	914/915	2.92	5.84	–5.41

[a]BMI was adjusted for age, sex, study site, and socioeconomic status. [b]Genome-wide lod scores >2 are formally significant; scores >1 are suggestive. [c]Locus-specific lod scores >5 are formally significant; scores >4 are suggestive.

statistic (Z-score) in the genome is −5.09 (Bonferroni-adjusted $P = 3.6 \times 10^{-4}$), and was exactly at the same location as the peak locus-specific lod of 4.11.

We also performed admixture scans on nondiabetic subjects only. The peak locus-specific lod increased from 4.11 to 4.65 (**Figure 3b**) and the genome-wide lod score increased from 1.14 to 1.73, despite the smaller sample size. The best case–control Z-score is −5.41 (Bonferroni-adjusted $P = 6.3 \times 10^{-5}$). For weight, a peak locus-specific lod of 4.07 was identified at the same location with BMI at 2p23.3, with a corresponding genome-wide lod score of 1.22 (**Table 4**), again reaching the thresholds for suggestive significance for a genome-wide analysis. The case–control statistic for weight at this peak is −5.11, corresponding to a Bonferroni-adjusted $P = 3.2 \times 10^{-4}$. We failed to find any evidence of associations with the other six obesity-related traits. However, the best locus-specific lod scores for waist circumference and hip circumference were also at 2p23.3.

We next performed an analysis to determine whether the observed association on 2p23.3 was due to our case–control definition, by examining the locus-specific lod score for each individual. As expected, we found that the evidence of association to BMI was contributed mostly by the subjects with the highest BMIs (those in the top 15–45% of BMI) (**Figure 4**). The cumulative score increases gradually after the top 15% of adjusted BMI values, reaches its maximum for the top 30%, and then drops after the top 45%. We also ranked all nondiabetic subjects according to their adjusted BMI values. Similarly, we found that the admixture association is contributed by the top 12–50% of nondiabetic subjects. The cumulative lod

Table 4 Summary of locus–genome statistic from the admixture mapping scans of obesity-related traits except BMI

Traits[a]	No. of cases/controls	Genome-wide lod score[b]	Best locus		Locus-specific lod[c] at 2p23.3
			Locus-specific lod[c]	Position	
Weight	1,053/1,054	1.22	4.07	2p23.3	4.07
Waist circumference	1,053/1,054	–0.09	2.55	2p23.3	2.55
Hip circumference	1,054/1,055	0.01	2.58	2p23.3	2.58
Waist-to-hip ratio	1,053/1,054	–0.31	2.16	5q35.3	–0.39
Waist circumference adjusted for BMI	1,053/1,053	–0.75	1.39	3p12.3	–2.10
Subscapular skinfold thickness	1,050/1,050	–0.24	1.69	16p13.3	0.84
Triceps skinfold thickness	1,053/1,054	–0.45	1.66	2q37.3	0.92

[a]Adjusted for age, sex, study site, and socioeconomic status. [b]Genome-wide lod scores >2 are formally significant; scores >1 are suggestive. [c]Locus-specific lod scores >5 are formally significant; scores >4 are suggestive.

Table 5 Linear regression models of BMI[a] on global European ancestry and local European ancestry at 2p23.3 peak

Subjects and variables in models	Model 1		Model 2		Model 3	
	Regression coefficient	*P* value	Regression coefficient	*P* value	Regression coefficient	*P* value
All African-American subjects						
Global Europe ancestry	–0.07[b]	7.3×10^{-5}			–0.03[b]	0.098
Local European ancestry			–0.19[c]	3.2×10^{-8}	–0.16[c]	2.9×10^{-5}
Nondiabetic African-American subjects						
Global Europe ancestry	–0.07[b]	2.3×10^{-4}			–0.02[b]	0.255
Local European ancestry			–0.22[c]	4.3×10^{-9}	–0.19[c]	2.5×10^{-6}

[a]BMI was normal-quantile transformed after adjustment for age, sex, study site, and socioeconomic status, including personal education, occupation, and family income. [b]Change in BMI *Z*-score units for 0.1 increase in proportion of European ancestry. [c]Change in BMI *Z*-score units for one additional copy of European ancestral allele at the 2p23.3 peak.

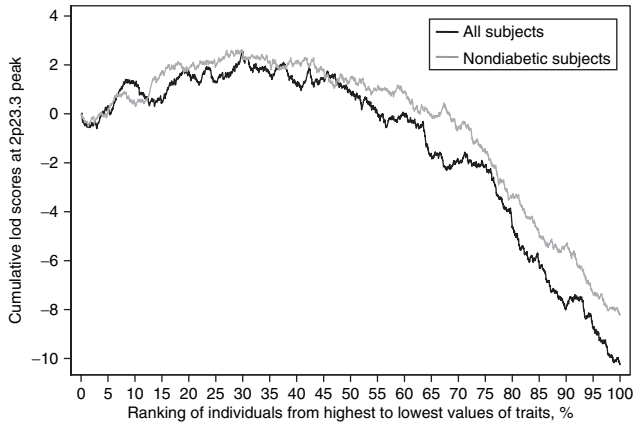

Figure 4 The observed association was contributed mostly by the top 15–45% of subjects with the highest BMI, adjusted for age, sex, study site, and socioeconomic status. The software ANCESTRYMAP was used to calculate the lod score of disease association for each individual at the peak locus on 2p23.3 under the risk model of 0.79-fold increased risk for each copy of the European allele, with a control risk of 1.25. The cumulative lod score for BMI increased markedly for the top 15–45% of the subjects with highest BMI, and then decreased. For nondiabetic subjects, the individual's lod score was calculated at the same peak under the risk model of 0.79-fold increased risk for European ancestry, with a control risk of 1.30. In nondiabetic subjects, the cumulative lod score increased dramatically for the top 12–50% with highest BMI, and then decreased.

score is generally higher in nondiabetic subjects after the top 12% (Figure 4), which suggests the admixture association is stronger when we exclude diabetic subjects.

To further examine the robustness of the admixture-generated signal on 2p23.3 and whether the signal on 2p23.3 contributes to the overall association between global PEA and BMI, we next carried out a series of linear regression analysis to assess the association between the normal-quantile transformed BMI and both global PEA and local estimates of European ancestry at the 2p23.3 peak in all individuals (Table 5). As expected, global PEA was strongly associated with transformed BMI ($P = 7.3 \times 10^{-5}$; model 1 in Table 5). Similarly, local European ancestry alone was strongly, inversely associated with transformed BMI ($P = 3.2 \times 10^{-8}$; model 2). To assess how much of the association between global PEA and transformed BMI was accounted by association with ancestry at 2p23.3, we modeled transformed BMI as a function of both global and local European ancestry (model 3). We found that the locus European ancestry at the 2p23.3 peak almost eliminated the association between BMI and individual global PEA. Conversely, after adjustment for each individual's global PEA, there was still significant evidence of residual association ($P = 2.9 \times 10^{-5}$) between transformed BMI and the local European ancestry (model 3). Each additional copy of

a European ancestral allele at the 2p23.3 peak was associated with a BMI decrease of 0.16 *Z*-score units on average (equivalent to ~0.92 kg/m^2). Similarly in nondiabetic subjects, after adjustment for global PEA, the residual association with the local European ancestry remained significant ($P = 2.5 \times 10^{-6}$) and the size effect of the local ancestry was greater than that in all subjects.

DISCUSSION

We have conducted the largest admixture mapping scan to date in African Americans, and found evidence for association to BMI at 2p23.3. Among individuals with BMI in the upper 30th percentile of the population, there was suggestive evidence of increased African ancestry on 2p23.3 compared to the rest of the genome (genome-wide lod score = 1.14; locus-specific lod = 4.11). We observed concordant results when BMI was examined as a continuous variable, with higher levels of European ancestry being significantly associated with lower levels of BMI. The significant evidence of association with the local ancestry at the 2p23.3 peak was above and beyond the contribution of global European ancestry.

In addition to BMI, we found significant inverse correlations between European ancestry and three other continuous obesity-related traits, including weight, hip circumference, and subscapular skinfold thickness after adjusting for SES. On the other hand, although waist circumference was not associated with genetic admixture after adjusting for SES, BMI-adjusted waist circumference, potentially as a marker of visceral adiposity, was significantly and positively correlated with European ancestry. Clinical studies using computer tomography to determine abdominal fat have shown that African-American adults tend to have lower visceral fat compared to whites, despite similar BMI and anthropometric measurements (36–39). Our results are consistent with these studies because waist circumference adjusted for BMI is a good indicator of central adiposity, adjusted for overall fatness (40).

Evidence from previous studies have been equivocal with some showing that European ancestry was inversely correlated with BMI in Pima Indians (41) and in African Americans (20,21) and others failing to detect such associations (42,43). Differences in the previous studies might have been due to the small number and the lower informativeness of ancestry-informative markers (20–40 markers) (20,21,42,43) used to determine genetic admixture, thus leading to imprecise estimates of ancestry. In contrast, in >3,500 African Americans, using a genome-wide panel of 1,350 highly informative markers for differentiating West African and European populations, we were able to demonstrate a significant inverse correlation between BMI and other obesity-related traits and European ancestry even after accounting for measures of SES, which is consistent with the higher prevalence of obesity in African American, compared to whites (4,5,44). It is noteworthy, however, that none of the above studies reported a high correlation between genetic ancestry and BMI. The proportion of variance explained by genetic ancestry is typically weak to modest (20). In the present study, using linear regression models, we noticed that the local ancestry alone can only account for 0.9% of the variation in BMI measurements in African Americans. This may suggest that the influence of environmental factors, either measured or unmeasured, is much greater than the influence of genetics on the disparity of obesity risk between ethnic/racial groups.

Despite the significant correlations of genetic admixture to BMI and weight, the evidence for association between the locus on 2p23.3 and both BMI and weight from admixture mapping scans was only suggestive and did not reach genome-wide significance. Although our sample size was quite large, the lack of statistical significance of this association was not surprising given the fact that the effect sizes for common variants influencing obesity are generally small. A single variant in the *FTO* gene, the first locus identified by genome-wide association studies with impact on obesity, explains only ~1% of adult BMI variation (45). It will be important to follow-up this study by analysis of more samples to better understand the signal.

Given the suggestive statistical evidence for peak on 2p23.3, we constructed the 95% credible interval for this locus and examined whether any genes in this region might be biologically relevant to obesity-related phenotypes (13,16). The localization spans 21.6–29.7 Mb (~8.1 Mb) in build 35 of the human genome reference sequence and contained many genes that may have biological relevance to BMI and weight regulation. One such gene is the pro-opiomelanocortin (*POMC*) gene. In prior linkage studies, genomic regions containing the *POMC* gene have been linked to leptin levels, a predictor of obesity, and/or fat mass in African Americans (46) and other ethnic/racial groups (47,48). POMC is a precursor of several neuropeptides, such as adrenocorticotropin and α-melanocyte-stimulating hormone, acting at hypothalamic melanocortin receptors to reduce food intake and regulate energy homeostasis (49). Studies also showed that the effects of leptin on controlling appetite are mainly mediated by the hypothalamic melanocortin pathway (50). Despite these various lines of evidence, association studies of the *POMC* gene and obesity-related traits have been inconsistent (51–53). To our knowledge, only one small association study ($N = 242$) was conducted in African Americans (53), in which children homozygous for *POMC* A7429G had a nonsignificant trend for greater BMI, compared to wild-type or heterozygote children. Further mapping work is also needed to determine whether other variants in the interval are causative for higher levels of BMI.

In conclusion, we have carried out a genome-wide admixture mapping scan in 3,531 African Americans and identified a risk locus for obesity at 2p23.3 where the correlation between local ancestry and BMI was strong and suggested a genetic effect beyond the effects of global ancestry on BMI. Follow-up fine mapping or haplotype tagging across the peak will be necessary to determine whether this region harbors genetic variations that may partially account for differences in obesity risk between African Americans and whites.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/oby

ACKNOWLEDGMENTS

We thank the staff and participants of the Atherosclerosis Risk in Communities (ARIC) study for their important contributions. This work is supported by the National Institute of Diabetes and Digestive and Kidney Diseases grants R21DK073482 and K01DK067207 (W.H.L.K.). The ARIC study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022 and infrastructure is partly supported by grant UL1RR025005, a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. D.R. is supported by a Burroughs Wellcome Career Development Award in the Biomedical Sciences and D.R., N.P. and A.T. are supported by the NIH grant U01-HG004168. Genotyping services were provided by the Johns Hopkins University under federal contract number (N01-HV-48195) from the NHLBI and by the Center for Inherited Disease Research (CDIR). CDIR is fully funded through a federal contract from the NIH to the Johns Hopkins University, contract number N01-HG-65403.

DISCLOSURE

The authors declared no conflict of interest.

REFERENCES

1. Hill JO, Peters JC. Environmental contributions to the obesity epidemic. *Science* 1998;280:1371–1374.
2. Bouchard C. Genetics of human obesity: recent results from linkage studies. *J Nutr* 1997;127:1887S–1890S.
3. Comuzzie AG, Allison DB. The search for human obesity genes. *Science* 1998;280:1374–1377.
4. Ogden CL, Carroll MD, Curtin LR *et al*. Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* 2006;295:1549–1555.
5. Wang Y, Beydoun MA. The obesity epidemic in the United States—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiol Rev* 2007;29:6–28.
6. Rife DC. Populations of hybrid origin as source material for the detection of linkage. *Am J Hum Genet* 1954;6:26–33.
7. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 1988;85:9119–9123.
8. McKeigue PM. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am J Hum Genet* 1997;60:188–196.
9. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. *Am J Hum Genet* 2004;74:965–978.
10. Patterson N, Hattangadi N, Lane B *et al*. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004;74:979–1000.
11. Smith MW, Patterson N, Lautenberger JA *et al*. A high-density admixture map for disease gene discovery in african americans. *Am J Hum Genet* 2004;74:1001–1013.
12. Carlson CS, Eberle MA, Rieder MJ *et al*. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet* 2003;33:518–521.
13. Reich D, Patterson N, De Jager PL *et al*. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 2005;37:1113–1118.
14. Freedman ML, Haiman CA, Patterson N *et al*. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 2006;103:14068–14073.
15. Zhu X, Luke A, Cooper RS *et al*. Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* 2005;37:177–181.
16. Reich D, Patterson N, Ramesh V *et al*. Admixture mapping of an allele affecting interleukin 6 soluble receptor and interleukin 6 levels. *Am J Hum Genet* 2007;80:716–726.
17. Nalls MA, Wilson JG, Patterson NJ *et al*. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am J Hum Genet* 2008;82:81–87.
18. Kao WH, Klag MJ, Meoni LA *et al*. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nat Genet* 2008;40:1185–1192.
19. Haiman CA, Patterson N, Freedman ML *et al*. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638–644.
20. 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;119:624–633.
21. Fernández JR, Shriver MD, Beasley TM *et al*. Association of African genetic admixture with resting metabolic rate and obesity among women. *Obes Res* 2003;11:904–911.
22. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 1989;129:687–702.
23. Ferrario M, Carpenter MA, Chambless LE. Reliability of body fat distribution measurements. The ARIC Study baseline cohort results. Atherosclerosis Risk in Communities Study. *Int J Obes Relat Metab Disord* 1995;19:449–457.
24. Steffes MW, Gross MD, Schreiner PJ *et al*. Serum adiponectin in young adults—interactions with central adiposity, circulating levels of glucose, and insulin resistance: the CARDIA study. *Ann Epidemiol* 2004;14:492–498.
25. Bureau of the Census. 1980 Census of population: classified index of industries and occupations. Bureau of the Census: Washington, DC, 1982.
26. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr* 1982;36:936–942.
27. Willett WC, Sampson L, Stampfer MJ *et al*. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51–65.
28. Anderson JT, Jacobs DR Jr, Foster N *et al*. Scoring systems for evaluating dietary pattern effect on serum cholesterol. *Prev Med* 1979;8:525–537.
29. Hinds DA, Stuve LL, Nilsen GB *et al*. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072–1079.
30. Frazer KA, Ballinger DG, Cox DR *et al*. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861.
31. Fan JB, Oliphant A, Shen R *et al*. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 2003;68:69–78.
32. Deo RC, Patterson N, Tandon A *et al*. A high-density admixture scan in 1,670 African Americans with hypertension. *PLoS Genet* 2007;3:e196.
33. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:e190.
34. Price AL, Patterson NJ, Plenge RM *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909.
35. Keys A, Anderson JT, Grande F. Prediction of serum-cholesterol responses of man to changes in fats in the diet. *Lancet* 1957;273:959–966.
36. Carroll JF, Chiapa AL, Rodriquez M *et al*. Visceral fat, waist circumference, and BMI: impact of race/ethnicity. *Obesity (Silver Spring)* 2008;16:600–607.
37. Lovejoy JC, de la, Bretonne JA, Klemperer M, Tulley R. Abdominal fat distribution and metabolic risk factors: effects of race. *Metabolism* 1996;45:1119–1124.
38. Conway JM, Yanovski SZ, Avila NA, Hubbard VS. Visceral adipose tissue differences in black and white women. *Am J Clin Nutr* 1995;61:765–771.
39. Kanaley JA, Giannopoulou I, Tillapaugh-Fay G, Nappi JS, Ploutz-Snyder LL. Racial differences in subcutaneous and visceral fat distribution in postmenopausal black and white women. *Metab Clin Exp* 2003;52:186–191.
40. Katzmarzyk PT, Pérusse L, Rao DC, Bouchard C. Familial risk of obesity and central adipose tissue distribution in the general Canadian population. *Am J Epidemiol* 1999;149:933–942.
41. Williams RC, Long JC, Hanson RL, Sievers ML, Knowler WC. Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. *Am J Hum Genet* 2000;66:527–538.
42. Shaffer JR, Kammerer CM, Reich D *et al*. Genetic markers for ancestry are correlated with body composition traits in older African Americans. *Osteoporos Int* 2007;18:733–741.
43. 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;121:565–575.

44. Hedley AA, Ogden CL, Johnson CL *et al*. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *JAMA* 2004;291:2847–2850.
45. Frayling TM, Timpson NJ, Weedon MN *et al*. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;316:889–894.
46. Rotimi CN, Comuzzie AG, Lowe WL *et al*. The quantitative trait locus on chromosome 2 for serum leptin levels is confirmed in African-Americans. *Diabetes* 1999;48:643–644.
47. Comuzzie AG, Hixson JE, Almasy L *et al*. A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet* 1997;15:273–276.
48. Hager J, Dina C, Francke S *et al*. A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nat Genet* 1998;20:304–308.
49. Pritchard LE, Turnbull AV, White A. Pro-opiomelanocortin processing in the hypothalamus: impact on melanocortin signalling and obesity. *J Endocrinol* 2002;172:411–421.
50. Boston BA, Blaydon KM, Varnerin J, Cone RD. Independent and additive effects of central POMC and leptin pathways on murine obesity. *Science* 1997;278:1641–1644.
51. Baker M, Gaukrodger N, Mayosi BM *et al*. Association between common polymorphisms of the proopiomelanocortin gene and body fat distribution: a family study. *Diabetes* 2005;54:2492–2496.
52. Sutton BS, Langefeld CD, Williams AH *et al*. Association of proopiomelanocortin gene polymorphisms with obesity in the IRAS family study. *Obes Res* 2005;13:1491–1498.
53. Feng N, Adler-Wailes D, Elberg J *et al*. Sequence variants of the POMC gene and their associations with body composition in children. *Obes Res* 2003;11:619–624.