RESEARCH ARTICLE

Comparing genetic ancestry and self-reported race/ethnicity in a multiethnic population in New York City

YIN LENG LEE[1,2†], SUSAN TEITELBAUM[1†], MARY S. WOLFF[1], JAMES G. WETMUR[2,3] and JIA CHEN[1,4,5*]

[1]*Departments of Preventive Medicine,* [2]*Department of Microbiology,*
[3]*Department of Genetics and Genomic Sciences,* [4]*Department of Pediatrics, and*
[5]*Department of Oncological Science, Mount Sinai School of Medicine, New York, NY 10029, USA*

Abstract

Self-reported race/ethnicity is frequently used in epidemiological studies to assess an individual's background origin. However, in admixed populations such as Hispanic, self-reported race/ethnicity may not accurately represent them genetically because they are admixed with European, African and Native American ancestry. We estimated the proportions of genetic admixture in an ethnically diverse population of 396 mothers and 188 of their children with 35 ancestry informative markers (AIMs) using the STRUCTURE version 2.2 program. The majority of the markers showed significant deviation from Hardy–Weinberg equilibrium in our study population. In mothers self-identified as Black and White, the imputed ancestry proportions were 77.6% African and 75.1% European respectively, while the racial composition among self-identified Hispanics was 29.2% European, 26.0% African, and 44.8% Native American. We also investigated the utility of AIMs by showing the improved fitness of models in paraoxanase-1 genotype–phenotype associations after incorporating AIMs; however, the improvement was moderate at best. In summary, a minimal set of 35 AIMs is sufficient to detect population stratification and estimate the proportion of individual genetic admixture; however, the utility of these markers remains questionable.

[Lee Y. L., Teitelbaum S., Wolff M. S., Wetmur J. G. and Chen J. 2010 Comparing genetic ancestry and self-reported race/ethnicity in a multiethnic population in New York City. *J. Genet*. **89**, 417–423]

Introduction

Race and ethnicity are widely used interchangeably in population research. The ambiguity and interchangeable nature of these terms raises concerns about the scientific validity of statistical comparisons that use variables based on them (LaVeist 1994; Lee *et al*. 2001; Braun 2002). Race is linked to individual genetic makeup, which is not necessarily related to cultural or environmental characteristics. To the contrary, ethnicity refers to shared customs, beliefs and traditions among population subgroups that may or may not have a common genetic origin.

According to US Census Bureau, Hispanics, who can be of any race, constitute the largest ethnic minority in United States. They constitute 15.5% of the current US population and are expected to reach 24% by 2050. The Hispanic population is genetically diverse, represening a heterogeneous mix of European, African and Native American ancestry (Gonzalez Burchard *et al*. 2005; Salari *et al*. 2005). Therefore, a Hispanic individual may selfidentify as a single race or multiple races. There are wide variations across and within Hispanic ethnic groups in terms of genetic, socioeconomic, cultural and geographic origin.

The complicated genetic structure of contemporary admixed populations has several important implications for conducting genetic epidemiology studies. Population stratification, referring to the difference in allele frequencies between cases and controls, may give rise to false association of genes with disease (Pritchard and Rosenberg 1999). Population stratification exists in ethnically admixed populations; without proper statistical adjustment, such stratification may lead to false-positive or false-negative results and produce incorrect associations between genes and disease. Self-reported ancestry can be inaccurate for determin-

*For correspondence. E-mail: jia.chen@mssm.edu.
†These authors contributed equally to this work.

Keywords. population stratification; ancestry informative markers (AIMs); race; ethnicity; genetic epidemiology; forensic genetics; Hispanics.

ing an individual's actual genetic ancestry (Choudhry *et al*. 2007). The use of self-reported race and/or ethnicity for Hispanic study participants in epidemiologic studies is particularly problematic due to the possibility that these individuals may not be fully aware of their own complex ancestry mixture. Genetic markers, such as ancestry informative markers (AIMs), may provide more accurate information on potential population stratification. AIMs and newly developed statistical methods are making the genetic estimation of ancestry increasingly more feasible and accurate (Hoggart *et al*. 2003; Pritchard *et al*. 2000a,b). Using a panel of genetic polymorphisms that present large differences in allelic frequencies (> 0.40) between Europeans and Africans, it is possible to estimate the degree of European and African admixture among Hispanics (Ziv *et al*. 2006).

We used a panel of 35 AIMs to estimate genetic ancestry in an ethnically diverse population in New York City, USA. These markers were selected to maximize the difference between European, African American and Native American populations. The objective is to estimate the extent of genetic admixture in this population with a minimal number of single nucleotide polymorphisms (SNPs). The selection of such a small number of markers also minimizes the cost for genotyping. To demonstrate the utility of these markers, we re-examined the genotype–phenotype relationship of paraoxonase-1 in this population (Chen *et al*. 2003) by adjusting for population stratification as determined by AIMs.

Materials and methods

Population

We utilized the resources of an ongoing prospective study conducted at the Mount Sinai Center for Children's Environmental Health and Disease Prevention Research (Berkowitz *et al*. 2003). The cohort consists of an ethnically diverse group of mother–infant pairs participating in a longitudinal study assessing infant growth and neurodevelopment associated with environmental exposures in urban New York City. Subjects were recruited during early pregnancy from the prenatal clinic and two private practices at Mount Sinai Hospital from March 1998 to March 2002. The details of the study design have been previously described (Berkowitz *et al*. 2004). Maternal blood samples were obtained in heparin-treated vacutainers, and DNA was extracted and purified using Hi-Pure PCR Template Preparation kits (Roche Applied Scheme, Indianapolis, USA) as described by the manufacturer. The study population consists of 404 prenatal patients, 396 with sufficient DNA for genotyping. We also obtained sufficient DNA from 188 cord blood samples from the infants born to the study participants.

Race/ethnicity information

A structured questionnaire was administered by interviewers in English or Spanish to collect parent's demographic information such as country of birth, socio-demographic characteristics, maternal health, lifestyle habits and residential history. Participants were asked to specify a single race/ethnic group based on six categories; White, Black, Black Hispanic, White Hispanic, Asian or Other.

Selection of AIMs

AIMs were selected to maximize the absolute difference in allele frequency among ancestral populations. A set of 35 AIMs (table 1) was selected based on previously published literatures (Salari *et al*. 2005; Choudhry *et al*. 2006; Ziv *et al*. 2006). These markers were selected because they exhibited largest difference in allele frequency between any two of the three ancestral groups, i.e. European, African and Native American. Europeans were sampled from Utah residents with ancestry from northern and western Europe while Africans from Yoruba in Nigeria. Native American samples were from western United States, Mexico and Central America. These AIMs were distributed across the genome but distant from functional domains of the genome so it is unlikely that these markers would result in any disease phenotype. Detailed information about the flanking sequences and other information for all 35 AIMs are available from the HapMap database (www.hapmap.org).

Genotyping for AIMs

Genotyping was performed at the DNA core (Shared Research Facility) of Mount Sinai School of Medicine using SNPlex method under standard conditions (Applied Biosystems, Foster City, USA). It utilizes pre-optimized universal assay reagents kits and a set of SNP-ligation probes to perform genotyping up to 48-plex (48 SNPs genotyped in a single reaction). The analysis system collects and manages raw data and provides automated allele calling and quality metrics (Tobler *et al*. 2005).

Characterization of population structure and admixture

Using the exact test, we examined deviations from Hardy–Weinberg equilibrium (HWE) of each AIM using a software developed by Paul Lewis (http://www.eeb.uconn.edu/people/plewis/software.php). We estimated individual genetic ancestry and number of ancestral populations (K) among mothers and infants, separately, by Bayesian Markov chain Monte Carlo (MCMC) method implemented in the program STRUCTURE 2.2 available at http://pritch.bsd.uchicago.edu/software.html (Pritchard *et al*. 2000a; Falush *et al*. 2003). STRUCTURE arbitrarily groups individuals into clusters based on a Bayesian approach that uses the distribution and likelihood estimation to determine the distribution of population frequency. STRUCTURE was run with a burn-in length of 100,000 and 100,000 iterations after burn in without any prior population assignment under the admixture model using correlated allele frequencies and the number of populations ($K = 3$), designated as European, African, and Native American.

Statistical analysis

Results on genotype and phenotype of *PON1* in the same population have been published previously (Chen *et al*. 2003). Both race-stratified (based on self-reported category) and race-adjusted genotype–phenotype relationships had been reported in the original paper, in which standard multiple regression techniques were employed using PROC REG and PROC GLM of SAS software, version 9.1(SAS Institute, Carolina, USA). Here, we report the adjusted coefficient of determination, R^2, the percentage of the total variation in *PON1* phenotype explained by various models. To test for improvements of multivariate models, we contrasted the R^2 of these models with and without AIMs.

Results

Our study population consists of multiracial/multiethnic mother–infant pairs in New York City. Of 396 mothers included in these analyses, the composition of the self-reported race/ethnicity was as follows (table 2): White (19.9%), Black (28.3%), Hispanics (47.5%), and non-specified (4.3%). Among 188 selfidentified Hispanics, most of whom were of Caribbean origin, the majority (64.0%) identified their race as 'other Hispanic', not Black or White.

We used a panel of 35 AIMs to examine population stratification of this population. As expected, the majority of the markers (27/35) showed significant deviation ($P < 0.05$) from HWE indicating significant population

Table 1. Thirty-five ancestry informative markers with NCBI reference numbers, chromosomal locations and allele frequencies in European, African and Native Americans.

Marker name	dbSNP rs no.	Location	Allele frequencies		
			African	European	Native American
TYR-192	rs1042602	11q21	0.00	0.47	0.05
OCA2	rs1800404	15q13.1	0.14	0.72	0.48
DRD2-Taq	rs1800498	11q23.1	0.14	0.65	0.09
TSC0003523	rs1985080	7p14.3	0.10	0.64	0.97
TSC0745571	rs203096	17q21.33	0.65	0.72	0.28
TSC1102055	rs2065160	1q32.1	0.50	0.92	0.17
TSC0042312	rs2077863	18p11	0.51	0.93	0.93
WI-4019	rs2161	7q22.1	0.44	0.30	0.62
MC1R-314	rs2228478	16q24.3	0.51	0.14	0.04
	rs223830	16q13	0.03	0.19	0.64
TSC1023401	rs235936	21q21.3	0.18	0.49	0.37
WI-11909	rs2695	9q21.31	0.81	0.86	0.22
WI-11392	rs2752	1q42.2	0.88	0.43	0.63
FY-null	rs2814778	1q23.2	0.00	0.99	0.99
WI-7423	rs2816	17p12	0.00	0.49	0.08
LPL	rs285	8p21.3	0.97	0.52	0.45
WI-14319	rs2862	15Q14	0.38	0.17	0.69
WI-14867	rs2891	17p13.2	0.02	0.51	0.43
CRH	rs3176921	8q13.1	0.32	0.93	0.98
WI-16857	rs3287	2p16.1	0.73	0.20	0.21
SGC30610	rs3309	5q11.2	0.40	0.28	0.69
SGC30055	rs3317	5q23.1	0.05	0.59	0.73
WI-17163	rs3340	5q33.2	0.06	0.19	0.65
CYP19	rs4646	15q21.2	0.32	0.29	0.72
TSC0879894	rs518116	9q33.3	0.13	0.67	0.58
GNB3	rs5443	12p13.31	0.80	0.33	0.36
TSC1157118	rs584059	3q22.3	0.49	0.14	0.47
F13B	rs6003	1q31.3	0.70	0.08	0.03
GC1	rs7041	4q13.3	0.93	0.41	0.45
TSC0050288	rs722098	21q21.1	0.90	0.18	0.72
TSC0053441	rs723632	1q32.3	0.10	0.92	0.67
TSC0000409	rs7349	10p11.22	0.04	0.87	0.96
TSC0030203	rs736394	14q32.12	0.52	0.74	0.99
TSC0255473	rs930072	5p13.2	0.96	0.10	0.45
TSC0316844	rs994174	10q23.1	0.76	0.25	0.26

AIMs markers and allele frequencies are compiled from Salari *et al*. (2005); Choudhry *et al*. (2006); Ziv *et al*. (2006).

Table 2. Estimate of European, African and Native American admixture in mothers of Children Environmental Cohort study, 1998–2002 using STRUCTURE with 35 AIMs in mothers and 34 AIMs in infants.

Race (selfreported)	Percentage population ancestry							
	European				African		Native American	
	n	%	Mean	SD	Mean	SD	Mean	SD
Mothers								
White	79	19.9	75.1	19.0	6.6	12.0	18.3	12.8
Black	112	28.3	10.2	14.0	77.6	24.2	12.2	14.6
Hispanic	188		29.2	23.0	26.0	25.9	44.8	27.0
Black	28	7.1	18.5	20.2	55.5	31.6	26.0	23.3
White	39	9.8	31.3	21.8	19.6	20.3	49.1	26.6
Other	121	30.6	30.9	23.4	21.3	21.3	47.8	26.3
Non-specified	17	4.3						
Infants								
White	40	23.7	68.8	21.6	7.6	13.4	23.7	16.8
Black	45	26.6	11.5	11.9	73.5	22.4	15.0	14.1
Hispanic	80		28.0	19.8	30.0	25.3	42.0	23.0
Black	8	4.7	29.1	29.9	45.6	37.1	25.4	23.1
White	13	7.7	28.5	23.4	17.8	15.8	53.7	24.5
Other	59	34.9	27.8	17.6	31.1	24.1	41.2	21.3
Non-specified	4	2.4						

stratification. However, the number of markers that depart from HWE drastically decreased when analyses were restricted to each racial group, with 2, 4 and 4 markers in Blacks, White and Hispanics, respectively. This small number of genes that depart from HWE may reflect gene flow between and among different racial groups. Our study population is racially and ethnically mixed; the overall mean proportions of European, African, and Native American in mothers as estimated by the STRUCTURE program were 32.8%, 37.6% and 29.6%, respectively. We then examined the imputed cluster distribution against self-reported race/ethnicity. Agreement between self-identified White and Black and population cluster assignments by STRUCTURE was fairly consistent; imputed ancestry proportions were 75.1% European and 77.5% African for these two groups, respectively. However, there was a less consistent pattern among the self-identified Hispanics, where the imputed cluster distribution was 29.2% European, 26.0% African and 44.8% Native American. Further categorization by race (e.g., Black Hispanics, White Hispanics, Others) among the Hispanics revealed a pattern that is moderately reflective of the self-report of combined race-ethnicity. For example, the women who reported their ancestry as Black Hispanic had a relatively high African cluster (55.5%) while those who self-identified as White Hispanics had higher Native American cluster (49.1%) than the European cluster (31.3%).

The assay for one of the SNPs in our AIMs panel, rs2077863, failed for cord blood, leaving 34 markers. Similar to the mothers, the majority of the markers (22/34) for the newborns showed significant deviation ($P < 0.05$) from HWE indicating significant population stratification. Not surprisingly, the racial/ethnic composition of the overall infant population was similar to that of their mothers (data not shown). Among the self-identified Hispanics, we also observed a similar pattern when comparing the child's race/ethnicity inferred from the mother's self-report to the STRUCTURE estimated population ancestry proportion (table 2). The African ancestry cluster among Black Hispanic infants was 45.6%, whereas the European ancestry cluster among the White Hispanic infants was 28.5%, much lower than the Native American ancestry cluster of 53.7%. This difference may partially reflect the ancestry of the infants' fathers, which was not included in the self-reported race–ethnicity categorization.

We have previously reported the genotype–phenotype association of paraoxonase-1 in the same population (Chen *et al*. 2003). The measurements were carried out with phenylacetate as the substrate so that activity could be independent of the *PON1* Q192R polymorphism (Chen *et al*. 2003). In the original report, we observed that the enzymatic activity of *PON1* varied by race/ethnicity as well as the distribution of SNPs in *PON1*. As a result, we reported genotype–phenotype associations either stratified or adjusted by race/ethnicity. Herein, we reanalysed the data by incorporating AIMs into the models. The analysis was restricted to the individuals with complete data for *PON1* genotype and phenotype as well as AIMS. To illustrate the utility of AIMs, we included only results on one promoter SNP (−108) and one coding SNP (L55M) as examples. Although R^2 values differed, similar trends were observed for −909 and −162 SNPs that were in the original report. In addition, −909 SNP was tightly linked with −108 SNP. L55M is in linkage disequilibrium

with the −108 promoter SNP with $D' = 0.5$ (Chen *et al*. 2003). There are 350 mothers and 132 children who have complete data on AIMs, *PON1* genotypes and phenotypes. Similar to the original report, the R^2 of infants was much larger than that of the mothers (table 3) reflecting the stronger genetic / less environmental influence on the enzymatic activity of *PON1*. Also shown in table 3, with respect to both *PON1* SNPs, models without adjusting for race or AIMs always yielded the smallest R^2. A gradual increase in R^2 was observed by incorporating self-reported race alone, AIMs alone, and finally race and AIMs together, indicating overall improvement in terms of fitness of these models. The improvement, however, is very modest in mothers and infants. For example, with respect to the L55M SNP, the R^2 was 0.082 in the unadjusted model and became 0.123 in the model that was adjusted for race and AIMs in mothers. In comparison, the R^2 went from 0.435 to 0.513 in infants. We also performed similar analyses restricted to self-identified Hispanics with additional adjustment by AIMs (table 3, lower panel). Again, AIMs offered very modest improvement in mothers and infants. Lastly, when analyses were restricted to the 132 mother–infant pairs for whom we had complete genotype as well as phenotype data, similar results were observed (data not shown).

Discussion

Our goal was to estimate the extent of individual ancestry proportion using a small panel of AIMs in our New York City population and to examine the reliability of self-reported race/ethnicity information from questionnaires. Results from this investigation demonstrate that this panel of 35 AIMs provides reliable estimates of ancestry proportion in individuals who self-identified as White or Black. Further, this panel of AIMs highlights the multiracial makeup of the Hispanic women who participated in our study and reflects their heterogeneous origins. However, it also demonstrates that the use of a single 'Hispanic' category may be insufficient for characterizing genetic background associated with race in epidemiologic or other analyses.

There are numerous reports on using informative markers for ancestry inference; these reports vary drastically in terms of selection criteria, the nature of the marker (microsatellite versus SNPs) as well as the number of AIMs, especially with respect to Hispanic populations. The minimal cutoff point for AIMs varies from $\delta = 0.6$ (Shaffer *et al*. 2007) to $\delta = 0.5$ (Reiner *et al*. 2005; Yaeger *et al*. 2008) to $\delta = 0.3$ (Choudhry *et al*. 2006; Ziv *et al*. 2006). The number of markers varied from 326 microsatellite markers in Tang *et al*. (2005), 75–100 AIMs in Parra *et al*. (1998) and Shriver *et al*. (1997), to 744 AIMs in Smith *et al*. (2001). Risch *et al*. (2002) showed that 20 AIMs provide the same information as 120 randomly selected markers when trying to discriminate European from Native American or African. In our study, we demonstrated that the use of a small number of AIMs ($n = 35$) allowed us to satisfactorily distinguish the ancestral population in our racially diverse population while minimize genotyping costs and DNA usage.

Estimates of Native American ancestry in our cohort are different from other studies among self-identified Hispanics. The mean proportion of Native American ancestry among our self-identified Hispanic participants was 48.8% by using AIMs, which is more than two-fold higher than what was observed in another study of Hispanics of Puerto Rican origin (17.6%) (Bonilla *et al*. 2004). This may be due to

Table 3. Proportion of variance (R^2) in PON1 activity explained by *PON1* genotype adjusted for self-reported race/ethnicity and AIMs.

		R^2	
PONI genotype		Mothers ($n = 350$)	Infants ($n = 132$)
−108 (C/T)	AIMs and race	0.169	0.438
	AIMs only	0.152	0.407
	Race only	0.156	0.426
	None	0.110	0.407
55 (L/M)	AIMs and race	0.123	0.513
	AIMs only	0.106	0.453
	Race only	0.106	0.510
	None	0.082	0.435
Self-identified Hispanics		Mothers ($n = 172$)	Infants ($n = 62$)
−108 (C/T)	AIMs	0.216	0.343
	no AIMs	0.205	0.336
55 (L/M)	AIMs	0.159	0.390
	no AIMs	0.144	0.374

misclassification associated with restricted choices for reporting race/ethnicity in our study questionnaire, in which Puerto Rican, Dominican, Latin American or other Hispanic origins were grouped together as a single response. In addition, the questionnaire did not offer an opportunity for our participants to choose multiple responses on their ancestral heritage, thus prevented detailed analysis of women with differing or multiple Hispanic origins. Salari *et al*. (2005) showed that Mexicans and Puerto Ricans are different in term of ancestry. Study location may have also contributed to the differences observed between our study and others, since the Hispanic population that resides our study area (East Harlem, New York, USA) is likely to have different origins than those living on the West Coast of USA, where majority of other studies were conducted (Ziv *et al*. 2006).

While self-identified information on Black and White is generally reliable, self-identified Hispanics are genetically admixed with three ancestral populations (Gonzalez Burchard *et al*. 2005). Therefore, including 'Hispanic' as a race category does not identify a genetically distinct racial group because individuals grouped together under this category do not necessarily possess a similar genetic ancestry. On the other hand, these individuals may share customs, beliefs, traditions, and lifestyles. If there are a large number of individuals who identify as Hispanics a population, AIMs may be necessary to correct for population stratification to prevent bias in estimates of genetic associations which could possibly lead to false-positive or false-negative associations. This small number of AIMs is cost effective as it can be done on small amounts of DNA with a readily available platform. However, the additional information provided by AIMs was limited in our data. Although we demonstrated improved fit in *PON1* genotype–phenotype associations after incorporating AIMs into the models, the improvement is far from substantial.

Race is a complex term and when used to define a population for research studies, it may mistakenly imply a biological explanation for social and health disparities. The way we define the population can have significant implications for how we interpret the scientific meaning of genetic findings. Without appropriate control for race and ethnicity, observed associations may be biased. Hence, detailed characterization with regard to self-reported ethnicity, place of birth and country of origin should be given detailed attention in the Hispanic population for future epidemiology and genetic studies.

Acknowledgements

This work was supported by grants from the National Institutes of Health (ES09584) and Environmental Protection Agency (R827039). Susan Teitelbaum is supported by grants from the National Institutes of Health (K01-ES012645) and Department of Defense (W81XWH-04-1-0507). Jia Chen is partially supported by National Institutes of Health (CA109753).

References

Berkowitz G. S., Obel J., Deych E., Lapinski R., Godbold J., Liu Z. *et al*. 2003 Exposure to indoor pesticides during pregnancy in a multiethnic, urban cohort. *Environ. Health Perspect*. **1**, 79–84.

Berkowitz G. S., Wetmur J. G., Birman-Deych E., Obel J., Lapinski R. H. *et al*. 2004 In utero pesticide exposure, maternal paraoxonase activity, and head circumference. *Environ. Health Perspect*. **3**, 388–391.

Bonilla C., Shriver M. D., Parra E. J., Jones A. and Fernandez J. R. 2004 Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. *Hum. Genet*. **1**, 57–68.

Braun L. 2002 Race, ethnicity, and health: can genetics explain disparities? *Perspect. Biol. Med*. **2**, 159–174.

Chen J., Kumar M., Chan W., Berkowitz G. and Wetmur J. G. 2003 Increased influence of genetic variation on PON1 activity in neonates. *Environ. Health Perspect*. **11**, 1403–1409.

Choudhry S., Coyle N. E., Tang H., Salari K., Lind D., Clark S. L. *et al*. 2006 Population stratification confounds genetic association studies among Latinos. *Hum. Genet*. **5**, 652–664.

Choudhry S., Seibold M. A., Borrell L. N., Tang H., Serebrisky D., Chapela R. *et al*. 2007 Dissecting complex diseases in complex populations: asthma in latino americans. *Proc. Am. Thorac. Soc*. **3**, 226–233.

Falush D., Stephens M. and Pritchard J. K. 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **4**, 1567–1587.

Gonzalez Burchard E., Borrell L. N., Choudhry S., Naqvi M., Tsai H. J., Rodriguez Santana J. R. *et al*. 2005 Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *Am. J. Public Health* **12**, 2161–2168.

Hoggart C. J., Parra E. J., Shriver M. D., Bonilla C., Kittles R. A., Clayton D. G. and McKeigue P. M. 2003 Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet*. **6**, 1492–1504.

LaVeist T. A. 1994 Beyond dummy variables and sample selection: what health services researchers ought to know about race as a variable. *Health Serv. Res*. **1**, 1–16.

Lee S. S., Mountain J. and Koenig B. A. 2001 The meanings of "race" in the new genomics: implications for health disparities research. *Yale J. Health Policy Law Ethics* **1**, 33–75.

Parra E. J., Marcini A., Akey J., Martinson J., Batzer M. A., Cooper R. *et al*. 1998 Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet*. **6**, 1839–1851.

Pritchard J. K. and Rosenberg N. A. 1999 Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet*. **1**, 220–228.

Pritchard J. K., Stephens M., Rosenberg N. A. and Donnelly P. 2000a Association mapping in structured populations. *Am. J. Hum. Genet*. **1**, 170–181.

Pritchard J. K., Stephens M. and Donnelly P. 2000b Inference of population structure using multilocus genotype data. *Genetics* **2**, 945–959.

Reiner A. P., Ziv E., Lind D. L., Nievergelt C. M., Schork N. J., Cummings S. R. *et al*. 2005 Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am. J. Hum. Genet*. **3**, 463–477.

Risch N., Burchard E., Ziv E. and Tang H. 2002 Categorization of humans in biomedical research: genes, race and disease. *Genome Biol*. **7**, comment 2007.

Salari K., Choudhry S., Tang H., Naqvi M., Lind D., Avila P. C. *et al*. 2005 Genetic admixture and asthma-related phenotypes

in Mexican American and Puerto Rican asthmatics. *Genet. Epidemiol*. **1**, 76–86.

Shaffer J. R., Kammerer C. M., Reich D., McDonald G., Patterson N., Goodpaster B. *et al*. 2007 Genetic markers for ancestry are correlated with body composition traits in older African Americans. *Osteoporos. Int*. **6**, 733–741.

Shriver M. D., Smith M. W., Jin L., Marcini A., Akey J. M., Deka R. and Ferrell R. E. 1997 Ethnic-affiliation estimation by use of population-specific DNA markers. *Am. J. Hum. Genet*. **4**, 957–964.

Smith M. W., Lautenberger J. A., Shin H. D., Chretien J. P., Shrestha S., Gilbert D. A. and O'Brien S. J. 2001 Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. *Am. J. Hum. Genet*. **5**, 1080–1094.

Tang H., Quertermous T., Rodriguez B., Kardia S. L., Zhu X., Brown A. *et al*. 2005 Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am. J. Hum. Genet*. **2**, 268–275.

Tobler A. R., Short S., Andersen M. R., Paner T. M., Briggs J. C., Lambert S. M. *et al*. 2005 The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J. Biomol. Tech*. **4**, 398–406.

Yaeger R., Avila-Bront A., Abdul K., Nolan P. C., Grann V. R., Birchette M. G. *et al*. 2008 Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. *Cancer Epidemiol. Biomarkers Prevent*. **6**, 1329–1338.

Ziv E., John E. M., Choudhry S., Kho J., Lorizio W., Perez-Stable E. J. and Burchard E. G. 2006 Genetic ancestry and risk factors for breast cancer among Latinas in the San Francisco Bay Area. Cancer *Epidemiol. Biomarkers Prevent*. **10**, 1878–1885.

Received 14 January 2010, in revised form 26 March 2010; accepted 28 April 2010
Published on the Web: 13 October 2010