

## Exploration of the utility of ancestry informative markers for genetic association studies of African Americans with type 2 diabetes and end stage renal disease

Keith L. Keene · Josyf C. Mychaleckyj · Tennille S. Leak · Shelly G. Smith · Peter S. Perlegas · Jasmin Divers · Carl D. Langefeld · Barry I. Freedman · Donald W. Bowden · Michèle M. Sale

Received: 3 March 2008 / Accepted: 15 July 2008 / Published online: 25 July 2008  
© Springer-Verlag 2008

**Abstract** Admixture and population stratification are major concerns in genetic association studies. We wished to evaluate the impact of admixture using empirically derived data from genetic association studies of African Americans (AA) with type 2 diabetes (T2DM) and end-stage renal disease (ESRD). Seventy ancestry informative markers (AIMs) were genotyped in 577 AA with T2DM–ESRD, 596 AA controls, 44 Yoruba Nigerian (YRI) and 39 European American (EA) controls. Genotypic data and association results for eight T2DM candidate gene studies in our AA population were included. Ancestral estimates were calculated using FRAPPE, ADMIXMAP and STRUCTURE for all AA samples, using varying numbers of AIMs (25, 50, and 70). Ancestry estimates varied significantly across all three programs with the highest estimates

obtained using STRUCTURE, followed by ADMIXMAP; while FRAPPE estimates were the lowest. FRAPPE estimates were similar using varying numbers of AIMs, while STRUCTURE estimates using 25 AIMs differed from estimates using 50 and 70 AIMs. Female T2DM–ESRD cases showed higher mean African proportions as compared to female controls, male cases, and male controls. Age showed a weak but significant correlation with individual ancestral estimates in AA cases ( $r^2 = 0.101$ ;  $P = 0.019$ ) and in the combined set ( $r^2 = 0.131$ ;  $P = 3.57 \times 10^{-5}$ ). The absolute difference between frequencies in parental populations, absolute  $\delta$ , was correlated with admixture impact for dominant, additive, and recessive genotypic models of association. This study presents exploratory analyses of the impact of admixture on studies of AA with T2DM–ESRD and supports the use of ancestral proportions as a means of reducing confounding effects due to admixture.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-008-0532-6) contains supplementary material, which is available to authorized users.

K. L. Keene · J. C. Mychaleckyj · T. S. Leak · S. G. Smith · D. W. Bowden · M. M. Sale  
Center for Human Genomics,  
Wake Forest University School of Medicine,  
Winston-Salem, NC, USA

K. L. Keene · J. C. Mychaleckyj · M. M. Sale (✉)  
Center for Public Health Genomics, University of Virginia,  
PO Box 800717, Charlottesville, VA 22908, USA  
e-mail: msale@virginia.edu

K. L. Keene · M. M. Sale  
Department of Biochemistry and Molecular Genetics,  
University of Virginia, Charlottesville, VA, USA

J. C. Mychaleckyj  
Public Health Sciences, University of Virginia,  
Charlottesville, VA, USA

J. C. Mychaleckyj · B. I. Freedman · D. W. Bowden · M. M. Sale  
Department of Internal Medicine,  
Wake Forest University School of Medicine,  
Winston-Salem, NC, USA

P. S. Perlegas · D. W. Bowden  
Department of Biochemistry,  
Wake Forest University School of Medicine,  
Winston-Salem, NC, USA

J. Divers · C. D. Langefeld  
Department of Biostatistical Sciences,  
Wake Forest University School of Medicine,  
Winston-Salem, NC, USA

M. M. Sale  
Department of Medicine, University of Virginia,  
Charlottesville, VA, USA

## Introduction

In the USA, 20.8 million people, or 7% of the total population, have diabetes (Kaiser Family Foundation 2007). Diabetes is a growing epidemic whose risk factors include obesity, race, and family history. An African American (AA) is twice as likely to develop diabetes as compared to a Caucasian peer (Brancati et al. 2000). Evidence of a genetic contribution to the susceptibility of type 2 diabetes mellitus (T2DM) has been established through twin (Hawkes 1997), familial aggregation (Rotimi et al. 1994), and segregation studies (Goodman and Chung 1975).

There are a limited number of genetic association studies evaluating the susceptibility of minority populations, especially AA, to developing T2DM (Elbein 2007). Genetic association studies using admixed populations are presented with the additional challenge of addressing the contributions due to admixture. An admixed population is the product of mating between individuals from reproductively isolated ancestral populations (Tang et al. 2005). The two largest minority groups in the United States, African Americans and Hispanic Americans, are both examples of admixed populations. Markers that are differentially distributed between two parental populations provide more information regarding their ancestral origin (Pfaff et al. 2004). This divergence in frequencies may result in deviations from Hardy Weinberg Equilibrium (HWE) and produce apparent but spurious association results with disease traits in the presence of admixture linkage disequilibrium (ALD) (Long 1991).

One approach for locating and identifying disease genes is through pedigree-linkage analysis followed by subsequent genetic-association studies. Association studies often attempt to identify susceptibility loci using population samples containing unrelated individuals. Spurious associations may occur if the samples are obtained from two or more subpopulations (population substructure) and if the distribution of the trait and the allele frequency of the polymorphism varies among the subpopulations (Reiner et al. 2005). Although there is a growing debate concerning the contributions of admixture in association studies, few projects address this issue using empirically derived (Fernandez and Shiver 2004), rather than simulated, data.

In an effort to evaluate the impact of admixture on individual estimates of ancestry and association studies of T2DM-ESRD, 70 AIMs were genotyped in a population of 577 AA with T2DM-ESRD, 596 AA controls, 44 Yoruba Nigerians and 39 EA controls. Additionally, genotypic results from studies of eight T2DM candidate genes (TCF7L2, PPAR $\gamma$ , CAPN10, KCNJ11, TCF1, HNF4 $\alpha$ , ESR1 and ENPP1) in AA with T2DM-ESRD were also included to further explore the impact of admixture on estimates of association. This study represents the first

exploratory study of the impact of admixture in genetic association studies of AA with T2DM-ESRD.

## Materials and methods

### Subjects

This study was conducted under Institutional Review Board approval from Wake Forest University School of Medicine, and adhered to the tenets of the Declaration of Helsinki. Identification, clinical characteristics, and recruitment of AA and European American (EA) patients and controls have been described previously (Freedman et al. 1997; Yu et al. 1996). Briefly, 577 unrelated AA patients with T2DM born in North Carolina, South Carolina, Georgia, Tennessee or Virginia were recruited from dialysis facilities. Individuals with type 1 diabetes, as diagnosed in patients with a history of diabetic ketoacidosis, or who developed diabetes mellitus prior to age 25 and received continuous insulin therapy since diagnosis, were excluded. A diagnosis of T2DM was based upon an initial diagnosis of diabetes mellitus after age 35 years, receiving oral hypoglycemic agents or dietary therapy without insulin for at least one year after initial diagnosis, and active treatment with diabetes medications. Cases had T2DM diagnosed at least 5 years prior to initiating renal replacement therapy, background or greater diabetic retinopathy and/or >3+ proteinuria on urinalysis in the absence of other causes of nephropathy. Five hundred ninety-six unrelated AA controls born in North Carolina, South Carolina, Georgia, Tennessee or Virginia, who did not have a current diagnosis of T2DM or renal disease, were recruited. Thirty-nine unrelated EA controls without known T2DM or renal disease were recruited using the same recruitment strategy as our AA controls. DNA extraction was performed using the PureGene system (Gentra Systems, Minneapolis, MN, USA). DNA was also obtained from 44 Yoruba Nigerians (YRI) from the National Institute of General Medical Sciences (NIGMS) Human Variation Collection (Coriell Cell Repositories, Camden, NJ).

### AIMs selection and genotyping

Seventy biallelic Ancestry Informative Markers (AIMs) were genotyped by Illumina Inc.'s Custom Genotyping Service (San Diego, CA, USA) or using a MassARRAY genotyping system (Sequenom Inc., San Diego, CA, USA) (Buetow et al. 2001) in 577 AA with T2DM-ESRD, 596 AA controls, 44 YRI and 39 EA controls (Supplementary Table 1). The AIMs were chosen from the previous literature (Shriver et al. 2003; Smith et al. 2004), or to ensure at least one SNP on each autosome, and adequate physical map separation to ensure linkage equilibrium. Fifty-three

AIMs were genotyped using Illumina Inc.'s Custom Genotyping Service (San Diego, CA, USA). The remaining seventeen AIMs were genotyped using iPLEX or hME methodology on a MassARRAY genotyping system (Sequenom Inc., San Diego, CA, USA) (Buetow et al. 2001): rs2814778, rs6003, rs2065160, rs2752, rs3287, rs3340, rs455726, rs2763, rs2161, rs3176921, rs680273, rs1042602, rs1800498, rs1800404, rs4646, rs2228478, and rs2816. Primer sequences are available on request.

#### Statistical analyses

Hardy Weinberg equilibrium (HWE) values were determined by calculating a  $\chi^2$  statistic and corresponding *P* value.

Ancestral proportions were calculated using the programs STRUCTURE (Pritchard et al. 2000), FRAPPE (Frequentist Estimation of individual ancestry proportion) (Tang et al. 2005) and ADMIXMAP (McKeigue et al. 2000) under a two population model. Estimates of “pseudo-ancestral” allele frequencies were obtained from genotyped YRI and EA samples.

Delta values ( $\delta$ ), defined as the difference between the YRI major allele frequency and corresponding allele frequency in EA, were calculated for each AIM using the genotyped data from the 44 YRI and 39 EA individuals. Absolute  $\delta$  values were used to rank the AIMs, with the highest and lowest absolute  $\delta$  values earning a ranking of 1 and 70, respectively (Supplementary Table 1). AIMs were then grouped into three sets, representing the subset of AIMs with the highest  $\delta$  values: “Top 25” (the 25 SNPs with the largest absolute  $\delta$  values), “Top 50” (50 highest absolute  $\delta$  values), and all 70 AIMs (“All”) with all programs.

Statistical analysis for correlation and analysis of variance (ANOVA) were determined using Graph Pad Prism 5 ([www.graphpad.com](http://www.graphpad.com)). Non parametric (Spearman) correlation coefficients and corresponding *P*-values were calculated for each pair of data sets. Non parametric ANOVA (Kruskal–Wallis) tests were performed to determine differences in means across groups.

#### African American association with T2DM-ESRD

Genotypic association results for 208 SNPs from eight T2DM candidate genes: 2 Calpain 10 (CAPN10) SNPs, 1 Peroxisome proliferator-activated receptor  $\gamma$  (PPARG) SNP, 5 Transcription factor 7-like 2 (TCF7L2) SNPs, 1 ATP-sensitive inwardly rectifying potassium channel subunit Kir6.2 (KCNJ11) SNP, 1 Hepatic transcription factor 1 (TCF1) SNP, 3 Hepatocyte nuclear factor 4- $\alpha$  (HNF4A) SNPs (Sale et al. 2007), 146 Estrogen receptor  $\alpha$  gene (ESR1) SNPs (Keene et al. 2008b) and 49 Ectonucleotide

Pyrophosphatase/Phosphodiesterase 1 (ENPP1) SNPs, (Keene et al. 2008a), genotyped in this same African American (AA) case-control population, were included in this analysis to determine the impact of admixture adjustments in AA genetic association studies of T2DM-ESRD. Briefly, 150 SNPs, spanning more than 476 kb, located in the coding and flanking regions of ESR1 were genotyped in 577 AA T2DM-ESRD cases and 596 AA controls. The genotyping success rates for the 150 SNPs in the AA cases and controls range from 94.1 to 100%. Forty-nine ENPP1 SNPs were genotyped in the same 577 AA T2DM-ESRD cases and 596 controls. The genotyping success rates for the 49 SNPs in the AA cases and controls range from 90.4 to 100%. TCF7L2 SNP genotyping success rates were 100% in case subjects and >99.8% in control subjects. Genotyping success rates for the remaining candidate SNPs were >97.7% in case subjects and >96.5% in control subjects.

Unadjusted genotypic tests for the dominant, additive, and recessive models were evaluated by calculating a  $\chi^2$  statistic and corresponding *P* value using the program SNP-GWA (C. Langefeld, personal communication). Due to a lack of validity of the large sample  $\chi^2$  test statistic, only the dominant model *P* values were considered for SNPs with ten or fewer individuals that were homozygous for the minor allele. Admixture adjusted *P* values were subsequently determined by performing covariate adjustment of individual estimates of African ancestry using 70 AIMs (using FRAPPE estimates) for AA subjects in logistic regression tests of dominant, additive, and recessive models of association as implemented in the program SNPADMIX (C. Langefeld, personal communication).

## Results

### Population characteristics

Characteristics of the AA case and control populations have been published previously (Sale et al. 2007) and are shown in Supplementary Table 2. Controls were significantly younger than cases ( $P < 0.0001$ ), although they were significantly older than the mean age at T2DM diagnosis in cases ( $P < 0.0001$ ). Additionally, age data were unavailable for 148 controls recruited during the early phase of the study. There was a higher proportion of females cases (61%) than female controls (51%).

### Admixture informative markers

Genotyping success rates for the 70 AIMs were 97.2–100% in AA cases, 94.1–100% in AA controls, 95.5–100% in Yoruban Africans, and 94.9–100% in Caucasian Europeans (CE). The mean absolute delta values ( $\pm$ SD) were

$0.836 \pm 0.054$ ,  $0.743 \pm 0.120$ , and  $0.617 \pm 0.237$  for 25, 50, and 70 AIMS, respectively. Sixty-one of the 70 AIMS had an absolute delta value  $\geq 0.300$ , while the remaining nine had values ranging from 0.055 to 0.278.

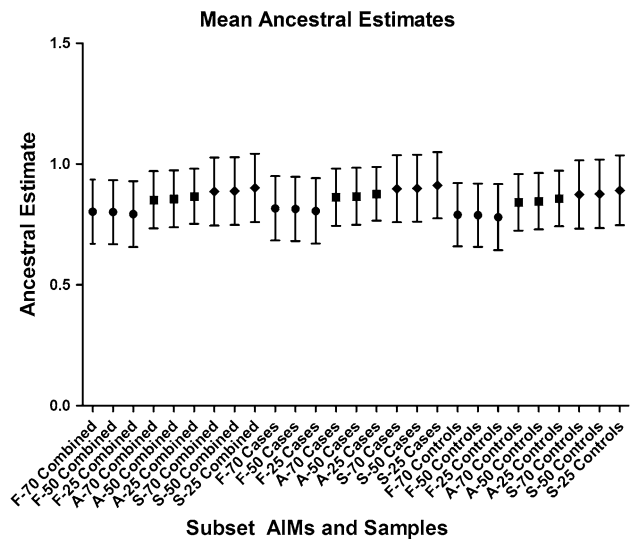
#### Deviations from HWE

Using a Hardy Weinberg Equilibrium (HWE) threshold of  $P < 0.01$ , three AIMS deviated from expected HWE proportions in the AA cases (rs2228478,  $P = 2.80 \times 10^{-9}$ ; rs2814778,  $P = 0.00008$ ; and rs4821667,  $P = 0.0022$ ) and eight SNPs (rs2228478,  $P = 1.5 \times 10^{-10}$ ; rs1941141,  $P = 0.00033$ ; rs7957445,  $P = 0.00094$ ; rs3287,  $P = 0.0014$ ; rs3176921,  $P = 0.0015$ ; rs1374092,  $P = 0.0054$ ; rs7349,  $P = 0.0065$ ; and rs4436849,  $P = 0.0086$ ) were inconsistent with HWE in the AA controls.

For SNPs genotyped in candidate genes (without regard to ancestral frequencies), two ENPP1 SNPs, rs858341 ( $P = 0.004$ ) and rs9493105 ( $P = 0.001$ ) deviated from expected HWE proportions in the AA controls while all SNPs were consistent with HWE in the AA cases. Five ESR1 SNPs deviated from expected HWE proportions in the AA cases (rs9397459, rs9341070, rs9340969, rs1569788, and rs722208) and one SNP (rs9341070) was inconsistent with HWE in the AA controls. One additional SNP was inconsistent with HWE in both case and control subjects (HNF4A SNP rs4810424;  $P < 0.0001$  in both case and control subjects).

#### Estimates of ancestral proportions

Estimates of African ancestral proportions from STRUCTURE, FRAPPE and ADMIXMAP, calculated using varying numbers of AIMS (25, 50, and 70), are shown in Fig. 1. The mean proportion ( $\pm$ SD) of African ancestry in the combined sample set was  $0.91 \pm 0.14$ ,  $0.90 \pm 0.14$ , and  $0.90 \pm 0.14$  for STRUCTURE, 0.79  $\pm$  0.14, 0.80  $\pm$  0.13, and 0.80  $\pm$  0.13 for FRAPPE and 0.87  $\pm$  0.11, 0.86  $\pm$  0.12, and 0.85  $\pm$  0.12 for ADMIXMAP using 25, 50, and 70 AIMS, respectively. The individual ancestral estimates for each program and each subset were highly correlated ( $r^2 > 0.90$ ;  $P < 0.0001$ ) (Fig. 2 a–c), however, the mean group ancestral estimates were highest using STRUCTURE, followed by ADMIXMAP, while FRAPPE estimates were the lowest. Supplementary Figure 1 displays the distribution of admixture estimates for the combined set for all programs using the three subsets of AIMS. Ancestral estimates obtained from FRAPPE were similar across the three subsets, as determined by performing one way ANOVA analysis ( $P = 0.070$ ). In contrast, ADMIXMAP estimates comparing 70 AIMS versus 25 AIMS were statistically different ( $P < 0.05$ ), while 50 AIMS versus 25 AIMS and 70 AIMS versus 50 AIMS were similar. However, using



**Fig. 1** Mean ancestral estimates using FRAPPE (F circles), ADMIXMAP (A squares), and STRUCTURE (S diamonds) with varying numbers of AIMS (25, 50, and 70) in AA cases and controls (Combined,  $n = 1,173$ ), AA cases only ( $n = 577$ ), and AA controls only ( $n = 596$ )

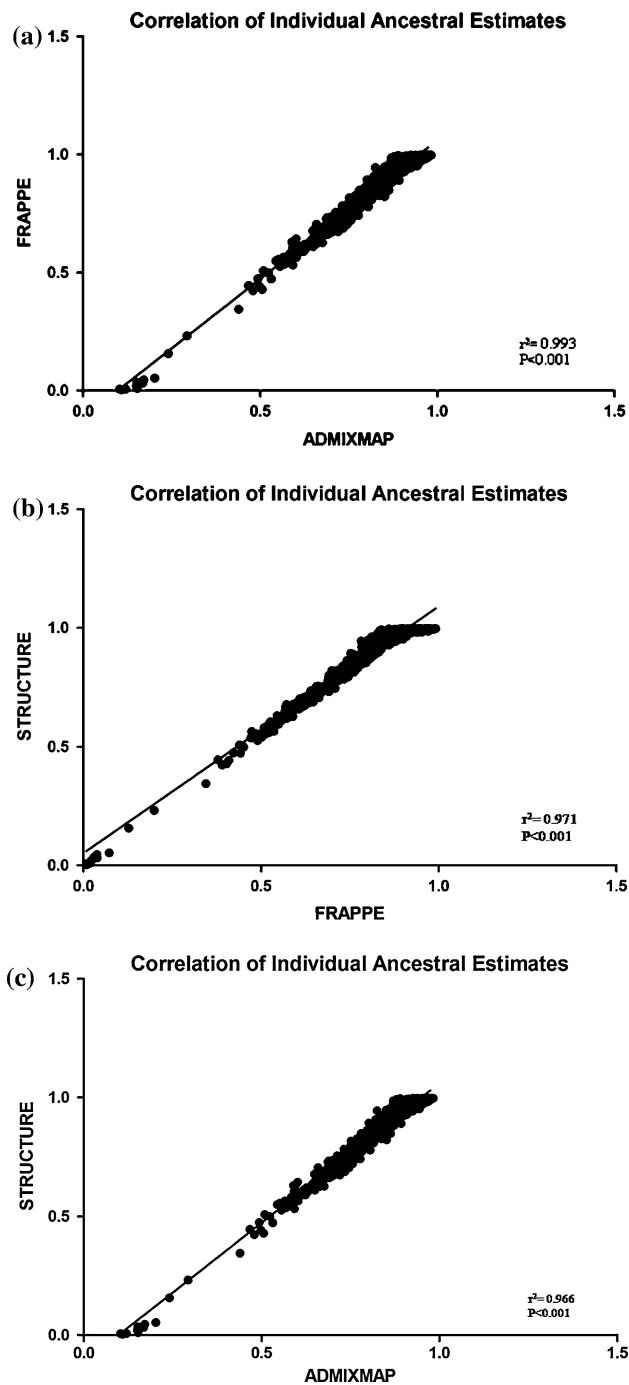
STRUCTURE, only estimates for 70 and 50 AIMS were similar.

Gender stratified distribution of admixture estimates for male and female cases and controls are presented in Fig. 3. The mean proportion of African ancestry (as determined using FRAPPE, 70 AIMS  $\pm$  SD) was  $0.781 \pm 0.144$  ( $n = 306$ ) for the female controls, and  $0.826 \pm 0.123$  ( $n = 352$ ) for the female cases. The mean proportion of African ancestry was  $0.800 \pm 0.115$  ( $n = 290$ ) for the male controls, and  $0.803 \pm 0.148$  ( $n = 225$ ) for the male cases. Using ANOVA, the mean African ancestry estimates for the female cases differed significantly from the female controls ( $P < 0.0001$ ), all samples combined ( $P < 0.05$ ), all controls ( $P < 0.0001$ ), and male controls ( $P < 0.01$ ), but were not significantly different from male case estimates.

There was a significant but weak correlation between age and individual admixture estimates in the AA cases ( $r^2 = 0.101$ ;  $P = 0.019$ ) and combined (AA cases and controls) ( $r^2 = 0.131$ ;  $P = 3.57 \times 10^{-5}$ ) that was not observed in the AA controls ( $r^2 = 0.016$ ;  $P = 0.733$ ) alone (Supplementary Figure 2a–c).

#### Impact of admixture on genotypic tests of association

Admixture adjustments were performed for all genotypic associations of candidate genes. Seven of the nine ENPP1 SNPs and 25 of the 31 ESR1 SNPs remained significantly associated with T2DM–ESRD in one or more genotypic models of association, after adjusting for admixture. Of the remaining 13 SNPs, seven SNPs were associated with T2DM–ESRD, six of which remained significant after



**Fig. 2** Correlation of individual ancestral estimates for all samples using 70 AIMs: **a** FRAPPE and ADMIXMAP; **b** FRAPPE and STRUCTURE; and **c** ADMIXMAP and STRUCTURE

adjusting for admixture. The impact of admixture on association results is depicted in Fig. 4, showing the difference between the  $-\log(\text{unadjusted})$  and  $-\log(\text{admixture adjusted})$   $P$  value, plotted against the corresponding absolute  $\delta$  value for that SNP. The impact of admixture was significantly correlated with absolute delta value for the dominant ( $r^2 = 0.1997$ ;  $P = 0.0051$ ), additive ( $r^2 = 0.2662$ ;

$P = 0.0015$ ), and recessive ( $r^2 = 0.1735$ ;  $P = 0.0410$ ) models of association.

## Discussion

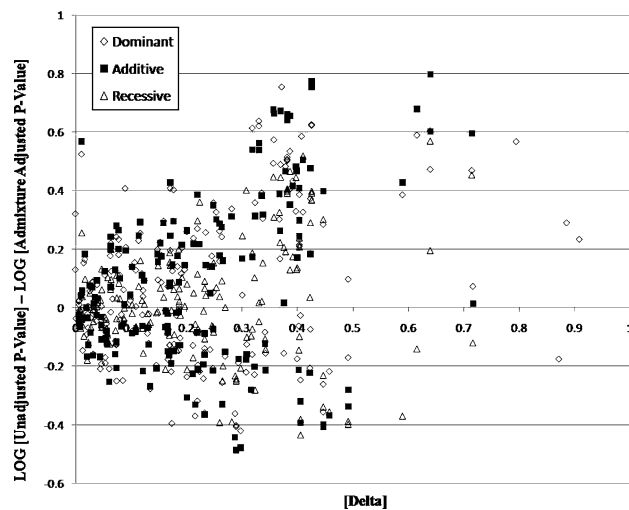
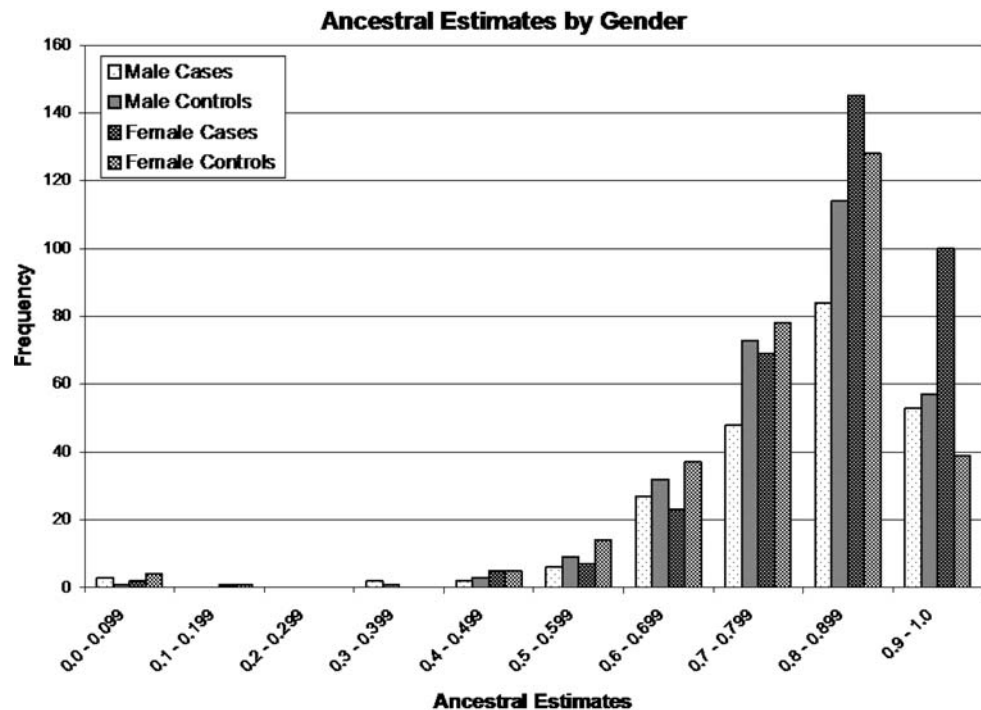
Several studies have investigated the theoretical impact of admixture association with disease traits using simulated data. We have evaluated the impact of admixture on individual estimates of ancestry generated using three methods, correlations with gender and age, and the impact on association results.

Individual ancestral estimates, calculated using FRAPPE, ADMIXMAP, and STRUCTURE were highly correlated ( $r^2 > 0.90$ ;  $P < 0.0001$ ). However, the overall group means for STRUCTURE were significantly higher compared to ADMIXMAP and FRAPPE, which is consistent with reports using simulated data (Tsai et al. 2005). These differences in estimations can be contributed to the methodologies implemented in the various programs (Maximum Likelihood Estimation (FRAPPE) and Bayesian methodologies (ADMIXMAP and STRUCTURE)) as well as the way each program generates ancestral allele frequencies from the two parental populations. Using ANOVA, intra-program mean ancestral estimates using FRAPPE were similar for all three subsets of AIMS (25, 50, 70, Fig. 1). In contrast mean ancestral estimates from STRUCTURE were similar only for estimates obtained using 70 and 50 AIMS. However, ADMIXMAP estimates were similar when comparing 70 and 50 AIMS, as well as 50 and 25 AIMS but the mean group estimates using 25 AIMS were significantly higher than the estimates obtained using the total set of 70 AIMS, suggesting that 25 highly informative AIMS may not be sufficient to determine “accurate” individual ancestral proportions, but 50 highly informative AIMS may suffice. Although using 100 AIMS for estimating ancestry in Latino subjects has been proposed by Tsai et al. (2005), their data also suggests that more informative markers ( $F_{ST}$  higher than 0.5) could reduce the number of markers needed to decrease the type I error rate. When they simulated 25 AIMS with  $F_{ST}$  higher than 0.5, the number of AIMS required for controlling the excess of false positives was less than 50 (Tsai et al. 2005).

The mean ancestral estimate (using FRAPPE, 70 AIMS) for the combined AA in this study ( $0.80 \pm 0.13$ ) was similar to previous AA estimates from Winston Salem, NC (0.791) (Reiner et al. 2005), Washington, DC (0.787) (Shriver et al. 2003), Chicago, IL (.806) (Smith et al. 2004), Birmingham, AL (0.812) (Reiner et al. 2007), and New York (0.802), but different from Gullah speaking AA from the Sea Islands of coastal South Carolina (0.965) (Parra et al. 2001).

Gender stratified ancestral estimates ( $\pm$ SD) (as determined using FRAPPE, 70 AIMS) were  $0.781 \pm 0.145$  for

**Fig. 3** Histogram displaying the distribution of individual ancestral estimates from FRAPPE for AA male cases ( $n = 225$ ), female cases ( $n = 352$ ), male controls ( $n = 290$ ), and female controls ( $n = 306$ )



**Fig. 4** Relationship between absolute  $\delta$  and the impact of admixture (using FRAPPE generated individuals ancestral estimates) on genotypic tests of association for dominant, additive, and recessive models in 208 SNPs from eight candidate gene studies for T2DM. Absolute  $\delta$ , defined as the absolute difference between the Yoruban African major allele frequency and corresponding allele frequency in European Americans, is plotted along the x-axis while the admixture impact, defined as the difference between the  $-\log(\text{unadjusted})$  and  $-\log(\text{admixture adjusted})$   $P$  value is plotted along the y-axis

female controls ( $n = 306$ ),  $0.826 \pm 0.123$  for female cases ( $n = 352$ ),  $0.800 \pm 0.115$  for male controls ( $n = 290$ ), and  $0.803 \pm 0.148$  for male cases ( $n = 225$ ). ANOVA analysis showed the mean ancestral estimate in female cases differed significantly from male ( $P < 0.01$ ), and female controls ( $P < 0.0001$ ), but were similar to male case estimates.

These findings also hold true when performing logit transformations of ancestry estimates (data not shown). Although ancestral lineage differences have been supported by studies utilizing X, Y, and mitochondrial markers, (Lind et al. 2007; McLean et al. 2005), the divergent frequencies in the female cases and controls suggests that females from our recruitment region with higher African proportion may have an increased risk for developing T2DM–ESRD. This theory is supported by the unexplained increase in risk for having a family history of ESRD that is present among African American women, relative to men (Freedman et al. 2005; McClellan et al. 2007). The higher proportion of ancestry in female cases may hint at a biological relationship between gender and greater African ancestry. A modification of the thrifty gene hypothesis (Neel 1999) is one possible explanation for the higher African ancestry observed in AA female T2DM–ESRD patients. If AA females with T2DM have genes of “African descent” that are involved in energy storage and expenditure, but also influence or are influenced by gender-specific pathways, when exposed to a westernized lifestyle, these individuals will be at an increased risk of obesity and subsequently developing T2DM. African ancestry and markers for ancestry have previously shown to be correlated with body composition traits in older AA (Shaffer et al. 2007) and associated with resting metabolic rate and obesity in AA women ( $n = 145$ ) (Fernandez et al. 2003). One possible pathway, suggested by our previous association studies, is through estrogen receptor mediated actions. Estrogen has shown to have antidiabetic actions in both rodent models as well as clinical trials (Louet et al. 2004; Margolis et al.

2004; Seed 2002). Our association results suggest a differential association with ESR1 variants and T2DM–ESRD in males (7 SNPs) versus females (35 SNPs) (Keene et al. 2008b).

There was a significant but weak correlation between age and individual admixture estimates in the AA cases ( $r^2 = 0.101$ ;  $P = 0.019$ ) and combined (AA cases and controls) ( $r^2 = 0.131$ ;  $P = 3.57 \times 10^{-5}$ ), that was not observed in the AA controls ( $r^2 = -0.016$ ;  $P = 0.733$ ) alone. The slope of the trend line for the cases suggests that older AA patients may have higher African ancestry. The lack of significance in the controls may be due to missing age data for 148 controls. Eighty-two individuals with missing age data had an individual ancestry estimate greater than 0.79 (the control group mean) while only 12 individuals with missing age data had an individual ancestry estimate less than 0.60, suggesting a bias towards missing age data in AA controls with higher ancestral data, which may have impacted our ability to detect the modest relationship observed in cases. The trend in the AA cases does suggest that younger generations of AA may have increasing levels of European admixture, therefore it may be important to test for this correlation before performing age- and admixture adjusted analyses.

The impact of admixture on association results was significantly correlated with absolute  $\delta$  value for the dominant, additive, and recessive genotypic models of association. Additionally, there was no correlation between minor allele frequency or genotypic frequency with the impact of admixture, suggesting that there is no bias of admixture effect based on the actual allele frequency in the admixed population. Of the 208 SNPs, 47 (22.6%) were nominally associated with T2DM–ESRD (before adjusting for admixture) in this AA population. Of those 47 SNPs, nine were no longer associated with T2DM–ESRD after adjusting for African ancestry (Supplementary Table 3). Although these nine SNPs only represent 4% of the total SNP count, they account for nearly 20% of all of the associated SNPs, thus demonstrating the increased chance of type I error due to admixture. Admixture also affected the most highly associated SNP in this study, TCF7L2 SNP rs7903146, with an admixture-adjusted dominant model  $P$  value of  $2.55 \times 10^{-8}$ , compared with the unadjusted dominant model  $P$  value of  $1.79 \times 10^{-8}$ . Associations for this SNP have been replicated in T2DM association studies from several populations (Sale et al. 2007) and the SNP has very similar allele frequencies for Africans and Caucasians, with an absolute  $\delta$  value of only 0.0149. We conclude that admixture adjustments should be performed for all markers in an admixed population regardless of parental population frequencies,  $\delta$  values, or actual allele and genotypic frequencies in the admixed population; however it is especially important to account for admixture in genetic

association studies where the frequency of the trait/disease of interest and the allele frequency of the polymorphism varies significantly among the ancestral populations (sub-populations).

Increasing the sample size of the parental populations may provide more accurate estimates of the parental frequencies for each marker, which in turn could improve individual estimates of ancestry. Additionally, sampling from more divergent African populations may also increase the accuracy of these estimates, although historical records and genetic studies (Reed and Tishkoff 2006) suggest that the majority of AA have ancestral origins from West Africa. Ancestry estimates can be determined and may vary depending on the DNA source (i.e. autosomes, sex chromosomes and mitochondrial DNA). One limitation of this study was the lack of ancestral estimates using mitochondrial DNA and sex chromosomes, however the estimates obtained using 70 autosomes appear to be informative for ancestry and useful for association studies.

This study is the first such admixture study of its kind to analyze the effects of admixture in AA populations with T2DM–ESRD. Investigation of differing number of AIMs suggest as few as 50 AIMs may be sufficient to generate reliable ancestral estimates in AA, although more studies are necessary to verify this finding. Additionally, this study provides evidence that higher proportions of African ancestry show a modest correlation with age in AA cases from this population, and even greater correlation with T2DM–ESRD status in AA females. The higher ancestry in AA females suggests that there may be a biological relationship between African ancestry and sex as it relates to T2DM–ESRD, however further evaluations are necessary in order to identify the exact relationship. This study provides insight into the influence and critical importance of considering admixture in AA genetic association studies.

**Acknowledgments** We thank the individuals recruited as cases and controls for their participation, recruiters Joyce Byers and Mitzie Spainhour, technician Candace Gordon, programmer Matt Stiegert, and Mark Hansen and colleagues at Illumina Inc. This work was supported by grants DK066358, DK072550, DK070941, the Wake Forest University General Clinical Research Center M01 RR07122, and a Career Development Award from the American Diabetes Association (MMS).

## References

- Brancati FL, Kao WH, Folsom AR, Watson RL, Szklo M (2000) Incident type 2 diabetes mellitus in African American and white adults: the atherosclerosis risk in communities study. *JAMA* 283:2253–2259
- Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A (2001) High-throughput development and characterization of a genome-wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization

- time-of-flight mass spectrometry. *Proc Natl Acad Sci USA* 98:581–584
- Elbein SC (2007) Evaluation of polymorphisms known to contribute to risk for diabetes in African and African-American populations. *Curr Opin Clin Nutr Metab Care* 10:415–419
- Fernandez JR, Shriver MD (2004) Using genetic admixture to study the biology of obesity traits and to map genes in admixed populations. *Nutr Rev* 62:S69–S74
- Fernandez JR, Shriver MD, Beasley TM, Rafla-Demetriou N, Parra E, Albu J, Nicklas B, Ryan AS, McKeigue PM, Hoggart CL, Weinsier RL, Allison DB (2003) Association of African genetic admixture with resting metabolic rate and obesity among women. *Obes Res* 11:904–911
- Freedman BI, Yu H, Spray BJ, Rich SS, Rothschild CB, Bowden DW (1997) Genetic linkage analysis of growth factor loci and end-stage renal disease in African Americans. *Kidney Int* 51:819–825
- Freedman BI, Volkova NV, Satko SG, Krisher J, Jurkovicz C, Soucie JM, McClellan WM (2005) Population-based screening for family history of end-stage renal disease among incident dialysis patients. *Am J Nephrol* 25:529–535
- Goodman MJ, Chung CS (1975) Diabetes mellitus: discrimination between single locus and multifactorial models of inheritance. *Clin Genet* 8:66–74
- Hawkes CH (1997) Twin studies in diabetes mellitus. *Diabet Med* 14:347–352
- Lind JM, Hutcheson-Dilks HB, Williams SM, Moore JH, Essex M, Ruiz-Pesini E, Wallace DC, Tishkoff SA, O'Brien SJ, Smith MW (2007) Elevated male European and female African contributions to the genomes of African American individuals. *Hum Genet* 120:713–722
- Long JC (1991) The genetic structure of admixed populations. *Genetics* 127:417–428
- Kaiser Family Foundation (2007) State Health Facts Online. <http://www.statehealthfacts.org/>
- Keene KL, Mychaleckyj JC, Smith SG, Leak TS, Perlegas PS, Langefeld CD, Freedman BI, Rich SS, Bowden DW, Sale MM (2008a) Association of the distal region of the ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) gene with type 2 diabetes in an African American population enriched for nephropathy. *Diabetes* 57(4):1057–1062
- Keene KL, Mychaleckyj JC, Smith SG, Leak TS, Perlegas PS, Langefeld CD, Herrington DM, Freedman BI, Rich SS, Bowden DW, Sale MM (2008b) Comprehensive evaluation of the estrogen receptor alpha gene reveals further evidence for association with type 2 diabetes enriched for nephropathy in an African American population. *Hum Genet* 123(4):333–341
- Louet JF, LeMay C, Mauvais-Jarvis F (2004) Antidiabetic actions of estrogen: insight from human and genetic mouse models. *Curr Atheroscler Rep* 6:180–185
- Margolis KL, Bonds DE, Rodabough RJ, Tinker L, Phillips LS, Allen C, Bassford T, Burke G, Torrens J, Howard BV (2004) Effect of oestrogen plus progestin on the incidence of diabetes in postmenopausal women: results from the Women's Health Initiative Hormone Trial. *Diabetologia* 47:1175–1187
- McClellan W, Speckman R, McClure L, Howard V, Campbell RC, Cushman M, Audhya P, Howard G, Warnock DG (2007) Prevalence and characteristics of a family history of end-stage renal disease among adults in the United States population: reasons for geographic and racial differences in stroke (REGARDS) renal cohort study. *J Am Soc Nephrol* 18:1344–1352
- McKeigue PM, Carpenter JR, Parra EJ, Shriver MD (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet* 64:171–186
- McLean DC Jr, Spruill I, Argyropoulos G, Page GP, Shriver MD, Garvey WT (2005) Mitochondrial DNA (mtDNA) haplotypes reveal maternal population genetic affinities of Sea Island Gullah-speaking African Americans. *Am J Phys Anthropol* 127:427–438
- Neel JV (1999) Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? 1962. *Bull World Health Organ* 77:694–703 Discussion 692–693
- Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 114:18–29
- Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC (2004) Information on ancestry from genetic markers. *Genet Epidemiol* 26:305–315
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959
- Reed FA, Tishkoff SA (2006) African human diversity, origins and migrations. *Curr Opin Genet Dev* 16:597–605
- Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, Phong A, Burchard EG, Harris TB, Psaty BM, Kwok PY (2005) Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet* 76:463–477
- Reiner AP, Carlson CS, Ziv E, Iribarren C, Jaquish CE, Nickerson DA (2007) Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African-American participants in the CARDIA Study. *Hum Genet* 121:565–575
- Rotimi C, Cooper R, Cao G, Sundarum C, McGee D (1994) Familial aggregation of cardiovascular diseases in African-American pedigrees. *Genet Epidemiol* 11:397–407
- Sale MM, Smith SG, Mychaleckyj JC, Keene KL, Langefeld CD, Leak TS, Hicks PJ, Bowden DW, Rich SS, Freedman BI (2007) Variants of the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in an African-American population enriched for nephropathy. *Diabetes* 56(10):2638–2642
- Seed M (2002) The choice of hormone replacement therapy or statin therapy in the treatment of hyperlipidemic postmenopausal women. *Atheroscler Suppl* 3:53–63
- Shaffer JR, Kammerer CM, Reich D, McDonald G, Patterson N, Goodpaster B, Bauer DC, Li J, Newman AB, Cauley JA, Harris TB, Tyllavsky F, Ferrell RE, Zmuda JM (2007) Genetic markers for ancestry are correlated with body composition traits in older African Americans. *Osteoporos Int* 18:733–741
- Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 112:387–399
- Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL, Mignault AA, Yi Z, De The G, Essex M, Sankale JL, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De La Vega FM, Woodage T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, O'Brien SJ, Reich D (2004) A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 74:1001–1013
- Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 28:289–301
- Tsai HJ, Choudhry S, Naqvi M, Rodriguez-Cintron W, Burchard EG, Ziv E (2005) Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. *Hum Genet* 118:424–433
- Yu H, Bowden DW, Spray BJ, Rich SS, Freedman BI (1996) Linkage analysis between loci in the renin-angiotensin axis and end-stage renal disease in African Americans. *J Am Soc Nephrol* 7:2559–2564