

## SOUNDING BOARD

## Race and Genomics

Richard S. Cooper, M.D., Jay S. Kaufman, Ph.D., and Ryk Ward, Ph.D.

Race is a thoroughly contentious topic, as one might expect of an idea that intrudes on the everyday life of so many people. The modern concept of race grew out of the experience of Europeans in naming and organizing the populations encountered in the rapid expansion of their empires.<sup>1</sup> As a way to categorize humans, race has since come to take on a wide range of meanings, mixing social and biologic ingredients in varied proportions. This plasticity has made it a tool that fits equally well in the hands of demagogues who want to justify genocide and eugenics and of health scientists who want to improve surveillance for disease. It is not surprising, therefore, that diametrically opposing views have been voiced about its scientific and social value.<sup>2,3</sup> Indeed, few other concepts used in the conduct of ordinary science are the subject of a passionate debate about whether they actually exist.

Into this storm of controversy rides genomics. With the acknowledgment that race is the product of a marriage of social and biologic influences, it has been proposed that genomics now at least offers the opportunity to put its biologic claims to an objective test.<sup>4</sup> If those claims are validated, race will become a way to choose drug therapy for patients, categorize persons for genetic research, and understand the causes of disease. Genomics, with its technological innovations and authority as “big science,” might thereby solve the conundrum of race and bring peace to the warring factions.

Of the many implications that flow from the claim that race categorizes humans, few have more immediate clinical relevance than the choice of drug therapy. Remarkable progress has been made with drugs for cardiovascular disease, and the rapidly evolving patterns of care now include, for the first time, a candidate for “race-specific” therapy.<sup>5</sup> Randomized trials have been interpreted to show that a combination of vasodilators is more effective in treating heart failure in black persons than in white persons<sup>6</sup> and that angiotensin-converting-enzyme (ACE) inhibitors have little efficacy in blacks.<sup>7</sup> The results of the vasodilator trials were inconsistent,

however, and never achieved statistical significance for an interaction between treatment and race, the outcome of interest. Moreover, another analysis of the data from the aforementioned ACE-inhibitor trial<sup>7</sup> demonstrated that the original result showing a racial difference was unique to the end point that was chosen; in the portion of the study focusing on prevention, the drug had equal efficacy in blacks and in whites in reducing the incidence of the combined end point of death or development of a new onset of heart failure.<sup>8</sup> Similarly, no interaction with race was observed for the relative benefit of ACE inhibitors in preventing heart failure in the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial, supporting the view that the original observation of an effect of race was a type I error.<sup>9</sup>

Promotion of a drug for a race-specific “niche market” could distract physicians from therapies for which unequivocal evidence of benefit already exists. Race-specific therapy draws its rationale from the presumption that the frequencies of genetic variants influencing the efficacy of the drug are substantially different among races. This result is hard to demonstrate for any class of drugs, including those used to treat heart failure. Although a study of polymorphisms in drug-metabolizing enzymes did, in fact, show statistically significant variation in allele frequencies according to race, neither racial categories nor genetic clusters were sufficiently precise to make them clinically useful in guiding the choice of drugs.<sup>10</sup> What is lost in these arguments is the difficulty of translating differences among groups into a test that has adequate predictive value to help with clinical decisions. Race can help to target screening for a disease-associated mutation that is present at a high frequency in one population and is virtually absent in another,<sup>11</sup> but it is impossible for race as we recognize it clinically to provide both perfect sensitivity and specificity for the presence of a DNA-sequence variant. For this reason, race has never been shown to be an adequate proxy for use in choosing a drug; if you really need

to know whether a patient has a particular genotype, you will have to do the test to find out.

The availability of high-throughput genotyping creates the opportunity for increasingly sophisticated analyses of the extent to which continental populations vary genetically. Analysis of a large set of multiallelic microsatellite loci has shown that it is possible to cluster persons into population groups with high statistical accuracy.<sup>12</sup> Although clustering persons according to geographic origin has been accomplished most effectively with the use of highly informative, rapidly mutating, microsatellite loci, the use of single-nucleotide polymorphisms (SNPs) or their corresponding haplotypes also results in some degree of classification according to continent.<sup>13</sup>

However, the public health relevance of these data remains controversial. One view holds that the ability to categorize persons according to continental "race" validates the clinical and epidemiologic use of self-reported racial ancestry in terms of the categories of white, black, Asian, Pacific Islander, and Native American used by the U.S. Census.<sup>2</sup> We disagree. The success of microsatellite loci in classifying persons according to continental group depends in part on the cumulative effect of minor differences in the frequencies of common alleles and in part on the effect of population-specific alleles. In neither case is it apparent that such differences have relevance for traits that are important to health. Most population-specific microsatellite alleles are unlikely to be functional; rather, like a last name, they merely help to verify the geographic origin of a person's ancestry. Accumulated small differences in common alleles will yield differences in population risk only if a disease is caused primarily by interactions among multiple loci, and this is both mathematically and biologically implausible.

The same points apply to noncoding SNPs. In addition, coding sequences are highly conserved across groups. Moreover, in cases in which common polymorphisms occur, they are old and tend to be shared.<sup>14</sup> Categorizing people on the basis of differences in allele frequencies is therefore not the same as apportioning the whole of human diversity into medically relevant categories. The more relevant outcome — that the sets of common functional polymorphisms are distributed in discrete racial categories — has not been demonstrated. Furthermore, most population geneticists concur that the bulk of genetic variation (90 to 95 percent) occurs within, not among, continental populations.<sup>12-16</sup>

The central observations remain: variation is continuous and discordant with race, systematic variation according to continent is very limited, and there is no evidence that the units of interest for medical genetics correspond to what we call races.

The real effect of the biologic concept of race has always been its implications for common quantitative traits. Marked differences in the rates of cardiovascular diseases, for example, have been held up as examples of how race matters.<sup>17-20</sup> Reframed in genomic terms, it is argued that if "biological is defined by susceptibility to, and natural history of, a chronic disease, then . . . numerous studies . . . have documented biological differences among the races."<sup>2</sup> However, there is no body of evidence to support these broad claims about chronic diseases. Although it is obvious that many genetic diseases vary markedly among populations, those conditions are generally rare. Tay-Sachs disease, cystic fibrosis, and hemoglobinopathies, for example, are absent in many populations but present in others. But for these conditions, continental populations are not the categories of interest: persons of Jewish descent, not "whites," share a risk of Tay-Sachs disease; the frequency of cystic fibrosis varies widely within Europe; and thalassemia occurs in a variety of populations distributed from Italy to Thailand.

Many single-gene disorders have now been defined at the molecular level, and the emerging challenge faced by geneticists is to "make the genome relevant to public health."<sup>21</sup> Defining the molecular underpinnings of common chronic diseases has therefore become the central focus of genetic epidemiology. By extension, some investigators have turned with renewed enthusiasm to race as a tool for categorizing population risk. This approach draws on the practice, of long standing in the public health field in the United States, of granting priority to race or ethnic background as a demographic category — a surveillance practice, it is worth noting, that is virtually unique in the world. At the present time, however, very little is known about the genetic component of diseases of complex causation. Few, if any, well-characterized susceptibility genes have been identified for any of the degenerative conditions that kill at least 5 percent of the population, and we do not even know whether the individual variants are common or rare or whether they affect a protein's structure or its level of expression.

Since we do not know about the genetic variants that predispose persons to common chronic diseases, one might assume that arguments for the exist-

ence of genetic predispositions would be made for all population groups equally. The reality is very different. Minority groups, particularly blacks in the United States, are assumed to be genetically predisposed to virtually all common chronic diseases.<sup>17-20,22-26</sup> Genes are regularly proposed as the cause when no genetic data have been obtained, and the social and biologic factors remain hopelessly confounded.<sup>23-26</sup> Even when molecular data are collected, causal arguments are based on nonsignificant findings or genetic variation that does not have an established association with the disease being studied.<sup>19,20,22</sup> Coincidence is not a plausible explanation of the widespread occurrence of this practice over time and across subdisciplines. The correlation between the use of unsupported genetic inferences and the social standing of a group is glaring evidence of bias and demonstrates how race is used both to categorize and to rank order subpopulations.<sup>27,28</sup>

Not only are the relevant genetic data absent, but the distribution of polygenic phenotypes does not suggest that race is a useful category. Consider as an example height, a continuous trait that is highly heritable in all populations. Does continental race tell us something useful about average height? People who attain both the tallest stature (the Masai) and the shortest (the Biaka) are found in sub-Saharan Africa; Swedish people have traditionally been much taller than Sicilians; and although Japanese people used to be short, the current generation of children in Japan cannot fit in the desks in schools. The concept of race does not summarize this information effectively. If that complexity is multiplied by thousands of traits, which are randomly distributed among groups within continents, one gets an idea of the limitations of race as a classification scheme.

Although the rapid pace of change in genomics makes today's conclusions obsolete tomorrow, some predictions are in order. We can expect genomics increasingly to negate the old-fashioned concept that differences in genetic susceptibility to common diseases are racially distributed. In any common disease, many genes are likely to be involved, and each gene will have many variants. All the current data indicate that susceptibility alleles tend to be old, have moderate-to-small effects, and are shared among many populations. The APOE  $\epsilon 4$  allele, a well-studied example that contributes to a small extent to individual and potential risk for traits such as heart disease and dementia, is found in virtually all populations, albeit at varying rates.

Recent genomic surveys have also shown that as few as three to five common haplotypes capture the bulk of segregating variation at any specific locus throughout the genome, and those haplotypes are generally represented in the populations of all continents (Fig. 1).<sup>29</sup> Therefore, if susceptibility alleles for chronic diseases are located on common haplotypes, those alleles must be shared by members of all populations. Measuring the net effect of these genetic influences in a given population will require summing the frequencies of these susceptibility alleles in all genomic regions, while taking into account the environmental factors that are either difficult to measure or wholly unknown. Given these daunting epidemiologic challenges, it will be very difficult to calculate the "genetic susceptibility score" for any particular racial category.

This point requires further attention. There is no doubt that there are some important biologic differences among populations, and molecular techniques can help to define what those differences are. Some traits, such as skin color, vary in a strikingly systematic pattern. The inference does not follow, however, that genetic variation among human populations falls into racial categories or that race, as we currently define it, provides an effective system for summarizing that variation. The confused nature of this debate is apparent when we recognize that although everyone, from geneticists to laypersons, tends to use "race" as if it were a scientific category; with rare exceptions,<sup>15</sup> no one offers a quantifiable definition of what a race is in genetic terms. The free-floating debate that results, while entertaining, has little chance of advancing this field.

What is at stake is a more practical question — namely, has genomics provided evidence that race can act as a surrogate for genetic constitution in medicine or public health? Our answer is no. Race, at the continental level, has not been shown to provide a useful categorization of genetic information about the response to drugs, diagnosis, or causes of disease.

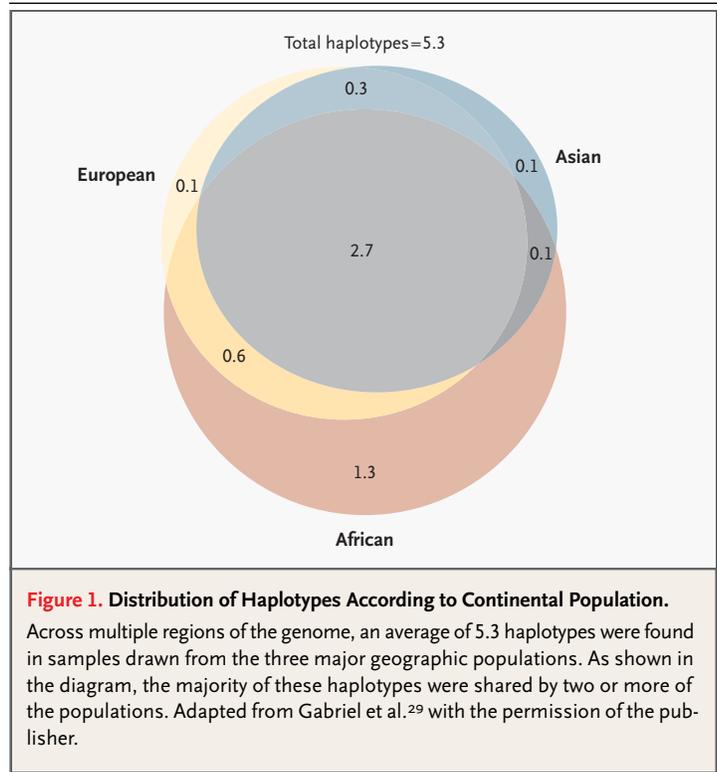
But in the United States, there is substantial variation in health status among major population subgroups. This self-evident truth has been the driving force behind the use of racial or ethnic categories in surveillance for disease. Among persons who are less convinced by the genetic data, variation in environmental exposure is seen as the cause of this phenomenon, and it follows that differences in health occur because privilege and power are unequal in racially stratified societies. The globalization of com-

plex chronic diseases seems to confirm the view that all populations are susceptible and that variation in rates can be understood as the result of differential exposure to environmental causes.

Although we acknowledge the salience of these arguments, the value of continental race as a classification scheme must be questioned in this context much as it was in the context of genetics. For example, persons who could be classified as having “African ancestry” have wide variation in rates of hypertension and diabetes, as do all large continental populations.<sup>30,31</sup> Without the context provided by such variables as the level of education, occupation, type of diet, and place of residence, race as a social category is not a useful predictor of health outcomes. Just as most genetic heterogeneity occurs within populations, there is enormous variation in the patterns of culture-derived behavioral and risk factors. An unintended result of categorizing people according to race can be to foreclose the question of why they have ill health, leaving us blind to the meaning of the more relevant local and individual context.

Race, in the metaphor introduced above, is the product of an arranged marriage between the social and biologic worlds. Although it often seems to travel back and forth between these parallel universes, it maintains a home in both. From the social sphere, race has inherited certain attributes that cannot be alienated from its meaning, no matter how hard we might try. The concept of race has currency in everyday discourse and is an epistemological category independent of the action of geneticists. From the beginning, it has been used not just to organize populations, but to create a classification scheme that explains the meaning inherent in the social order, according to which some groups dominate others.<sup>1,28</sup> There is a tendency for scientists to ignore the messy social implications of what they do. At the extreme, the argument is made that “we just tell the truth about nature,” and its negative consequences are political problems that do not concern us. Whether or not such a position is defensible from an ethical point of view, the debate over race cannot be sidestepped so easily. Race already has a meaning. To invoke the authority of genomic science in the debate over the value of race as a category of nature is to accept the social meaning as well.

In the 20th century, physics promised us knowledge of how the universe works, space travel, and the ability to harness the atom as an infinite source of energy. Although vast amounts of knowledge did



flow from research in those areas, the consequences in practice were not always benign. The accumulated record of peaceful and nonpeaceful atomic energy subsequently led many physicists to understand more fully that science is a part of society. In this century, biology — especially genomics — has emerged as the beacon of science leading us into the future, where data on the genetic sequence will unlock the secret of life. For genomics to fall in lock step with the socially defined use of race is not a propitious beginning to that journey. The ability to catalogue molecular variants in persons and populations has thrust genetics into a new relationship with society. Interpreting that catalogue within the existing framework of race, as was done in the case of eugenics, violates the principles that give science its unique status as a force outside the social hierarchy, one that does not take sides in factional contests. Racial affiliation draws on deep emotions about group identity and the importance of belonging. The discovery that races exist is not an advance of genomic science into uncharted territory; it is an extension of the atavistic belief that human populations are not just organized, but ordered.

From the Department of Preventive Medicine and Epidemiology, Loyola Stritch School of Medicine, Maywood, Ill. (R.S.C.); the Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill (J.S.K.); and the Department of Biological Anthropology, Oxford University, Oxford, United Kingdom (R.W.). Address reprint requests to Dr. Cooper at the Department of Preventive Medicine and Epidemiology, Loyola Stritch School of Medicine, 2160 S. First Ave., Maywood, IL 60153, or at rcooper@lumc.edu.

This article is dedicated to the memory of Dr. Ryk Ward, whose contributions to the genetics community will long be remembered.

1. Montagu A, ed. *The concept of race*. New York: Free Press of Glencoe, 1964.
2. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol* 2002; 3(7):comment2007.1–comment2007.12.
3. Muntaner C, Nieto FJ, O'Campo P. The Bell Curve: on race, social class, and epidemiologic research. *Am J Epidemiol* 1996;144:531-6.
4. Wade N. Race is seen as real guide to track roots of disease. *New York Times*. July 30, 2002:F1.
5. Franciosa JA, Taylor AL, Cohn JN, et al. African-American Heart Failure Trial (A-HeFT): rationale, design, and methodology. *J Card Fail* 2002;8:128-35.
6. Carson P, Ziesche S, Johnson G, Cohn JN. Racial differences in response to therapy for heart failure: analysis of the vasodilator-heart failure trials. *J Card Fail* 1999;5:178-87.
7. 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.
8. Dries DL, Strong M, Cooper RS, Drazner MH. Efficacy of angiotensin-converting enzyme inhibition in reducing progression of asymptomatic left ventricular dysfunction to symptomatic heart failure in black and white patients. *J Am Coll Cardiol* 2002;40:311-7. [Erratum, *J Am Coll Cardiol* 2002;40:1019.]
9. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002;288:2981-97. [Erratum, *JAMA* 2003;289:178.]
10. Wilson JE, Weale ME, Smith AC, et al. Population genetic structure of variable drug response. *Nat Genet* 2001;29:265-9.
11. Splawski I, Timothy KW, Tateyama M, et al. Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. *Science* 2002;297:1333-6.
12. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science* 2002;298:2381-5.
13. Stephens JC, Schneider JA, Tanguay DA, et al. Haplotype varia-

tion and linkage disequilibrium in 313 human genes. *Science* 2001; 293:489-93. [Erratum, *Science* 2001;293:1048.]

14. Halushka MK, Fan J-B, Bentley K, et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood pressure homeostasis. *Nat Genet* 1999;22:239-47.
15. Templeton AR. Human races: a genetic and evolutionary perspective. *Am Anthropol* 1998;100:632-50.
16. Romualdi C, Balding D, Nasidze IS, et al. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 2002;12:602-12.
17. Freedman BI. End-stage renal failure in African Americans: insights in kidney disease susceptibility. *Nephrol Dial Transplant* 2002;17:198-200.
18. Yancy CW. The role of race in heart failure therapy. *Curr Cardiol Rep* 2002;4:218-25.
19. Hajjar RJ, MacRae CA. Adrenergic-receptor polymorphisms and heart failure. *N Engl J Med* 2002;347:1196-9.
20. Henderson SO, Coetzee GA, Ross RK, Yu MC, Henderson BE. Elevated mortality rates from circulatory disease in African American men and women of Los Angeles County, California — a possible genetic susceptibility? *Am J Med Sci* 2000;320:18-23.
21. The HGP: end of phase 1. *Nat Genet* 2002;30:125.
22. Kimm SYS, Glynn NW, Aston CE, et al. Racial differences in the relation between uncoupling protein genes and resting energy expenditure. *Am J Clin Nutr* 2002;75:714-9.
23. Robbins AS, Whittmore AS, Thom DH. Differences in socioeconomic status and survival among white and black men with prostate cancer. *Am J Epidemiol* 2000;151:409-16.
24. Brancati FL, Whelton PK, Kuller LH, Klag MJ. Diabetes mellitus, race, and socioeconomic status: a population-based study. *Ann Epidemiol* 1996;6:67-73.
25. Brewster LM, Clark JF, van Montfrans GA. Is greater tissue activity of creatine kinase the genetic factor increasing hypertension risk in black people of sub-Saharan African descent? *J Hypertens* 2000; 18:1537-44.
26. Grim CE, Robinson M. Blood pressure variation in blacks: genetic factors. *Semin Nephrol* 1996;16:83-93.
27. Chase A. *The legacy of Malthus: the social costs of the new scientific racism*. New York: Alfred A. Knopf, 1977.
28. Gould SJ. *The mismeasure of man*. Rev. ed. New York: W.W. Norton, 1996.
29. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.
30. Cooper R, Rotimi C, Ataman S, et al. The prevalence of hypertension in seven populations of West African origin. *Am J Public Health* 1997;87:160-8.
31. Cooper RS, Rotimi CN, Kaufman JS, et al. Prevalence of NIDDM among populations of the African diaspora. *Diabetes Care* 1997;20: 343-8.

Copyright © 2003 Massachusetts Medical Society.

## The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice

Esteban González Burchard, M.D., Elad Ziv, M.D., Natasha Coyle, Ph.D., Scarlett Lin Gomez, Ph.D., Hua Tang, Ph.D., Andrew J. Karter, Ph.D., Joanna L. Mountain, Ph.D., Eliseo J. Pérez-Stable, M.D., Dean Sheppard, M.D., and Neil Risch, Ph.D.

A debate has recently arisen over the use of racial classification in medicine and biomedical research. In particular, with the completion of a rough draft of the human genome, some have suggested that racial classification may not be useful for biomed-

ical studies, since it reflects “a fairly small number of genes that describe appearance”<sup>1</sup> and “there is no basis in the genetic code for race.”<sup>2</sup> In part on the basis of these conclusions, some have argued for the exclusion of racial and ethnic classification