The Racial Genetics Paradox in Biomedical Research and Public Health

SEYMOUR GARTE, PHD[a]

While the concept of race is of obvious social, historical, and cultural importance in the United States, it remains very difficult to define. In fact, a strict biological definition of race among humans does not exist, and most useful definitions of race involve social or cultural parameters as well as referring to specific continental populations such as Europeans, Asians, and Africans. There are no good biological criteria on a phenotypic level to determine the race of any individual, or even to determine with any precision exactly how many races exist.

Can genetics be used to more accurately define race? The answer is no. On a genetic level, human variation is a smooth continuum with very little evidence for sharp racially defined heterogeneities. The availability of data on thousands of DNA polymorphisms from the various genome re-sequencing projects has clearly shown that the largest part of genetic variability within the human population is due to differences among individuals within populations, rather than to differences between populations. According to most researchers in the field, these results effectively discredit a genetic basis for the concept of "race."[1,2] This makes sense from the perspective of human migration and admixture over the past hundreds of thousands of years. Present definitions of race based on superficial characteristics, or on other phenotypes strongly influenced by natural selection, such as skin color, simply do not correlate with data from the whole genome.

However, even before the completion of the human DNA sequence, enough data had been generated on polymorphisms in certain genes to demonstrate that large differences exist in certain allele frequencies between the three major "racial" groups. The sequence of the human genome may eventually lead to genetic definitions of human potential, disease risk, and so on, that could be linked in some cases with race-specific genotypes. For example, certain drug-metabolizing enzymes may have race-specific polymorphic alleles that confer an important phenotype (such as the inability to tolerate certain foods, alcohol, medicines, or other exposures), although it should be stressed that in no case are such alleles present in all members of any given group.

[a]Environmental and Occupational Health Sciences Institute, University of Medicine and Dentistry of New Jersey, Piscataway, NJ, and Genetics Research Institute, Milan, Italy

Address correspondence to: Seymour Garte, PhD, EOHSI Toxicology, 170 Frelinghuysen Rd., Piscataway, NJ 08854; tel. 732-445-0114; fax 732-445-0119; e-mail <garte@eohsi.rutgers.edu>.

RACIAL DIFFERENCES IN ALLELE FREQUENCIES

Significant differences in allele frequencies for the three major racial groups have been found for the metabolic genes involved in cancer susceptibility such as the Cytochrome P450s (CYP), N Acetyltransferase (NAT), and glutathione S transferase (GST) families.[3] At codon 72 of the p53 gene, 64% of whites had the arg/arg genotype, compared with 24% of blacks.[4] The 7/7 allele of UGT 1A1*28 was found in 11% of "Caucasians," but was not detected (<4%) in Asians.[5]

The frequency of the P-glycoprotein C/C genotype, which leads to overexpression of the protein product, was 83% in West Africans, 61% in African Americans, 26% in "Caucasians," and 34% in Japanese Americans.[6] This finding might have negative consequences for the clinical efficacy of protease inhibitors used to combat AIDS in people of African descent. The 5′ regulatory region of the gene for corticotropin-releasing hormone (CRH) plays a crucial role in stress response. The DNA sequence of this locus is highly conserved, meaning that few differences are seen between species. In a study of the distribution of CRH alleles A1 and A2, it was found that the gene frequencies for these alleles were extremely divergent between African and white populations. The African A1 frequency determined from three different populations ranged from 0.27 to 0.3, while the frequency in the white group was 0.9. Calculation of the value for genetic heterogeneity for this gene gave 0.612, a value that is seen only for loci under strong natural selection.[7] The basis for this selection is not yet clear, since the functional relevance of the two alleles is not yet known. Other genes exhibit similar differences. It should be noted that some of the cases of putative racial differences in allele frequencies mentioned above are preliminary and based on relatively small numbers of subjects (see below). Furthermore, very few if any of these allelic frequency differences have been convincingly linked to differences in disease risk or other outcomes.

EVIDENCE AGAINST A GENETIC DEFINITION OF RACE

While genetic diversity in certain genes as described above is large for the human population, exactly the opposite conclusion has been reached by population biologists. When populations are sampled for a variety of genetic markers, the lines between ethnic groups become blurred. In a study on 257 loci from each of the chromosomes, all variants were found in "Caucasians," African Americans, Chinese Americans, and Native Americans. Although there were some differences in allele frequencies, no ethnic-specific alleles were found.[8] The Y chromosome exhibited only three SNPs that could be attributed to ethnicity among Italians, Melanesians, Amazonian Indians, Africans, and Pygmies.[9]

Data from the landmark work of Cavalli-Sforza and his colleagues[10] and from the work of Relethford[11] suggest that Africans and Europeans share enough gene frequencies to enable them to be lumped together in a single group, as compared, for example, with American Indians or Australian Aborigines.[10] In other work, Mountain and Cavalli-Sforza found that genetic differences between any two Italians were five times as great as the difference between an Italian and a Japanese, African, or New Guinean.[12] In a recent report, Wilson et al. demonstrated that people from eight widely separated geographic and ethnic identities could be categorized into one of four genetic clusters.[13] While considerable overlap with ethnicity was found, each cluster also included some individuals of every ethnic group. The cluster containing most of the "Caucasians" also included most of the Ethiopians, about a quarter of the Afro-Caribbeans, and 9% of the Chinese. The authors further demonstrated that polymorphic variants of drug-metabolizing genes were associated more strongly with genetic clusters than with ethnicity. However, neither classification system was good enough to avoid the necessity of genotyping individuals for specific variants.

Even more revealing are the results when populations are further subdivided beyond the three largest "racial" groups defined by geographic origin. We observed large differences in the frequency of the "African-specific" CYP1A1 *3 allele among several tribal groups from Mali.[14] We have also reported on a significant difference in the frequency of the GSTT1 allele between Scandinavian and other European populations.[3]

THE PARADOX RESOLVED

Not all genes behave the same way with respect to allelic variation within the human population. For the great majority of genes, there has been little direct selection pressure since the origin of the human species, and allelic variants occur largely by genetic drift, a very slow process that in human beings can only be observed in populations that have been greatly isolated for many millennia. For some genes, on the other hand, the environment may exert a strong selective pressure that can result in a marked advantage for one allele over another. Since humans inhabit a large variety of geographical niches, people living in differ-

ent environments would be subject to different selection pressures, and therefore, for these genes, allele differences between populations from distinct geographical areas are to be expected. In this context it is useful to remember that members of a single "race" (as defined by continent of origin) may be subject to diverse environmental selection pressures. An interesting example is the geographic distribution of the duplicated form of the CYP2D6 gene, which codes for an enzyme that can detoxify many drugs and toxic compounds, including certain plant alkaloids. A high percentage of Ethiopians have multiple copies of this gene, possibly because of the selective advantage of being able to eat otherwise poisonous plants during periods of drought or famine. In West Africa, the frequencies of these alleles are about 1% to 3%, similar to most of the rest of the world, except for parts of southern Europe and Turkey, where the frequency is about 10%.[15] Clearly in this case, neither phenotypic race nor continent of origin is a useful parameter for predicting genotype.

The genes that fall into the group that are under high selection pressure include those related to all the surface characteristics that have been historically used to define race such as skin color, eye shape and color, body dimensions, facial structure, etc. The reasons for these human differences are fairly obvious, and have to do with the strong advantage of certain such characteristics in specific climates and conditions, such as white skin in cold dark regions and black skin in hot sunny regions. In addition, there are other genes whose allele frequencies may differ between populations because of selection pressure. Differences in the presence of infectious organisms (such as malaria-bearing mosquitoes), diet, and certain other exposures can affect allele selection for genes involved with metabolism, detoxification, pest resistance, and immune defense. One of the longest studied of these is the existence of the sickle cell allele in hemoglobin (which leads to protection against malaria in heterozygotes) in areas where malaria has been endemic. These areas include large parts of Africa as well as certain Mediterranean regions. As a result, sickle cell anemia, often thought of as an "African-specific" genetic condition, also affects "white" people with ancestry from certain areas of Greece, Italy, and the Arabian Peninsula.[16]

Among the genes that belong to the high selection category are many polymorphic genes of pharmacogenetic interest. Apparent ethnic specificities of several medically important metabolic traits have been known for many years, even before the genetic bases of these differences had been elucidated. These include glucose-6-phosphate dehydrogenase deficiency which mainly affects people of African, Mediterranean, and Asian descent, and the greater sensitivity of the Japanese population to alcohol compared with that of the white population due to differences in the ADH genes.[2]

Thus, the paradox referred to above may be resolved into two statements: For population genetics using representative genes of the human genome, genetic differences do not exist for different races, and the statement that race is not of any biological significance is correct. However, for a particular subset of genes, many with important biomedical function and significance, average allele frequency differences are in fact observed between populations originating in different geographic areas, or with different exposures, diets, or other factors. In many (but by no means all) of these cases, the allele frequency differences segregate with alleles that are responsible for the surface characteristics historically used to define race. However, skin color is not a genetic marker for disease susceptibility or any other biomedically relevant condition.

RACE AND GENETICS—A MATTER FOR URGENT ATTENTION

A recent paper published in *Science* analyzed 313 genes for 20 individuals from each of four ethnic groups—"Caucasians," African Americans, Asians, and Latinos.[17] This effort has provided a picture of genetic variation among and between population groups. However, there is no attempt in this paper to categorize genes or alleles on the basis of genetic diversity or possible selective pressure implications. Since the sizes of the groups examined were relatively small, it wasn't possible to do any more thorough analysis of geographic origin than the broad, ill-defined ethnic categories that are generally used. Another drawback of the small size of each population examined (20 individuals) is that allele frequencies less than 5% could not be detected. The authors present data on unique alleles, but such conclusions are not warranted unless one has confidence that the population examined is large enough to allow detection of alleles present at a frequency of 1% (the definition of a polymorphism). This usually requires around 100–200 people. Given the current feasibility of such large-scale sequencing efforts, and the importance of the results for pharmacogenetics and other biomedical applications, it appears likely that this work represents only the beginning of large amounts of such data generated by many laboratories.

Some doctors have decided not to treat black chronic heart failure patients with an ACE inhibitor drug[18] because of a report that such drugs worked less

well in patients who described themselves as black than in patients who described themselves as white.[19] Based on these and similar types of studies, pharmaceutical companies are considering the development of ethnic-specific medicines, as a sub-branch of individualized medicine or pharmacogenetics. In reality, 19% of the African American patients tested did respond to ACE inhibitor therapy, and 51% of the whites did not,[18] meaning that even in this study showing a difference in frequency of response, race cannot be used as a surrogate for the genetic marker of drug response.

There is a real potential for discrimination in health care based on the inappropriate use of race as a genetic marker. For example, if research results show that the frequency of a high-risk allele is higher among people of African than of European descent, it is possible that a stigmatization process, associated with overt and covert racism, could come into play related to "race" and high risk. As stated above, the real situation is much more likely to be some but not complete overlap of any allele with phenotypic race, and therefore it would be unfair and prejudicial to label people as high risk based on skin color.

ALTERNATIVES TO RACE

A number of scientists, social scientists, and philosophers have raised the legitimate question of why we need to divide the human population at all. Since there seems to be little biological basis to do so, perhaps it would be better to avoid any method of categorization of humans. Of course, people do tend to divide themselves, usually along cultural definitions of ethnicity or group identity, but the argument runs that this is no reason for a scientific effort at subdivision based on genetic or any other biological criteria. However, the application of the results of the Human Genome Project, both in terms of finding targets for drugs related to a variety of important diseases and in terms of finding markers of susceptibility to these diseases, will rely more and more heavily on a precise understanding of the population frequencies of specific alleles. As discussed above, we know that population differences exist for many of the genes with such biomedical importance. While it may be technically correct to simply not categorize anyone, in the real world, categorization will probably not go away, for a variety of reasons. The issue remains as to how to define these population differences without using the convenient but inaccurate and potentially harmful categories of race as they have been used historically.

If possible, it would be desirable to find alternative ways to label populations with specific allele frequencies of genes important in pharmacogenetics, metabolism, and disease susceptibility. The convenient division of the human population according to superficial characteristics such as skin color and eye shape only roughly approximates a classification that would be optimally useful for purposes of public health related to gene-environment interaction in specific populations. In other words, while it might be true that a particular high risk allele is more frequent in, say, people of African descent than in people of European descent, it could easily turn out that only one group of Africans (e.g., from West Africa) actually carry this allele, and that certain European populations also have a high frequency of the allele. This situation is far more likely to be the rule rather than the exception.

One potentially useful alternative to race is the idea of geographic origin. It is, after all, geography that exerts a great deal of the selective pressure leading to selection of new or variant alleles. Climate is one of the most important and obvious of such environmental factors. Others include diet, the presence of disease organisms or toxic substances, and lifestyle factors directly or indirectly related to the environment. Geographical definitions of populations should be as precise as possible. A high frequency allele discovered in African Americans, for example, should be studied in as many African populations as possible. Ultimately it would be useful to produce a map defining the frequencies of major variant alleles according to precise geographic location. Of course, geography alone may not be a much better surrogate than race, especially when one considers population migrations, admixture, and difficulties in tracing ancestral lineages.

Concepts of race are closely associated with feelings of identification, self-worth, and so on. The ultimate long-term goal in the field of population genetic diversity should be to open the debate on whether new, scientifically sound, socially acceptable, and culturally logical methods to divide the human population should be devised, when and if such division makes sense on medical or public health grounds. The issue of race definition in the post-genome era will not go away, and by beginning to address this issue now, we may be in a position to help avoid the potential for future confusion, uncertainty, anxiety, and mistrust on the part of many communities and individuals toward modern research into the human genome and its interaction with the human environment.

REFERENCES

1. Cann HM. Human genome diversity. *Comptes Rendus de l'Academie des Sciences. Serie III: Sciences de la Vie* 1998;32: 443-6.
2. Weber WW. Populations and genetic polymorphisms. Mol Diagn 1999;4:299-307.
3. Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, et al. Metabolic gene frequencies in control populations. Cancer Epidemiol Biomarkers Prev 2001;10:1239-48.
4. Shepherd T, Tolbert D, Benedetti J, Macdonald J, Stemmermann G, Wiest J, et al. Alterations in exon 4 of the p53 gene in gastric carcinoma. Gastroenterology 2000; 118:1039-44.
5. Lampe JW, Bigler J, Horner NK, Potter JD. UDP-glucuronosyltransferase (UGT1A1*28 and UGT1A6*2) polymorphisms in Caucasians and Asians: relationships to serum bilirubin concentrations. Pharmacogenetics 1999;9:341-9.
6. Schaeffeler E, Eichelbaum M, Brinkmann U, Penger A, Asante-Poku S, Zanger UM, Schwab M. Frequency of C3435T polymorphism of MDR1 gene in African people. Lancet 2001;358:383-4.
7. Baerwald CG, Mok CC, Fife MS, Tikly M, Lau CS, Wordsworth BP, et al. Distribution of corticotropin-releasing hormone promoter polymorphism in different ethnic groups: evidence for natural selection in human populations. Immunogenetics 1999;49:894-9.
8. Dean M, Stephens JC, Winkler C, Lomb DA, Ramsburg M, Boaze R, et al. Polymorphic admixture typing in human ethnic populations. Am J Hum Genet 1994;55: 788-808.
9. Whitfield LS, Sulston JE, Goodfellow PN. Sequence variation of the human Y chromosome. Nature 1995;378:379-80.
10. Cavalli-Sforza LL, Menozzi P, Piazza A. The history and geography of human genes. Princeton: Princeton University Press; 1994.
11. Relethford JH. Genetics and the search for modern human origins. New York: Wiley-Liss; 2001.
12. Mountain JL, Cavalli-Sforza LL. Multilocus genotypes, a tree of individuals, and human evolutionary history. Am J Hum Genet 1997;61:705-18.
13. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, et al. Population genetic structure of variable drug response. Nat Genet 2001;29:265-9.
14. Garte S. The role of ethnicity in cancer susceptibility gene polymorphisms: the example of CYP1A1. Carcinogenesis 1998;19:1329-32.
15. Ingelman-Sundberg M. Genetic variability in susceptibility and response to toxicants. Toxicol Lett 2001;120: 259-68.
16. Dyson SM. "Race," ethnicity and haemoglobin disorders. Soc Sci Med 1998;47:121-31.
17. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, et al. Haplotype variation and linkage disequilibrium in 313 human genes. Science 2001; 293:489-93.
18. Masoudi FA, Havranek EP. Race and responsiveness to drugs for heart failure. N Engl J Med 2001;345:767.
19. Exner DV, Dries DL, Domanski MJ, Cohn JN. Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. N Engl J Med 2001;344: 1351-7.