DOES  EXIST

By Michael J. Bamshad and Steve E. Olson

Look around on the streets of any major city, and you will see a sampling of the outward variety of humanity: skin tones ranging from milk-white to dark brown; hair textures running the gamut from fine and stick-straight to thick and wiry. People often use physical characteristics such as these—along with area of geographic origin and shared culture—to group themselves and others into "races." But how valid is the concept of race from a biological standpoint? Do physical features reliably say anything informative about a person's genetic makeup beyond indicating that the individual has genes for blue eyes or curly hair?

The problem is hard in part because the implicit definition of what makes a person a member of a particular race differs from region to region across the globe. Someone classified as "black" in the U.S., for instance, might be considered "white" in Brazil and "colored" (a category distinguished from both "black" and "white") in South Africa.

Yet common definitions of race do sometimes work well to divide groups according to genetically determined propensities for certain diseases. Sickle cell disease is usually found among people of largely African or Mediterranean descent, for instance, whereas cystic fibrosis is far more common among those of European ancestry. In addition, although the results have been controversial, a handful of studies have suggested that African-Americans are more likely to respond poorly to some drugs for cardiac disease than are members of other groups.

Over the past few years, scientists have collected data about the genetic constitution of populations around the world in an effort to probe the link between ancestry and patterns of disease. These data are now providing answers to several highly emotional and contentious questions: Can genetic information be used to distinguish human groups having a common heritage and to assign individuals to particular ones? Do such groups correspond well to predefined descriptions now widely used to specify race? And, more practically, does dividing people by familiar racial definitions or by genetic similarities say anything useful about how members of those groups experience disease or respond to drug treatment?

In general, we would answer the first question yes, the second no, and offer a qualified yes to the third. Our answers rest on several generalizations about race and genetics. Some groups do differ genetically from others, but how groups are divided depends on which genes are examined; simplistically put, you might fit into one group based on your skin-color genes but another based on a different characteristic. Many studies have demonstrated that roughly 90 percent of human genetic variation occurs within a population living on a given continent, whereas about 10 percent of the variation distinguishes continental populations. In other words, individuals from different populations are, on average, just slightly more different from one another than are individuals from the same population. Human populations are very similar, but they often can be distinguished.

Overview/*Genetics of Race*

- The outward signs on which most definitions of race are based—such as skin color and hair texture—are dictated by a handful of genes. But the other genes of two people of the same "race" can be very different. Conversely, two people of different "races" can share more genetic similarity than two individuals of the same race.
- Nevertheless, scientists can use genetics to sort most large populations according to their ancestral geographic origin. This approach does not work as well for populations resulting from recent mixing with other groups, however.
- The medical implications of racial genetic differences are still under debate.

Classifying Humans

AS A FIRST STEP to identifying links between social definitions of race and genetic heritage, scientists need a way to divide groups reliably according to their ancestry. Over the past 100,000 years or so, anatomically modern humans have migrated from Africa to other parts of the world, and members of our species have increased dramatically in number. This spread has left a distinct signature in our DNA.

To determine the degree of relatedness among groups, geneticists rely on tiny variations, or polymorphisms, in the DNA—specifically in the sequence of base pairs, the building blocks of DNA. Most of these polymorphisms do not occur within genes, the stretches of DNA that encode the information for making proteins (the molecules that constitute

Individuals from different populations are, on average, just slightly more different from one another than are individuals from the same population.

much of our bodies and carry out the chemical reactions of life). Accordingly, these common variations are neutral, in that they do not directly affect a particular trait. Some polymorphisms do occur in genes, however; these can contribute to individual variation in traits and to genetic diseases.

As scientists have sequenced the human genome (the full set of nuclear DNA), they have also identified millions of polymorphisms. The distribution of these polymorphisms across populations reflects the history of those populations and the effects of natural selection. To distinguish among groups, the ideal genetic polymorphism would be one that is present in all the members of one group and absent in the members of all other groups. But the major human groups have separated from one another too recently and have mixed too much for such differences to exist.

Polymorphisms that occur at different frequencies around the world can, however, be used to sort people roughly into groups. One useful class of polymorphisms consists of the *Alu*s, short pieces of DNA that are similar in sequence to one another. *Alu*s replicate occasionally, and the resulting copy splices itself at random into a new position on the original chromosome or on another chromosome, usually in a location that has no effect on the functioning of nearby genes. Each insertion is a unique event. Once an *Alu* sequence inserts itself, it can remain in place for eons, getting passed from one person to his or her descendants. Therefore, if two people have the same *Alu* sequence at the same spot in their genome, they must be descended from a common ancestor who gave them that specific segment of DNA.

One of us (Bamshad), working with University of Utah scientists Lynn B. Jorde, Stephen Wooding and W. Scott Watkins and with Mark A. Batzer of Louisiana State University, examined 100 different *Alu* polymorphisms in 565 people born in sub-Saharan Africa, Asia and Europe. First we determined the presence or absence of the 100 *Alu*s in each of the 565 people. Next we removed all the identifying labels (such as place of origin and ethnic group) from the data and sorted the people into groups using only their genetic information.

Our analysis yielded four different groups. When we added the labels back to see whether each individual's group assignment correlated to common, predefined labels for race or ethnicity, we saw that two of the groups consisted only of individuals from sub-Saharan Africa, with one of those two made up almost entirely of Mbuti Pygmies. The other two groups consisted only of individuals from Europe and East Asia, respectively. We found that we needed 60 *Alu* polymorphisms to assign individuals to their continent of origin with 90 percent accuracy. To achieve nearly 100 percent accuracy, however, we needed to use about 100 *Alu*s.

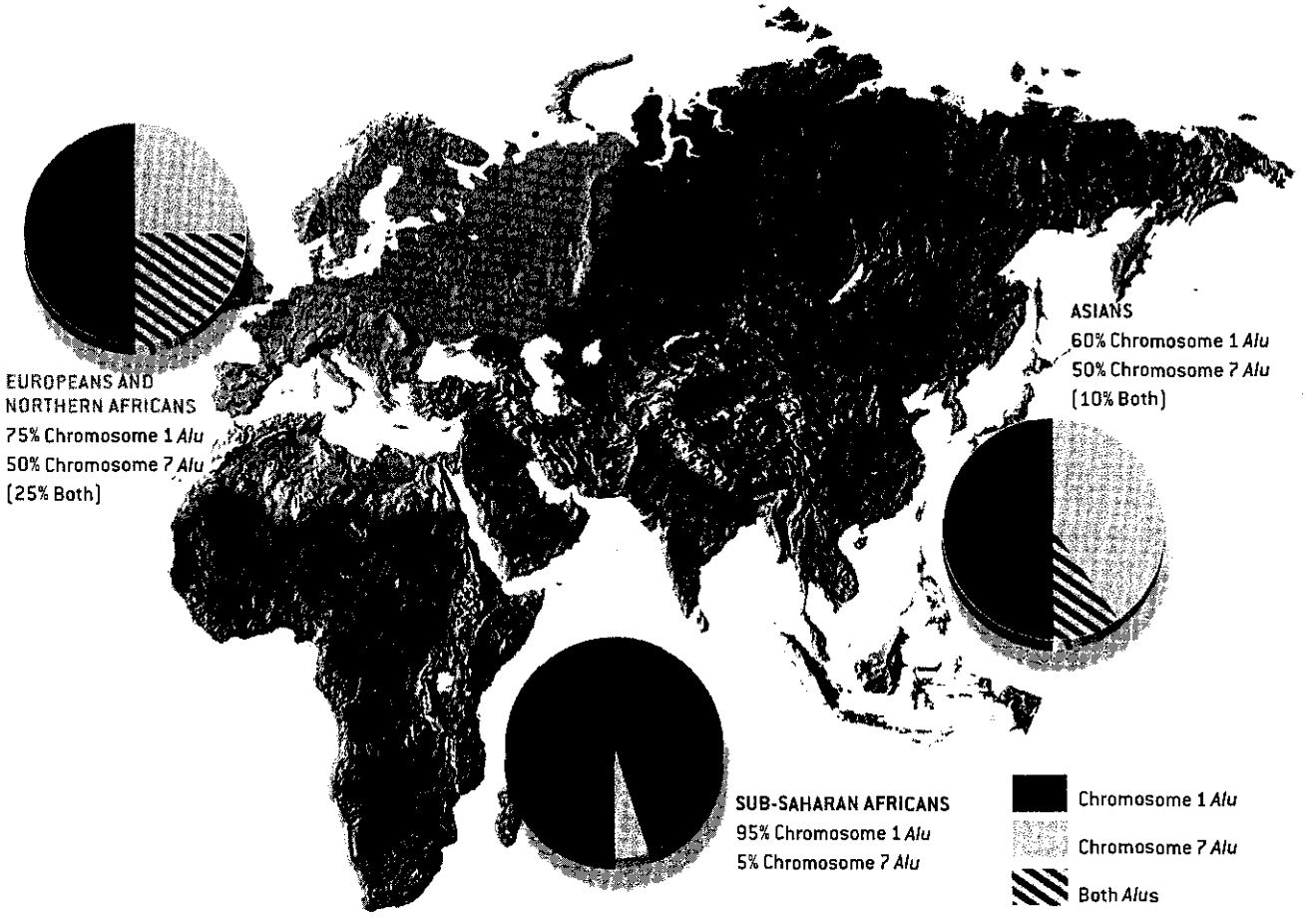
Other studies have produced comparable results. Noah A. Rosenberg and Jonathan K. Pritchard, geneticists formerly in the laboratory of Marcus W. Feldman of Stanford University, assayed approximately 375 polymorphisms called short tandem repeats in more than 1,000 people from 52 ethnic groups in Africa, Asia, Europe and the Americas. By look-

THE AUTHORS

MICHAEL J. BAMSHAD and *STEVE E. OLSON* have come to the subject of human genetic variation from different directions. Bamshad, a geneticist at the University of Utah School of Medicine, studies the genetics of populations to better understand human history and the origins of disease. Olson, a science writer who lives outside Washington, D.C., is the author of *Mapping Human History: Genes, Race, and Our Common Origins,* which was nominated for the 2002 National Book Award in nonfiction. They met during the writing of *Mapping Human History* and have been working together ever since to explore and explain patterns of human genetic variation worldwide.

HUMAN GENETIC DIVERSITY

RESEARCHERS OFTEN USE SHORT PIECES of DNA called *Alu* polymorphisms to determine whether various populations are related to one another. *Alu*s have no known function, yet they copy and insert themselves at random throughout a person's genome. Because previously inserted *Alu*s do not excise themselves, *Alu* patterns can be used as yardsticks to estimate how close two people—and, on average, two populations—are genetically. For example, an *Alu* polymorphism on chromosome 1 occurs in roughly 95 percent of sub-Saharan Africans who have been sampled, 75 percent of Europeans and northern Africans, and 60 percent of Asians, whereas a different *Alu* polymorphism on chromosome 7 is carried by approximately 5 percent of sub-Saharan Africans, 50 percent of Europeans and northern Africans, and 50 percent of Asians. Some individuals carry both polymorphisms. No single polymorphism can, by itself, distinguish all the members of one major human group from all the members of another group, but by analyzing hundreds of these polymorphisms, scientists can group individuals sampled from different locations on the basis of their genetic profiles. *—M.J.B. and S.E.O.*



ing at the varying frequencies of these polymorphisms, they were able to distinguish five different groups of people whose ancestors were typically isolated by oceans, deserts or mountains: sub-Saharan Africans; Europeans and Asians west of the Himalayas; East Asians; inhabitants of New Guinea and Melanesia; and Native Americans. They were also able to identify subgroups within each region that usually corresponded with each member's self-reported ethnicity.

The results of these studies indicate that genetic analyses can distinguish groups of people according to their geographic origin. But caution is warranted. The groups easiest to resolve were those that were widely separated from one another geographically. Such samples maximize the genetic variation among groups. When Bamshad and his co-workers used their 100 *Alu* polymorphisms to try to classify a sample of individuals from southern India into a separate group, the Indians instead had more in common with either Europeans or Asians. In other words, because India has been subject to many genetic influences from Europe and Asia, people on the subcontinent did not group into a unique cluster. We concluded that many hundreds—or perhaps thousands—of polymorphisms might have to be examined to distinguish between groups whose ancestors have historically interbred with multiple populations.

Genetic analyses can distinguish groups of people according to their geographic origin. But caution is warranted.

The Human Race

GIVEN THAT PEOPLE can be sorted broadly into groups using genetic data, do common notions of race correspond to underlying genetic differences among populations? In some cases they do, but often they do not. For instance, skin color or facial features—traits influenced by natural selection—are routinely used to divide people into races. But groups with similar physical characteristics as a result of selection can be quite different genetically. Individuals from sub-Saharan Africa and Australian Aborigines might have similar skin pigmentation (because of adapting to strong sun), but genetically they are quite dissimilar.

In contrast, two groups that are genetically similar to each other might be exposed to different selective forces. In this case, natural selection can exaggerate some of the differences between groups, making them appear more dissimilar on the surface than they are underneath. Because traits such as skin color have been strongly affected by natural selection, they do not necessarily reflect the population processes that have shaped the distribution of neutral polymorphisms such as *Alu*s or short tandem repeats. Therefore, traits or polymorphisms affected by natural selection may be poor predictors of group membership and may imply genetic relatedness where, in fact, little exists.

Another example of how difficult it is to categorize people involves populations in the U.S. Most people who describe themselves as African-American have relatively recent ancestors from West Africa, and West Africans generally have polymorphism frequencies that can be distinguished from those of Europeans, Asians and Native Americans. The fraction of gene variations that African-Americans share with West Africans, however, is far from uniform, because over the centuries African-Americans have mixed extensively with groups originating from elsewhere in Africa and beyond.

Over the past several years, Mark D. Shriver of Pennsylvania State University and Rick A. Kittles of Howard University have defined a set of polymorphisms that they have used to estimate the fraction of a person's genes originating from each continental region. They found that the West African contribution to the genes of individual African-Americans averages about 80 percent, although it ranges from 20 to 100 percent. Mixing of groups is also apparent in many individuals who believe they have only European ancestors. According to Shriver's analyses, approximately 30 percent of Americans who consider themselves "white" have less than 90 percent European ancestry. Thus, self-reported ancestry is not necessarily a good predictor of the genetic composition of a large number of Americans. Accordingly, common notions of race do not always reflect a person's genetic background.

Membership Has Its Privileges

UNDERSTANDING the relation between race and genetic variation has important practical implications. Several of the polymorphisms that differ in frequency from group to group have specific effects on health. The mutations responsible for sickle cell disease and some cases of cystic fibrosis, for instance, result from genetic changes that appear to have risen in frequency because they were protective against diseases prevalent in Africa and Europe, respectively. People who inherit one copy of the sickle cell polymorphism show some resistance to malaria; those with one copy of the cystic fibrosis trait may be less prone to the dehydration resulting from cholera. The symptoms of these diseases arise only in the unfortunate individuals who inherit two copies of the mutations.

Genetic variation also plays a role in individual susceptibility to one of the worst scourges of our age: AIDS. Some people have a small deletion in both their copies of a gene that encodes a particular cell-surface receptor called chemokine receptor 5 (CCR5). As a result, these indi-

viduals fail to produce CCR5 receptors on the surface of their cells. Most strains of HIV-1, the virus that causes AIDS, bind to the CCR5 receptor to gain entry to cells, so people who lack CCR5 receptors are resistant to HIV-1 infection. This polymorphism in the CCR5 receptor gene is found almost exclusively in groups from northeastern Europe.

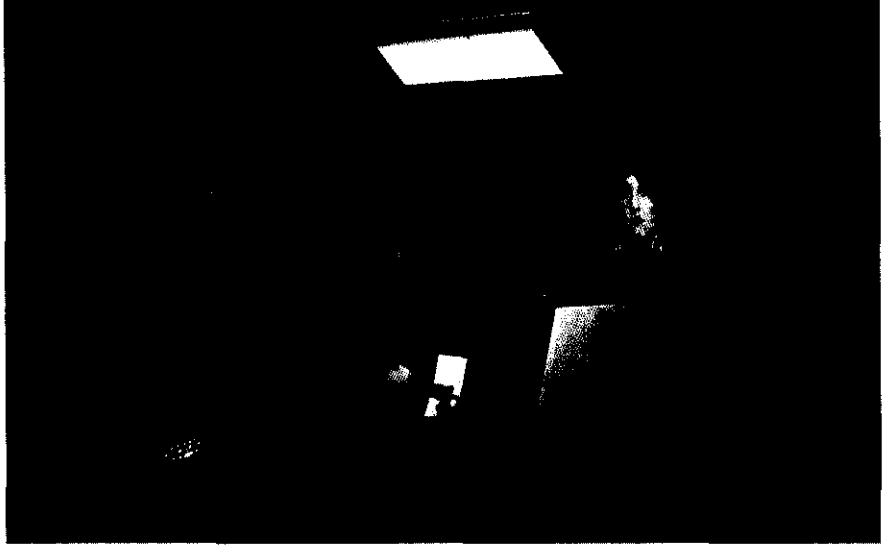
Several polymorphisms in CCR5 do not prevent infection but instead influence the rate at which HIV-1 infection leads to AIDS and death. Some of these polymorphisms have similar effects in different populations; others only alter the speed of disease progression in selected groups. One polymorphism, for example, is associated with delayed disease progression in European-Americans but accelerated disease in African-Americans. Researchers can only study such population-specific effects—and use that knowledge to direct therapy—if they can sort people into groups.

In these examples—and others like them—a polymorphism has a relatively large effect in a given disease. If genetic screening were inexpensive and efficient, *all* individuals could be screened for *all* such disease-related gene variants. But genetic testing remains costly. Perhaps more significantly, genetic screening raises concerns about privacy and consent: some people might not want to know about genetic factors that could increase their risk of developing a particular disease. Until these issues are resolved further, self-reported ancestry will continue to be a potentially useful diagnostic tool for physicians.

Ancestry may also be relevant for some diseases that are widespread in particular populations. Most common diseases, such as hypertension and diabetes, are the cumulative results of polymorphisms in several genes, each of which has a small influence on its own. Recent research suggests that polymorphisms that have a particular effect in one group may have a different effect in another group. This kind of complexity would make it much more difficult to use detected polymorphisms as a guide to therapy. Until further studies are done on the genetic and environmental contributions to complex diseases, physicians may have to rely on information about an individual's ancestry to know how best to treat some diseases.

About the Photoillustrations

NEW YORK ARTIST Nancy Burson used her invention, called the Human Race Machine, to generate the series of composite photographs that appears on page 79 (two also appear elsewhere in this article). The machine takes a photograph of an individual—in this case, a white woman—and adds and subtracts various outward features of racial identity to show what a person might look like if he or she were a member of another race. Burson says she seeks to use her work to underscore the commonality of humanity. And indeed, outward features often say little about the bulk of a person's genetic makeup. The Human Race Machine was one of the most popular attractions at the Millennium Dome in London, where it drew hundreds of thousands of people. More about the artist and her work can be found on her Web site, www.nancyburson.com



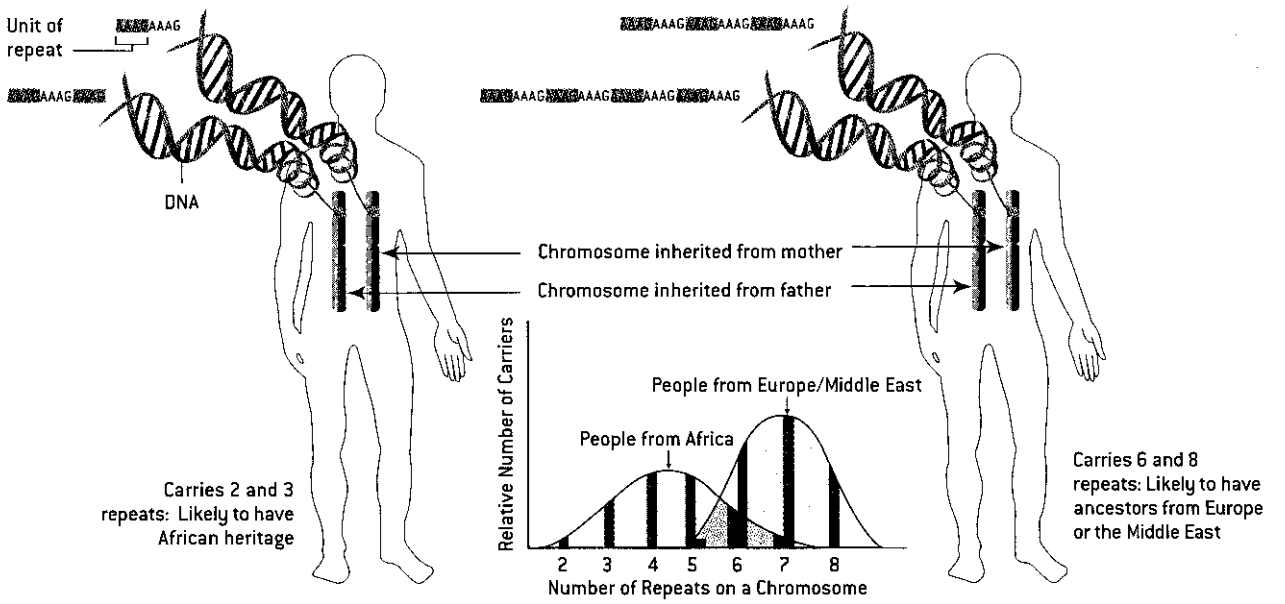
Race and Medicine

BUT THE IMPORTANCE of group membership as it relates to health care has been especially controversial in recent years. Last January the U.S. Food and Drug Administration issued guidelines advocating the collection of race and ethnicity data in all clinical trials. Some investigators contend that the differences between groups are so small and the historical abuses associated with categorizing people by race so extreme that group membership should play little if any role in genetic and medical studies. They assert that the FDA should abandon its recommendation and instead ask researchers conducting clinical trials to collect genomic data on each individual. Others suggest that only by using group membership, including common definitions of race based on skin color, can we understand how genetic and environmental differences among groups contribute to disease. This debate will be settled only by further research on the validity of race as a scientific variable.

A set of articles in the March 20 issue of the *New England Journal of Medicine* debated both sides of the medical implications of race. The authors of one article—Richard S. Cooper of the Loyola Stritch School of Medicine, Jay S. Kaufman of the University of North Carolina at Chapel Hill and Ryk Ward of the University of Oxford—argued that race is not an adequate criterion for physicians to use in choosing a particular drug for a given patient. They pointed out two findings of racial differences that are both

Tracing Human Origins

COUNTING THE NUMBER of DNA units called short tandem repeats on chromosomes can allow scientists to group individuals according to probable ancestry. One such repeat, AAAG, occurs between two and seven times in people with African heritage but between five and eight times in those whose ancestors came from Europe or the Middle East. (Every person inherits one set of repeats from their mother and one from their father.) Accordingly, someone who carries two and three repeats is likely to have African heritage, whereas someone with six and eight repeats probably has ancestors from Europe or the Middle East. People with between five and seven repeats occur in both populations, however, making these individuals more difficult to classify using only this particular repeat. *—M.J.B. and S.E.O.*



now considered questionable: that a combination of certain blood vessel–dilating drugs was more effective in treating heart failure in people of African ancestry and that specific enzyme inhibitors (angiotensin converting enzyme, or ACE, inhibitors) have little efficacy in such individuals. In the second article, a group led by Neil Risch of Stanford University countered that racial or ethnic groups can differ from one another genetically and that the differences can have medical importance. They cited a study showing that the rate of complications from type 2 diabetes varies according to race, even after adjusting for such factors as disparities in education and income.

The intensity of these arguments reflects both scientific and social factors. Many biomedical studies have not rigorously defined group membership, relying instead on inferred relationships based on racial categories. The dispute over the importance of group membership also illustrates how strongly the perception of race is shaped by different social and political perspectives.

In cases where membership in a geographically or culturally defined group has been correlated with health-related genetic traits, knowing something about an individual's group membership could be important for a physician. And to the extent that human groups live in different environments or have different experiences that affect health, group membership could also reflect nongenetic factors that are medically relevant.

Regardless of the medical implications of the genetics of race, the research findings are inherently exciting. For hundreds of years, people have wondered where various human groups came from and how those groups are related to one another. They have speculated about why human populations have different physical appearances and about whether the biological differences between groups are more than skin deep. New genetic data and new methods of analysis are finally allowing us to approach these questions. The result will be a much deeper understanding of both our biological nature and our human interconnectedness. SA

MORE TO EXPLORE

Mapping Human History: Genes, Race, and Our Common Origins. Steve Olson. Mariner Books, 2003.

Human Population Genetic Structure and Inference of Group Membership. Michael J. Bamshad et al. in *American Journal of Human Genetics,* Vol. 72, No. 3, pages 578–589; March 2003.

The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice. Esteban González Burchard et al. in *New England Journal of Medicine,* Vol. 348, No. 12, pages 1170–1175; March 20, 2003.

Race and Genomics. Richard S. Cooper, Jay S. Kaufman and Ryk Ward in *New England Journal of Medicine,* Vol. 348, No. 12, pages 1166–1170; March 20, 2003.