AMERICAN JOURNAL OF HUMAN BIOLOGY 15:814–823 (2003)

Locus-Specific Genetic Diversity Between Human Populations: An Analysis of the Literature

SEYMOUR GARTE*
School of Public Health, UMDNJ, New Brunswick, New Jersey 08903, and Genetics Research Institute, Milan, Italy

ABSTRACT The debate over classification of the human species according to racial or continental lines has involved reports on genetic differences in allele frequencies of a number of loci with important biomedical functions. Such differences are in contrast with the fact that, for human beings, intrapopulation genetic diversity is larger than that seen between populations. In an attempt to address the hypothesis that certain genes show high interpopulation diversity due to selective pressure, the literature was surveyed to quantify such diversity using Wrights Fst statistic. The gene-specific Fst values were then compared to pairwise population values of Fst taken over a large number of genes, which presumably reflect mostly neutral mechanisms of genetic diversity such as drift. The results showed that the majority of pairwise population values of Fst for over 30 genes of biomedical significance were either below or within the expected limits of Fst based on published values. These results do not support the idea that positive or diversifying natural selection plays an important role in increasing genetic diversity, even in genes that might be expected to be subject to selection pressure. Balancing selection, whereby the degree of genetic diversity is actually lower than that expected, appears to occur more frequently for these genes. The fact that allele frequency differences between populations might be "statistically significant" does not therefore necessarily imply a degree of genetic diversity greater than would be expected due to nonselective mechanisms. Am. J. Hum. Biol. 15:814–823, 2003.

Geographic differences in allele frequencies between populations are generally considered to be the result of neutral genetic drift and natural selection since the beginning of migrations of modern humans out of Africa (Hamblin et al., 2002; Hartl and Clark, 1997; Bowcock et al., 1991; Cavalli-Sforza et al., 1994). Not all genes behave the same way with respect to allelic variation within the human species. For the great majority of human genes, there is little evidence of natural selection since the emergence of modern humans, and these genes show interpopulation differences in allele frequency comparable or less than intrapopulation differences (Lewontin, 1972; Relethford, 2002; Steele, 2002; Romualdi et al., 2002). The role of balancing selection for certain genes of biomedical importance has been emphasized in a study of the AIDS-related gene CCR5 (Bamshad et al., 2002), in which such selection results in less than expected diversity. Gene variants whose functions strongly interact with geographic- or population-specific environmental factors may, on the other hand, be under strong positive or diversifying selection pressure. Examples of this form of selection in humans includes the genetic basis of skin color, body size and shape, the ability to tolerate lactose and alcohol, etc. It is to be expected that for such genes the level of diversity in allele frequency between the relevant populations will be significantly higher than that predicted for drift alone.

It could be hypothesized that genes subject to strong selection pressure from environmental factors such as sunlight, temperature, diet, infectious organisms, or environmental toxins would show evidence of positive selection by a greater level of interpopulation genetic diversity than that expected from neutral drift (Garte, 2002). Besides genes that code for skin color, and other obvious geographically based phenotypes, it is possible that many genes with biomedical significance such as those involved in xenobiotic (drug or dietary) metabolism, susceptibility

Contract grant sponsor: NIH; Contract grant number: HG002595 (from the ELSI program of the National Human Genome Research Institute).

*Correspondence to: Dr. Seymour Garte, Professor, School of Public Health UMDNJ, 335 George St., Suite 3700, New Brunswick, NJ 08903. E-mail: sjgarte@yahoo.com

Received 5 May 2003; Revision received 9 July 2003; Accepted 16 July 2003

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ajhb.10215

to infection, or detoxification could be part of this group of high diversity genes due to differences in exposures to infectious and toxic agents in different environments. If true, this would help explain the apparent paradox (Garte, 2002) between the fact that many genes that show large interpopulation differences in allele frequencies, and the fact that for most of the genome intrapopulation differences are much greater than those between populations (Lewontin, 1972).

The work of Cavalli-Sforza et al. (1994) has quantified a good deal of genetic diversity between human populations and has shown how migrations, isolation, intermarriage, and other historical patterns of human movement around the globe has caused complex patterns of geographic genetic diversity. These workers devote some discussion in their landmark work to the fact that there can be no biological significance to the concept of race in the human species. At the same time, their work, as well as a large literature, clearly documents population differences in many allele frequencies that seem to follow more or less continental lines.

A number of studies have examined the issue of population genetic diversity using pooled data from many loci (Smith et al., 2001; Watkins et al., 2001; Daly et al., 2001), or using data from single genes of biomedical interest. The hypothesis described above, namely, that the level of population genetic diversity is a function of the role of each particular gene and of its interaction with external factors, has not been directly addressed. Using data from the literature on a panel of over 30 genes of biomedical importance, all of which might be expected to be under selective pressure, this hypothesis has been addressed by computation of the genetic diversity statistic Fst, for all of these genes between a number of population pairs and over all human populations.

MATERIALS AND METHODS

A Medline search of the literature led to the discovery of several hundred articles in which allele frequencies were discussed according to populations. Articles using less than 50 individuals per population were excluded. Population definition follows no universal criteria, and most reports use self-definition for American ethnicity, or country of origin for other parts of the world. Certain groups, with primarily cultural and/or linguistic definition, such as Hispanics, were not included. Whenever possible, population definitions were fit to match those given in the prime reference used for genetic diversity comparison (Cavalli-Sforza et al., 1994). Thus, African Americans were treated as Africans, although it is well known that these two populations are not equivalent, and that considerable admixture (10–20%) with Europeans has occurred with the African American population (Chakraborty et al., 1992; Parra et al., 1998). However, since the use of Africans is itself an artificial amalgam of many populations with large interpopulation genetic heterogeneity, the relative error from including African Americans with Africans is expected to be small.

Pairwise population genetic diversity was determined by calculation of Wrights Fst, as described (Cavalli-Sforza et al., 1994). The formula used was:

$$F_{st} = \frac{\sum_{i=1}^{L} p^*_i(1 - p^*_i) - F_i}{\sum_{i=1}^{L} p^*_i(1 - p^*_i)}$$

where p^*_i is the average allele frequency (over all populations) of the i-th allele, L is the number of alleles, and F_i is the value of F_{st} for each allele. This is given for two populations by:

$$F_i = \frac{\sum_{j=1}^{2} (p_{ij} - p^*_i)^2}{p^*_i(1 - p^*_i)}$$

where p_{ij} is the frequency of the i-th allele in population j.

The ARLEQUIN program (Schneider et al., 2000) was used to check these calculations. Values of Fst were also calculated over the whole human species when such data were available from published sources for at least three populations of different continental origin, to include at least Asians, Europeans, and Africans. The reference Fst for each population pair was taken from Cavalli-Sforza et al. (1994), using confidence intervals calculated from the tables of Fst values and their errors presented in the text (see table 2.3.1A, p. 75, and table 2.3.2, p. 80). Values of Fst calculated from allele frequency data were compared to the standard reference value and its 95% upper and lower confidence intervals (CI). Values within these limits were considered to be equivalent to

the reference, while values below or above the 95% CI were considered to be below or above "expected" reference values.

RESULTS

The values of Fst for pairwise population comparisons of 34 genes among the three most commonly defined human continental population groups (African, Asians, and Europeans) are shown in Table 1. Each Fst was compared to that published for the appropriate population pair by Cavalli-Sforza et al. (1994). If the value fell within the 95% confidence limits of the published value, then the degree of genetic diversity for this gene between these two populations was considered to be not significantly different from that expected from the reference value (indicated as R in the table). Values greater than the upper confidence limit could reflect the effects of positive or "diversifying" natural selection leading to increased diversity

TABLE 1. Values of Fst for pairwise comparisons of three populations for 34 genes

Gene[a]	African-European Fst	African-Asian Fst	Asian-European Fst
HLA	0.030 (−)	0.106 (−)	0.061 (R)
CCR5	0.107 (R)	0.069 (−)	0.011 (−)
APO B			0.059 (−)
HER2	0.064 (−)		
ALPHP1	0.163 (R)		
NOS	0.146 (R)		
UGT1A1	0.025 (−)	0.242 (R)	0.128 (R)
LEPR	0.017 (−)	0.127 (R)	0.266 (+)
CYP3A4	0.625 (+)	0.759 (+)	0.041 (−)
LRP	0.014 (−)		
AHR	0.072 (−)	0.073 (−)	0 (−)
CYP1A1	0.118 (R)	0.046 (−)	0.145 (R)
CYP2E1			0.122 (R)
GSTM	0.145 (R)	0.143 (R)	0 (−)
FXIII	0.031 (−)	0.263 (R)	
LPL	0.016 (−)		
HFE	0.193 (R)		
FcgRIIIB	0.014 (−)	0.136 (R)	0.231 (+)
P53	0.325 (+)		
SFD1	0.096 (R)		
APO E	0.027 (−)		
ANGIO	0.010 (−)		
MTHFR	0.102 (R)	0.194 (R)	0.079 (R)
GPIIIa	0.004 (−)		
GSTP1	0.016 (−)	0.137 (R)	0.059 (−)
CAPN10			0.025 (−)
CYP1A2	0.001 (−)	0.077 (−)	0.092 (R)
LCT	0.674 (+)	0.021 (−)	0.538 (+)
OATP-C	0.260 (+)		
IL4RA	0.090 (R)		
AGT			0.219 (+)
MC1R	0.278 (+)	0.287 (R)	0.105 (R)
COMT			0.072 (R)
GSTT	0.038 (−)	0.050 (−)	0.169 (+)

[a]Abbreviations and references for genes: HLA (histocompatability; Dawson et al., 2001), CCR5 (CC Chemokine receptor 5; Gonzalez et al., 1999), APO B (apolipoprotein B; Iso et al., 1996), HER2 (Human epidermal growth factor receptor 2; Keshava et al., 2001), ALPHP1 (alpha 1-proteinase inhibitor; Halkas et al., 1998), NOS (endothelial nitric oxide synthase; Chen et al., 2001), UGT1A1 (UDP-glucurono syltransferase 1; Beutler et al., 1998), LEPR (leptin receptor gene; Chagnon et al., 2000), CYP3A4 (Cytochrome P4503A4; Ball et al., 1999), LRP (low density lipoprotein receptor-related protein; Harris et al., 1998), AHR (aryl hydrocarbon receptor; Wong et al. 2001), CYP1A1 (Cytochrome P4501A1; Garte et al 2001), CYP2E1 (Cytochrome P4502E1; Garte et al 2001), GSTM (glutathione S-transferase M1; Garte et al 2001), FXIII (factor XIII; Attie-Castro et al., 2000), LPL (human lipoprotein lipase; Clark et al., 1998), HFE (hemochromatosis; Beutler and Gelbart 2000), FcgRIIIB (neutrophil antigens NA1 and NA2; Matsuo et al., 2000), P53 (protein 53; Shepherd et al., 2000), SFD1 (stromal cell derived factor 1; Rabkin et al., 1999), APO E (Apolipoprotein E; Pablos-Méndez et al., 1997,), ANGIO (Angiogenin; Rivera et al., 2001), MTHFR (5,10-methylenetetrahydrofolate reductase; Rosenberg et al., 2002), GPIIIa (glycoprotein IIIa; Goldschmidt-Clermont et al., 1999), GSTP1 (glutathione S-transferase P1; Watson et al., 1998), CAPN10 (calpain-10; Fullerton et al., 2002), CYP1A2 (Cytochrome P4501A2; Aitchison et al., 2000), LCT (Lactase; Hollox et al., 2001), OATP-C (Human organic ion transporting polypeptide-C; Tirona et al., 2001), IL4RA (interleukin 4-receptor alpha; Ober et al., 2000), AGT (angiotensinogen; Nakajima et al., 2002), MC1R (Melanocortin 1 receptor; Harding et al., 2000), COMT (catechol-O-methyltransferase; Palmatier et al., 1999), GSTT (glutathione S-transferase T1; Garte et al 2001).

TABLE 2. Values of Fst between various population pairs for four genes

Population	LEPR Fst	AHR Fst	FXIII Fst	CAPN10 Fst
Japanese/Amerinds	0.184 (+)			0.016 (−)
African/Amerinds	0.271 (R)	0.047 (−)	0.104 (−)	0.179 (R)
European/Amerinds	0.186 (R)	0.010 (−)	0.009 (−)	0.018 (−)
Chinese/Amerinds	0.010 (−)			0.034 (−)
N.Asian/Amerinds				0.074 (R)
Pygmies/Amerinds				0.448 (R)
African/North Asians		0.073 (−)		
Chinese/North Asians		0 (−)		0.064 (R)
European/North Asians		0 (−)		0.086 (R)
Japanese/North Asians				0.091 (+)
Pygmies/North Asians				0.188 (R)

between populations. Values of Fst below the 95% confidence limit, on the other hand, might result from balancing selection (especially for alleles at or near fixation), or could be due to chance.

It is clear from Table 1 that for any population pair no pattern of genetic diversity can be seen that cuts across different genes. For example, when comparing Europeans and Africans (the most commonly examined pair) there were 9 genes with no significant deviation in diversity from expected, 12 genes with decreased diversity compared to that expected, and 2 genes with a higher than expected Fst. For the genes examined from more than a single population pair, very few showed a consistent pattern. The MTHFR gene had Fst values not different from expected for all three population pair comparisons, and the AHR gene had Fst values less than expected for the three population pairs, as well as for pairing of Northern Asians with Africans, Chinese and Europeans (Table 2). The other genes showed highly inconsistent patterns, suggesting that different selective factors might be exercising locus-specific effects in populations with different geographic origin. For example, the LEPR gene exhibited lower than expected diversity between Africans and Europeans, higher than expected diversity between Asians and Europeans, but between Africans and Asians the value of Fst was within that expected. As shown in Table 2, this pattern for LEPR was also found when comparing Amerinds with Japanese (positive), or with Africans or Europeans (reference).

It is clear from the population pair comparisons that the majority of Fst values fall within or below the range expected from the reference set, as illustrated in Figure 1. Only 26 out of 114 (23%) values of Fst for specific genes in all population pair comparisons showed a higher than expected degree of genetic diversity. A rare example of strong positive selection was seen for the metabolic detoxification gene GSTT1 comparing Northern Europeans to other Europeans (Table 3). Diversity in the allele frequencies of this gene is quite high between these very similar populations, suggesting an unknown selective mechanism in operation.

For those genes whose allele frequencies were available for at least three populations representing different geographical origins around the world, the human Fst was calculated over all populations available, as shown in Table 4. The Fst values for the genes analyzed covered a wide range (0.038–0.374), as expected, with a mean of 0.125, very close

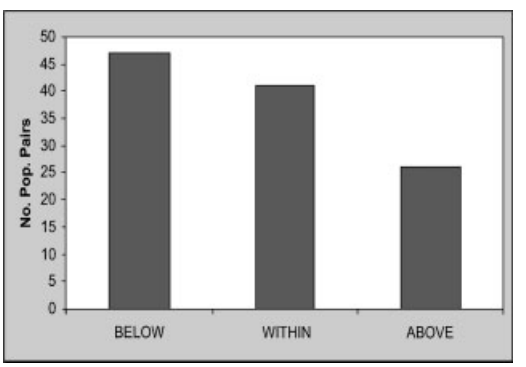

Fig. 1. Values of Fst compared to expected value for 114 population pair comparisons. Below, within, and above refer to the expected values for each population pair/gene combination. Total number of combinations was 114.

TABLE 3. Values of Fst and for pairwise comparisons of three European populations for GSTT1

	Observed Fst	Expected Fst	OBS/EXP
English/Danish	0.0209	0.0021	9.9
English/German	0.00031	0.0022	0.14
Danish/German	0.0160	0.0016	10.0

to the generally accepted value of Fst (0.1) for the human species (Relethford, 2001).

Analysis of two genes that could be considered to be under true positive selection provided interesting results. The ability of adults to tolerate lactose is due to a well-known polymorphism in the LCT gene found in Northern Europeans and people from Northern India. The Fst values between a number of populations confirm the positive selection, but only when comparing the specific populations in which this polymorphism has occurred. Another gene that could be considered under positive selection, the MC1R gene, involved in skin pigmentation, shows only limited evidence of such selection, and only between Africans and Europeans (Table 1). This result, while surprising, is in fact consistent with the conclusions of the article from which the data were extracted (Harding et al., 2000). These authors, using more sophisticated measures of selection and drift, also conclude that this particular gene (which is only one of several involved in skin pigmentation) does not show a degree of diversity consistent with positive selection.

DISCUSSION

The hypothesis put forward above has been clearly refuted by the results of the analysis of these genes among several populations. Instead of any pattern of increased diversity for biomedically important genes over that expected, a random pattern of diversity was observed, with large differences found even among single genes for different population pairs. Only a minority of these genes show any evidence of higher than expected diversity between populations which could be attributed to selective environmental or geographic influences. For many of the genes examined, much less diversity was observed than expected. For some genes, less diversity was observed for certain population pairs, and either equal or greater diversity for other pairwise comparisons.

TABLE 4. World wide Fst for 17 genes

Gene	Alleles	Populations	Fst
DRD2[a]	3	3	0.108
CCR5	8	3	0.101
HLA-A	12	5	0.062
AHR	2	5	0.038
LEPR	6	4	0.203
CYP3A4	1	3	0.374
UGT1A1	4	3	0.080
CYP1A1	4	3	0.064
GSTM1	2	3	0.062
CYP1A2	1	3	0.048
FXIII	1	4	0.162
FcgammaRIIIB	1	3	0.085
MTHFR	5	3	0.079
CAPN10	7	5	0.188
GSTP1	1	3	0.037
MC1R	8	3	0.144
LCT	1	11	0.285
MEAN (SEM)			0.125 ± 0.022

[a]DRD2 (D2 dopamine receptor; Gelernter et al., 1998). See Table 1 for gene abbreviations.

Several limitations of the approach taken here should be taken into consideration. The assessment of the degree of drift between specific populations is often a very rough estimate, based on fairly small sample sizes. Furthermore, the set of genes used by Cavalli-Sforza et al. (1994) to establish interpopulation genetic diversity includes some genes that could be influenced by selection (such as the immunoglobulin genes). It would have been preferable to use those genes which are known to be subject only to neutral drift, as has been done in other studies (Hamblin et al., 2002). However, such a set of genes is not available for the extensive worldwide population comparisons done by Cavalli-Sforza et al. (1994). Considering the fact that among the genes included in the control set, selection should not be biased in any particular direction, and the fact that most of the genes used (like most genes) are in fact under little or no selection, it is highly unlikely that the results of the analysis presented here have been greatly distorted by the use of this gene set for comparison. Indeed, the results are consistent with earlier estimates of the percentage of population specific contribution to human genetic diversity (Lewontin, 1972).

The determination of Fst values from data on allele frequencies presented in published reports usually does not allow for the statistical estimation of level of confidence. The allele frequencies presented here were often obtained from fairly small population samples, sometimes less than 100 individuals, and therefore may be inaccurate. From previous work on the determination of allele frequencies in very large population samples (Garte et al., 2001), it is clear that major errors can occur from small sample sizes. In an often-cited publication on population differences in over 300 genes, Stephens et al. (2001) used population samples of 20 individuals, from which they claimed to have found a number of ethnic-specific alleles. Clearly, such claims require far larger sample sizes to be substantiated. Allele frequencies reported to come from analysis of samples of less than 50 subjects were not used in this report, in an attempt to avoid such errors. Allele frequency estimation errors due to small sample sizes could result in the misclassification of a gene-specific Fst for some population pairs. However, it seems unlikely that any major change in the general conclusion would be necessary following correction of allele frequency data for particular loci and populations.

It should be noted that certain standard measures of natural selection such as Tajima's D (Wooding et al., 2002; Tajima, 1989) were not applied to these data, since such statistics have more value when examining the possible role of selection for a particular degree of heterogeneity within a population, and some require knowledge of haplotype, or larger sample sizes than were generally available from the literature sources used here.

Not all alleles produce changes in phenotype, and therefore it would not be expected that all alleles would be subject to selection pressure to the same extent. The majority of the allelic variants included in this analysis have clearly defined or indirectly established phenotypes. The only exceptions are CYP1A1, CYP2E1, LRP, and AHH. Unfortunately, the definition of phenotype is not always sufficiently clear to allow for a definitive decision to be made about the relevance of a particular variant to the effects of selection pressure. As an example, the CYP1A1*2A allele, which results from an SNP outside the coding region of the gene, has been widely assumed to be devoid of phenotypic consequences. However, several studies have shown this allele to be important in susceptibility to carcinogenesis (Vineis et al., 2003) and other effects (Garte et al., 2003). It is also far from clear that the inclusion of some possibly phenotypically neutral alleles has skewed the results toward neutrality. For GSTM1, where the variant genotype is complete gene deletion (with obvious loss of function phenotypic consequences), no significant departure from neutrality was found for the African/European pair, while between Asians and Africans the result was equivocal.

There is a debate around the issue of human categorization related to genetic profiles (Wilson et al., 2001; Goodman, 2000; Braun, 2002; Burchard et al., 2003; Cooper et al., 2003; Sankar and Cho, 2002). While some argue that such categorization is not justified by human genetic data, others point out that population differences do exist in many genes of biomedical importance. The analysis presented here does not directly address this debate, since differences due to drift are still differences. However, the fact that such differences appear to be lower or equal to (rather than

higher) than expected for many genes of biomedical importance could be used to argue that too much attention to the ethnic background or population origin of individual patients may not be warranted. A recent article has shown, for example, that even if one accepts continental geographic origin as a basis for making inferences about genetic diversity, it is not possible to accurately determine the continental origin of individuals based on such commonly used phenotypes as skin color or facial characteristics (Parra et al., 2003).

A review of the literature on human population genetic diversity reveals two broad categories of research studies. In articles devoted to population genetics, careful analysis of haplotypes, linkage, neutrality, selection, and other statistical measures of population structure are generally presented for one or more loci, often using large panels of well-defined population samples and sophisticated statistical tools of genetic analysis (Bonnen et al., 2000; Batzer et al., 1996; Knight et al., 1996; Goddard et al., 2000). On the other hand, a great deal of data on population diversity has been published in association studies dealing with a single locus, where the focus of the research is on pharmacogenetics, disease susceptibility, or other areas related to specific gene function in a biomedical context. In this latter category of publications population groups are sometimes ill-defined ("white" or "Hispanic") and allele frequencies of control populations are compared using chi-square or other simple statistical tests. Among the publications in this category is a large interdisciplinary pooled analysis study (led by the author) of metabolic gene allele frequencies in control populations (Garte et al., 2001). In this and many other articles (Sanghera et al., 1998; Chen et al., 2001; Beutler et al., 2000; Chagnon et al., 2000; Iso et al., 1996; Keshava et al., 2001; Tirona et al., 2001), population differences in allele frequencies are often discussed as being "large" or "significant," when in reality such differences were less than or not different from that expected between geographically distinct populations. It would be beneficial in general for biomedical researchers to perform a simple analysis to determine whether the population-based differences they detect in allele frequencies for their gene of interest are anything more (or in fact less) than that expected for all genes between those populations. As an example, the allele frequency of the GSTP EX5 allele was found to be 0.18, 0.33, and 0.42 in Asians, Europeans, and Africans, respectively, and these differences were statistically significant (Watson et al., 1998). The authors conclude that "The differences observed were highly significant . . . .[D]ifferences we observed between groups suggest the possibility of differences in susceptibility to electrophilic toxicants or effectiveness of drugs." In reality, the Fst values for this allele were 0.016 for European/African, 0.059 for Asian/European, and 0.14 for Asian/African pairs. The first two values are smaller than expected, while the third is within the expected range. While the conclusion by the authors quoted above might be warranted, it would be misleading to suggest that these significant differences point to a selection mechanism related to geographic factors or natural histories of populations, since such data probably correspond mostly to the result of genetic drift or other factors to be expected for any gene when comparing different populations.

While the assessment of population differences in allele frequencies of biomedically relevant genes is not incorrect, it is important to address the biological and medical meaning of these differences. Since the majority of these differences are less than, or the same as, that expected for any gene, it is not correct to presume that such differences reflect any meaningful or significant effects related to natural selection, including differential susceptibility to disease organisms, environmental exposures, or other specific geographical influences. The results of the analysis require a new hypothesis related to the degree of population-specific allele frequency differences for genes with biomedical significance. One possibility is that the emphasis in the literature on allele frequency differences between "racial" or ethnic groups in biomedically important genes might be due to a type of publication bias (biomedically relevant genes are more likely to be studied), coupled with a lack of awareness among many authors of human population genetics data concerning the expected levels of such differences.

Recent publications discussing the implications of racial or population differences in the frequency of certain alleles in genes of pharmacogenetic or susceptibility relevance have suggested that such variation needs to

be taken into account when designing treatment or prevention strategies (Kalow, 2001; Risch et al., 2002). It will probably remain true that for the majority of such susceptibility alleles, each individual will need to be genotyped regardless of apparent race. And consistent with work from many labs, the concept of race is of limited, if any, value in any field of genetic diagnosis. It is highly unlikely that specific genetic influences on treatment, diagnosis, toxicity, or other response will be predictable based on racial or ethnic identification alone (as determined by skin color, eye shape, self-identification, or in any other way). For almost all medicine–gene–disease situations, individual genotypes of patients will need to be assessed without regard to alleged race, in order to determine the patient's specific genetic profile.

LITERATURE CITED

Aitchison KJ, Gonzalez FJ, Quattrochi LC, Sapone A, Zhao JH, Zaher H, Elizondo G, Bryant C, Munro J, Collier DA, Makoffa AI, Kerwin RW. 2000. Identification of novel polymorphisms in the 5' flanking region of CYP1A2, characterization of interethnic variability, and investigation of their functional significance. Pharmacogenetics 10:695–704.

Attie-Castro FA, Zago MA, Lavinha J, Elion J, Rodriguez-Delfin L, Guerreiro JF, Franco RF. 2000. Ethnic heterogeneity of the factor XIII Val34Leu polymorphism. Thromb Haemost 84:601–603.

Ball SE, Scatina J, Kao J, Ferron GM, Fruncillo R, Mayer P, Weinryb I, Guida M, Hopkins PJ, Warner N, Hall J. 1999. Population distribution and effects on drug metabolism of a genetic variant in the 5' promoter region of CYP3A4. Clin Pharmacol Ther 66: 288–294.

Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS, Wooding S, Stone AC, Jorde LB, Weiss RB, Ahuja SK. 2002. A strong signature of balancing selection in the 5' *cis*-regulatory region of *CCR5*. Proc Natl Acad Sci USA 99:10539–10544.

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M. 1996. Genetic variation of recent Alu insertions in human populations. J Mol Evol 42:22–29.

Beutler E, Gelbart T. 2000. A common intron 3 mutation (IVS3 -48c→g) leads to misdiagnosis of the c.845G→A (C282Y) HFE gene mutation. Blood Cells Mol Dis 26:229–233.

Beutler E, Gelbart T, Demina A. 1998. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? Proc Natl Acad Sci USA 95: 8170–8174.

Bonnen PE, Story MD, Ashorn CL, Buchholz TA, Weil MM, Nelson DL. 2000. Haplotypes at ATM identify coding-sequence ariation and indicate a region of extensive linkage disequilibrium. Am J Hum Genet 67:1437–1451.

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839–843.

Braun L. 2002. Race, ethnicity, and health: can genetics explain disparities? Perspect Biol Med 45:159–174.

Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, Mountain JL, Perez-Stable EJ, Sheppard D, Risch N. 2003. The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med 348:1170–1175.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton, NJ: Princeton University Press.

Chagnon YC, Wilmore JH, Borecki IB, Gagnon J, Perusse L, Chagnon M, Collier GR, Leon AS, Skinner JS, Rao DC, Bouchard C. 2000. Associations between the leptin receptor gene and adiposity in middle-aged Caucasian males from the HERITAGE family study. J Clin Endocrinol Metab 85:29–34.

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE. 1992. Caucasian genes in American blacks: new data. Am J Hum Genet 50:145–155.

Chen W, Srinivasan SR, Elkasabany A, Ellsworth DL, Boerwinkle E, Berenson GS. 2001. Combined effects of endothelial nitric oxide synthase gene polymorphism (G894T) and insulin resistance status on blood pressure and familial risk of hypertension in young adults: the Bogalusa Heart Study. Am J Hypertens 14: 1046–1052.

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612.

Cooper RS, Kaufman JS, Ward R. 2003. Race and genomics. N Engl J Med 348:1166–1170.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. 2001. High-resolution haplotype structure in the human genome. Nat Genet 29:229–232.

Dawson DV, Ozgur M, Sari K, Ghanayem M, Kostyu DD. 2001. Ramifications of HLA class I polymorphism and population genetics for vaccine development. Genet Epidemiol Suppl 20:87–106.

Fullerton SM, Bartoszewicz A, Ybazeta G, Horikawa Y, Bell GI, Kidd KK, Cox NJ, Hudson RR, Di Rienzo A. 2002. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. Am J Hum Genet 70:1096–1106.

Garte S. 2002. The racial genetics paradox in biomedical research and public health. Public Health Rep 117: 421–425.

Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, Baranova H, Bathum L, Benhamou S, Boffetta P, Bouchardy C, Breskvar K, Brockmoller J, Cascorbi I, Clapper ML, Coutelle C, Daly A, Dell'Omo M, Dolzan V, Dresler CM, Fryer A, Haugen A, Hein DW, Hildesheim A, Hirvonen A, Hsieh LL, Ingelman-Sundberg M, Kalina I, Kang D, Kihara M, Kiyohara C, Kremers P, Philip Lazarus P, Le Marchand L, Lechner MC, van Lieshout EMM, London S, Manni JJ, Maugard CM, Morita S, Nazar-Stewart V, Noda K, Oda Y, Parl FF, Pastorelli R, Persson I, Peters WHM, Rannug A, Rebbeck T, Risch A, Roelandt L, Romkes M, Ryberg D, Salagovic J, Schoket B, Seidegard J, Shields PG, Sim E, Sinnet D, Strange RC, Stucker I, Sugimura H, To-Figueras J, Vineis P, Yu MC, Taioli E. 2001. Metabolic gene frequencies in control populations. Cancer Epidemiol Biomark Prevent 10:1239–1248.

Garte S, Ganguly S, Taioli E. 2003. Effect of genotype on steady state CYP1A1 gene expression in human peripheral lymphocytes. Biochem Pharmacol 65:441–445.
Gelernter J, Kranzler H, Cubells JF, Ichinose H, Nagatsu T. 1998. DRD2 allele frequencies and linkage disequilibria, including the -141CIns/Del promoter polymorphism, in European-American, African-American, and Japanese subjects. Genomics 51:21–26.
Goddard KA, Hopkins PJ, Hall JM, Witte JS. 2000. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. Am J Hum Genet 66:216–234.
Goldschmidt-Clermont PJ, Coleman LD, Pham YM, Cooke GE, Shear WS, Weiss EJ, Kral BG, Moy TF, Yook RM, Blumenthal RS, Becker DM, Becker LC, Bray PF. 1999. Higher prevalence of GPIIIa PlA2 polymorphism in siblings of patients with premature coronary heart disease. Arch Pathol Lab Med 123: 1223–1229.
Gonzalez E, Bamshad M, Sato N, Mummidi S, Dhanda R, Catano G, Cabrera S, McBride M, Cao XH, Merrill G, O'Connell P, Bowden DW, Freedman BI, Anderson SA, Walter EA, Evans JS, Stephan KT, Clark RA, Tyagi S, Ahuja SS, Dolan MJ, Ahuja SK. 1999. Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. Proc Natl Acad Sci USA 96:12004–12009.
Goodman AH. 2000. Why genes don't count (for racial differences in health). Am J Public Health 90: 1699–1702.
Halkas AC, Gaillard MC, Thomson PD, Green SL, Ludewick H, Kala U. 1998. Variants of alpha 1-proteinase inhibitor in black and white South African patients with focal glomerulosclerosis and minimal change nephrotic syndrome. J Med Genet 35:6–9.
Hamblin MT, Thompson EE, Di Rienzo A. 2002. Complex signatures of natural selection at the Duffy blood group locus. Am J Hum Genet 70:369–383.
Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL. 2000. Evidence for variable selective pressures at MC1R. Am J Hum Genet 66: 1351–1361.
Harris MR, Bunker CH, Hamman RF, Sanghera DK, Aston CE, Kamboh MI. 1998. Racial differences in the distribution of a low density lipoprotein receptor-related protein (LRP) polymorphism and its association with serum lipoprotein, lipid and apolipoprotein levels. Atherosclerosis 137:187–195.
Hartl DL, Clark AG. 1997. Principles of population genetics, 3rd ed. Sunderland, MA: Sinauer Associates.
Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM. 2001. Lactase haplotype diversity in the Old World. Am J Hum Genet 68:160–172.
Iso H, Harada S, Shimamoto T, Folsom AR, Koike K, Sato S, Iida M, Komachi Y. 1996. Polymorphism of the apolipoprotein B gene and blood lipid concentrations in Japanese and Caucasian population samples. Atherosclerosis 126:233–241.
Kalow W. 2001. Pharmacogenetics and population differences. Pharmacogenet Pharmacogenom Pharmacobiol 70:1–4.
Keshava C, McCanlies EC, Keshava N, Wolff MS, Weston A. 2001. Distribution of HER2 (V655) genotypes in breast cancer cases and controls in the United States. Cancer Lett 173:37–41.
Knight A, Batzer MA, Stoneking M, Tiwari HK, Scheer WD, Herrera RJ, Deininger PL. 1996. DNA sequences of Alu elements indicate a recent replacement of the human autosomal genetic complement. Proc Natl Acad Sci USA 93:4360–4364.
Lewontin RC. 1972. The apportionment of human diversity. Evol Biol 6:381–398.
Matsuo K, Procter J, Stroncek D. 2000. Variations in genes encoding neutrophil antigens NA1 and NA2. Transfusion 40:645–653.
Nakajima T, Jorde LB, Ishigami T, Umemura S, Emi M, Lalouel JM, Inoue I. 2002. Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations. Am J Hum Genet 70: 108–123.
Ober C, Leavitt SA, Tsalenko A, Howard TD, Hoki DM, Daniel R, Newman DL, Wu X, Parry R, Lester LA, Solway J, Blumenthal M, King RA, Xu J, Meyers DA, Bleecker ER, Cox NJ. 2000. Variation in the interleukin 4-receptor alpha gene confers susceptibility to asthma and atopy in ethnically diverse populations. Am J Hum Genet 66:517–526.
Pablos-Méndez A, Mayeux R, Ngai C, Shea S, Berglund L. 1997. Association of Apo E polymorphism with plasma lipid levels in a multiethnic elderly population arteriosclerosis. Thromb Vasc Biol 17:3534–3541.
Palmatier MA, Kang AM, Kidd KK. 1999. Global variation in the frequencies of functionally different catechol-O-methyltransferase alleles. Biol Psychiatry 46: 557–567.
Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD. 1998. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851.
Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. 2003. Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 100: 177–182.
Rabkin CS, Yang Q, Goedert JJ, Nguyen G, Mitsuya H, Sei S. 1999. Chemokine and chemokine receptor gene variants and risk of non-Hodgkin's lymphoma in human immunodeficiency virus-1-infected individuals. Blood 93:1838–1842.
Relethford JH. 2001. Genetics and the search for modern human origins. New York: Wiley-Liss.
Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393–398.
Risch N, Burchard E, Ziv E, Tang H. 2002. Categorization of humans in biomedical research: genes, race and disease. Genome Biol 3:1–12.
Rivera MA, Echegaray M, Rankinen T, Perusse L, Rice T, Gagnon J, Leon AS, Skinner JS, Wilmore JH, Rao DC, Bouchard C. 2001. TGF-beta(1) gene-race interactions for resting and exercise blood pressure in the HERITAGE Family Study. J Appl Physiol 91: 1808–1813.
Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12:602–612.
Rosenberg N, Murata M, Ikeda Y, Opare-Sem O, Zivelin A, Geffen E, Seligsohn U. 2002. The frequent 5,10-methylenetetrahydrofolate reductase C677T polymorphism is associated with a common haplotype in whites, Japanese, and Africans. Am J Hum Genet 70:758–762.
Sanghera DK, Saha N, Kamboh MI. 1998. The codon 55 polymorphism in the paraoxonase 1 gene is not associated with the risk of coronary heart disease in Asian Indians and Chinese. Atherosclerosis 136:217–223.
Sankar P, Cho MK. 2002. Toward a new vocabulary of human genetic variation. Science 298:1337–1338.
Schneider S, Roessli D, Excoffier L. 2000. Arlequin: a software for population genetics data analysis. v.2.000http://lgb.unige.ch/arlequin

Shepherd T, Tolbert D, Benedetti J, Macdonald J, Stemmermann G, Wiest J, DeVoe G, Miller MA, Wang J, Noffsinger A, Fenoglio-Preiser C. 2000. Alterations in exon 4 of the p53 gene in gastric carcinoma. Gastroenterology 118:1039–1044.

Smith MW, Lautenberger JA, Shin HD, Chretien JP, Shrestha S, Gilbert DA, O'Brien SJ. 2001. Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. Am J Hum Genet 69:1080–1094.

Steele FR. 2002. Genetic "differences." Genomics 79:145.

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. Science 293: 489–493.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tirona RG, Leake BF, Merino G, Kim RB. 2001. Polymorphisms in OATP-C: identification of multiple allelic variants associated with altered transport activity among European- and African-Americans. J Biol Chem 276:35669–35675.

Vineis P, Veglia F, Antilla S, Benhamou S, Butkievicz D, Cascorbi I, Clapper ML, Dolzan V, Haugen A, Hirvonen A, Ingelman-Sundberg M, Kihara M, Chikako Kiyohara C, Kremers P, Le Marchand L, Oshima S, Pastorelli R, Rannug A, Romkes M, Schoket B, Shields P, Strange RC, Stucker I, Sugimura H, Garte S, Gaspari L, Taioli E. 2003. Genetic modulation (CYP1A1) and duration of smoking in lung cancer: a pooled analysis of 2800 cases and 3700 controls. Int J Cancer 104:650–657.

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. Am J Hum Genet 68:738–752.

Watson MA, Stewart RK, Smith GB, Massey TE, Bell DA. 1998. Human glutathione S-transferase P1 polymorphisms: relationship to lung tissue enzyme activity and population frequency distribution. Carcinogenesis 19:275–280.

Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, Bradman N, Goldstein DB. 2001. Population genetic structure of variable drug response. Nat Genet 29:265–269.

Wong JM, Okey AB, Harper PA. 2001. Human aryl hydrocarbon receptor polymorphisms that result in loss of CYP1A1 induction. Biochem Biophys Res Commun 288:990–996.

Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, Jorde LB. 2002. DNA sequence variation in a 3.7-kb noncoding sequence 5' of the CYP1A2 gene: implications for human population history and natural selection. Am J Hum Genet 71: 528–542.