Am. J. Hum. Genet. 77:519–532, 2005

REVIEW ARTICLE
The Use of Racial, Ethnic, and Ancestral Categories in Human Genetics Research

Race, Ethnicity, and Genetics Working Group*

National Human Genome Research Institute, Bethesda

The global dispersal of anatomically modern humans over the past 100,000 years has produced patterns of phenotypic variation that have exerted—and continue to exert—powerful influences on the lives of individuals and the experiences of groups. The recency of our common ancestry and continued gene flow among populations have resulted in less genetic differentiation among geographically distributed human populations than is observed in many other mammalian species. Nevertheless, differences in appearance have contributed to the development of ideas about "race" and "ethnicity" that often include the belief that significant inherited differences distinguish humans. The use of racial, ethnic, and ancestral categories in genetics research can imply that group differences arise directly through differing allele frequencies, with little influence from socially mediated mechanisms. At the same time, careful investigations of the biological, environmental, social, and psychological attributes associated with these categories will be an essential component of cross-disciplinary research into the origins, prevention, and treatment of common diseases, including those diseases that differ in prevalence among groups.

Introduction

Human genetics research is generating unprecedented amounts of data about the genetic differences among individuals and groups. Investigation of these differences will transform our understanding of the origins and nature of human diseases (Collins et al. 2003).

Research into human genetic differences also has the potential to generate great controversy. In the past, concepts drawn from genetics have been used—both by geneticists and by individuals outside the field—to justify and perpetuate racial and ethnic discrimination (Kevles 1985; Provine 1986). The belief that racial and ethnic groups have substantial, well-demarcated biological differences and that these differences are important has contributed to many of the great atrocities of the 20th century and continues to shape personal interactions and social institutions (Mosse 1985; Shipler 1997). Because of the history of misuse of genetics ideas, geneticists have a special responsibility to examine carefully their use of racial and ethnic categories in their research. Investigations that fail to recognize and acknowledge the full range of mechanisms through which designations of race, ethnicity, and ancestry can correlate with personal traits and health outcomes threaten to reinforce widely held stereotypes. Yet genetics research also has the potential, by delineating the complex origins of traits and the close biological affinities between human groups, to help dispel these stereotypes.

The sequencing of the human genome (International Human Genome Sequencing Consortium 2001; Venter et al. 2001) and the ongoing international effort to catalog common haplotypes in several populations (International HapMap Consortium 2003) make this an opportune time to examine the complex relationships between genetics research and the categories of race, ethnicity, and ancestry. Although such a review inevitably draws on very different academic disciplines and literatures, a cross-disciplinary conversation is essential for reconciliation of the promise of genetics research with the historical and potential abuses of ideas drawn from genetics. This review summarizes what is known about patterns of human genetic variation; the historical development of widely held conceptions about race, ethnicity, and ancestry; and the interactions between these conceptions and human genetics research.

Received October 20, 2004; accepted for publication July 27, 2005; electronically published August 29, 2005.

Address for correspondence and reprints: Steve Olson, 7609 Sebago Road, Bethesda, MD 20817. E-mail: solson@nas.edu

* The Race, Ethnicity, and Genetics Working Group of the National Human Genome Research Institute includes Kate Berg, Vence Bonham, Joy Boyer, Larry Brody, Lisa Brooks, Francis Collins, Alan Guttmacher, Jean McEwen, Max Muenke, Steve Olson, Vivian Ota Wang, Laura Lyman Rodriguez, Nadarajen Vydelingum, and Esther Warshauer-Baker.

0002-9297/2005/7704-0002

The Origins, Patterns, and Physical Manifestations of Human Genetic Variation

The Origins of Modern Humans

Information about the history of our species comes from two main sources: the paleoanthropological record

and historical inferences based on current genetic differences observed in humans. Although both sources of information are fragmentary, they have been converging in recent years on the same general story.

The existing fossil evidence suggests that anatomically modern humans evolved in Africa, within the last ~200,000 years, from a pre-existing population of humans (Klein 1999). Although it is not easy to define "anatomically modern" in a way that encompasses all living humans and excludes all archaic humans (Lieberman et al. 2002), the generally agreed-upon physical characteristics of anatomical modernity include a high rounded skull, facial retraction, and a light and gracile, as opposed to heavy and robust, skeleton (Lahr 1996). Early fossils with these characteristics have been found in eastern Africa and have been dated to ~160,000–200,000 years ago (White et al. 2003; McDougall et al. 2005). At that time, the population of anatomically modern humans appears to have been small and localized (Harpending et al. 1998). Much larger populations of archaic humans lived elsewhere in the Old World, including the Neandertals in Europe and an earlier species of humans, *Homo erectus,* in Asia (Swisher et al. 1994).

Fossils of the earliest anatomically modern humans found outside Africa are from two sites in the Middle East and date to a period of relative global warmth, ~100,000 years ago, though this region was reinhabited by Neandertals in later millennia as the climate in the northern hemisphere again cooled (Lahr and Foley 1998). Groups of anatomically modern humans appear to have moved outside Africa permanently sometime >60,000 years ago. One of the earliest modern skeletons found outside Africa is from Australia and has been dated to ~42,000 years ago (Bowler et al. 2003), although studies of environmental changes in Australia argue for the presence of modern humans in Australia >55,000 years ago (Miller et al. 1999). To date, the earliest anatomically modern skeleton discovered from Europe comes from the Carpathian Mountains of Romania and is dated to 34,000–36,000 years ago (Trinkaus et al. 2003).

Existing data on human genetic variation support and extend conclusions based on the fossil evidence. African populations exhibit greater genetic diversity than do populations in the rest of the world, implying that humans appeared first in Africa and later colonized Eurasia and the Americas (Tishkoff and Williams 2002; Yu et al. 2002; Tishkoff and Verrelli 2003). The genetic variation seen outside Africa is generally a subset of the variation within Africa, a pattern that would be produced if the migrants from Africa were limited in number and carried just part of African genetic variability with them (Cavalli-Sforza and Feldman 2003). Patterns of genetic variation suggest an earlier population expansion in Africa followed by a subsequent expansion in non-African populations, and the dates calculated for the expansions generally coincide with the archaeological record (Jorde et al. 1998).

Aspects of the relationship between anatomically modern and archaic humans remain contentious. Studies of mtDNA (Ingman et al. 2000), the Y chromosome (Underhill et al. 2000), portions of the X chromosome (Kaessmann et al. 1999), and many (though not all) autosomal regions (Harpending and Rogers 2000) support the "Out of Africa" account of human history, in which anatomically modern humans appeared first in eastern Africa and then migrated throughout Africa and into the rest of the world, with little or no interbreeding between modern humans and the archaic populations they gradually replaced (Tishkoff et al. 2000; Stringer 2002). However, several groups of researchers cite fossil and genetic evidence to argue for a more complex account. They contend that humans bearing modern traits emerged several times from Africa, over an extended period, and mixed with archaic humans in various parts of the world (Hawks et al. 2000; Eswaran 2002; Templeton 2002; Ziętkiewicz et al. 2003). As a result, they say, autosomal DNA from archaic human populations living outside Africa persists in modern populations, and modern populations in various parts of the world still bear some physical resemblance to the archaic populations that inhabited those regions (Wolpoff et al. 2001).

However, distinguishing possible contributions to the gene pool of modern humans from archaic humans outside Africa is difficult, especially since many autosomal loci coalesce at times preceding the separation of archaic human populations (Pääbo 2003). In addition, studies of mtDNA from archaic and modern humans and extant Y chromosomes suggest that any surviving genetic contributions of archaic humans outside Africa must be small, if they exist at all (Krings et al. 1997; Nordborg 1998; Takahata et al. 2001; Serre et al. 2004). The observation that most genes studied to date coalesce in African populations points toward the importance of Africa as the source of most modern genetic variation, perhaps with some subdivision in the ancestral African population (Satta and Takahata 2002). Sequence data for hundreds of loci from widely distributed worldwide populations eventually may clarify the population processes associated with the appearance of anatomically modern humans (Wall 2000), as well as the amount of gene flow among modern humans since then.

The Distribution of Variation

A thorough description of the differences in patterns of genetic variation between humans and other species awaits additional genetic studies of human populations

and nonhuman species. But the data gathered to date suggest that human variation exhibits several distinctive characteristics. First, compared with many other mammalian species, humans are genetically less diverse—a counterintuitive finding, given our large population and worldwide distribution (Li and Sadler 1991; Kaessmann et al. 2001). For example, the chimpanzee subspecies living just in central and western Africa have higher levels of diversity than do humans (Ebersberger et al. 2002; Yu et al. 2003; Fischer et al. 2004).

The distribution of variants within and among human populations also differs from that of many other species. The details of this distribution are impossible to describe succinctly because of the difficulty of defining a "population," the clinal nature of variation, and heterogeneity across the genome (Long and Kittles 2003). In general, however, 5%–15% of genetic variation occurs between large groups living on different continents, with the remaining majority of the variation occurring within such groups (Lewontin 1972; Jorde et al. 2000*a*; Hinds et al. 2005). This distribution of genetic variation differs from the pattern seen in many other mammalian species, for which existing data suggest greater differentiation between groups (Templeton 1998; Kittles and Weiss 2003).

Our history as a species also has left genetic signals in regional populations. For example, in addition to having higher levels of genetic diversity, populations in Africa tend to have lower amounts of linkage disequilibrium than do populations outside Africa, partly because of the larger size of human populations in Africa over the course of human history and partly because the number of modern humans who left Africa to colonize the rest of the world appears to have been relatively low (Gabriel et al. 2002). In contrast, populations that have undergone dramatic size reductions or rapid expansions in the past and populations formed by the mixture of previously separate ancestral groups can have unusually high levels of linkage disequilibrium (Nordborg and Tavare 2002).

Many other geographic, climatic, and historical factors have contributed to the patterns of human genetic variation seen in the world today. For example, population processes associated with colonization, periods of geographic isolation, socially reinforced endogamy, and natural selection all have affected allele frequencies in certain populations (Jorde et al. 2000*b*; Bamshad and Wooding 2003). In general, however, the recency of our common ancestry and continual gene flow among human groups have limited genetic differentiation in our species.

Substructure in the Human Population

Although the genetic differences among human groups are relatively small, these differences nevertheless can be used to situate many individuals within broad, geographically based groupings. For example, computer analyses of hundreds of polymorphic loci sampled in globally distributed populations have revealed the existence of genetic clustering that roughly is associated with groups that historically have occupied large continental and subcontinental regions (Rosenberg et al. 2002; Bamshad et al. 2003).

Some commentators have argued that these patterns of variation provide a biological justification for the use of traditional racial categories. They argue that the continental clusterings correspond roughly with the division of human beings into sub-Saharan Africans; Europeans, western Asians, and northern Africans; eastern Asians; Polynesians and other inhabitants of Oceania; and Native Americans (Risch et al. 2002). Other observers disagree, saying that the same data undercut traditional notions of racial groups (King and Motulsky 2002; Calafell 2003; Tishkoff and Kidd 2004). They point out, for example, that major populations considered races or subgroups within races do not necessarily form their own clusters. Thus, samples taken from India and Pakistan affiliate with Europeans or eastern Asians rather than separating into a distinct cluster. However, samples from the Kalash, a small population living in northwestern Pakistan, form their own cluster on a level comparable with those of the major continental regions (Rosenberg et al. 2002).

Sampling design can have a critical influence on the results of such studies. Studies of genetic clustering often have relied on samples taken from widely separated and socially defined populations. When samples were analyzed from individuals who were more evenly distributed geographically, clustering was far less evident (Serre and Pääbo 2004). Furthermore, because human genetic variation is clinal, many individuals affiliate with two or more continental groups. Thus, the genetically based "biogeographical ancestry" assigned to any given person generally will be broadly distributed and will be accompanied by sizable uncertainties (Pfaff et al. 2004).

In many parts of the world, groups have mixed in such a way that many individuals have relatively recent ancestors from widely separated regions. Although genetic analyses of large numbers of loci can produce estimates of the percentage of a person's ancestors coming from various continental populations (Shriver et al. 2003; Bamshad et al. 2004), these estimates may assume a false distinctiveness of the parental populations, since human groups have exchanged mates from local to continental scales throughout history (Cavalli-Sforza et al. 1994; Hoerder 2002). Even with large numbers of markers, information for estimating admixture proportions of individuals or groups is limited, and estimates typically will have wide CIs (Pfaff et al. 2004).

Physical Variation in Humans

The distribution of many physical traits resembles the distribution of genetic variation within and between human populations (American Association of Physical Anthropologists 1996; Keita and Kittles 1997). For example, ~90% of the variation in human head shapes occurs within every human group, and ~10% separates groups, with a greater variability of head shape among individuals with recent African ancestors (Relethford 2002).

A prominent exception to the common distribution of physical characteristics within and among groups is skin color. Approximately 10% of the variance in skin color occurs within groups, and ~90% occurs between groups (Relethford 2002). This distribution of skin color and its geographic patterning—with people whose ancestors lived predominantly near the equator having darker skin than those with ancestors who lived predominantly in higher latitudes—indicate that this attribute has been under strong selective pressure. Darker skin appears to be strongly selected for in equatorial regions to prevent sunburn, skin cancer, the photolysis of folate, and damage to sweat glands (Sturm et al. 2001; Rees 2003). A leading hypothesis for the selection of lighter skin in higher latitudes is that it enables the body to form greater amounts of vitamin D, which helps prevent rickets (Jablonski 2004). However, the vitamin D hypothesis is not universally accepted (Aoki 2002), and lighter skin in high latitudes may correspond simply to an absence of selection for dark skin (Harding et al. 2000).

Because skin color has been under strong selective pressure, similar skin colors can result from convergent adaptation rather than from genetic relatedness. Sub-Saharan Africans, tribal populations from southern India, and Australian Aborigines have similar skin pigmentation, but genetically they are no more similar than are other widely separated groups. Furthermore, in some parts of the world in which people from different regions have mixed extensively, the connection between skin color and ancestry has been substantially weakened (Parra et al. 2004). In Brazil, for example, skin color is not closely associated with the percentage of recent African ancestors a person has, as estimated from an analysis of genetic variants differing in frequency among continent groups (Parra et al. 2003).

Considerable speculation has surrounded the possible adaptive value of other physical features characteristic of groups, such as the constellation of facial features observed in many eastern and northeastern Asians (Guthrie 1996). However, any given physical characteristic generally is found in multiple groups (Lahr 1996), and demonstrating that environmental selective pressures shaped specific physical features will be difficult, since such features may have resulted from sexual selection for individuals with certain appearances or from genetic drift (Roseman 2004).

The Social Interpretation of Physical Variation

The Development of the "Ideology of Race"

Given our visual acuity and complex social relationships, humans presumably have always observed and speculated about the physical differences among individuals and groups. But different societies have attributed markedly different meanings to these distinctions. Classical civilizations from Rome to China tended to invest much more importance in family or tribal affiliations than in physical appearance (Dikötter 1992; Goldenberg 2003). Some Roman writers adhered to an environmental determinism in which climate could affect the appearance and character of groups (Isaac 2004). But in many ancient civilizations, individuals with widely varying physical appearances could become full members of a society by growing up within that society or by adopting the society's cultural norms (Snowden 1983; Lewis 1990).

The English word "race" (possibly derived from the Spanish *raza,* meaning "breed" or "stock"), along with many of the ideas now associated with the term, were products of the European era of exploration (Smedley 1999). As Europeans encountered people from different parts of the world, they speculated about the physical, social, and cultural differences between human groups. The rise of the African slave trade, which gradually displaced an earlier trade in slaves from throughout the world, created a further incentive to categorize human groups to justify the barbarous treatment of African slaves (Meltzer 1993). Drawing on classical sources and on their own internal interactions—for example, the hostility between the English and Irish was a powerful influence on early thinking about the differences between people (Takaki 1993)—Europeans began to sort themselves and others into groups associated with physical appearance and with deeply ingrained behaviors and capacities. A set of "folk beliefs" took hold that linked inherited physical differences between groups to inherited intellectual, behavioral, and moral qualities (Banton 1977). Although similar ideas can be found in other cultures (Lewis 1990; Dikötter 1992), they appear not to have had as much influence on social structures as they did in Europe and the parts of the world colonized by Europeans.

In the 18th century, the differences between human groups became a focus of scientific investigation (Todorov 1993). Initially, scholars focused on cataloging and describing "The Natural Varieties of Mankind," as Johann Friedrich Blumenbach entitled his 1775 text (which established the five major divisions of humans

still reflected in some racial classifications). But as the science of anthropology took shape in the 19th century, European and American scientists increasingly sought explanations for the behavioral and cultural differences they attributed to groups (Stanton 1960). For example, they measured the shapes and sizes of skulls and related the results to group differences in intelligence or other attributes (Lieberman 2001). Both before and after the 1859 publication of *On the Origins of Species,* a debate raged in Europe over whether different human groups had the same origin or were the product of separate creations or evolutionary lineages (Wolpoff and Caspari 1997).

From the 17th through the 19th centuries, the merging of folk beliefs about group differences with scientific explanations of those differences produced what one scholar has called an "ideology of race" (Smedley 1999). According to this ideology, races are primordial, natural, enduring, and distinct. Some groups might be the result of mixture between formerly distinct populations, but careful study can distinguish the ancestral races that had combined to produce admixed groups.

The concept of race found wide application in many societies. The eugenics movement of the late 19th and early 20th centuries asserted as self-evident the biological inferiority of particular groups (Kevles 1985). In many parts of the world, the idea of race became a way of rigidly dividing groups by use of culture as well as physical appearances (Hannaford 1996). Campaigns of oppression and genocide often used supposed racial differences to motivate inhuman acts against others (Horowitz 2001).

The Incongruities of Racial Classifications

Even as the idea of "race" was becoming a powerful organizing principle in many societies, the shortcomings of the concept were apparent. In the Old World, the gradual transition in appearances from one group to adjacent groups emphasized that "one variety of mankind does so sensibly pass into the other, that you cannot mark out the limits between them," as Blumenbach observed in his writings on human variation (Marks 1995, p. 54). In parts of the Americas, the situation was somewhat different. The immigrants to the New World came largely from widely separated regions of the Old World—western and northern Europe, western Africa, and, later, eastern Asia and southern Europe. In the Americas, the immigrant populations began to mix among themselves and with the indigenous inhabitants of the continent. In the United States, for example, most people who self-identify as African American have some European ancestors—in one analysis of genetic markers that have differing frequencies between continents, European ancestry ranged from an estimated 7% for a sample of Jamaicans to ~23% for a sample of African Americans from New Orleans (Parra et al. 1998). Similarly, many people who identify as European American have some African or Native American ancestors, either through openly interracial marriages or through the gradual inclusion of people with mixed ancestry into the majority population. In a survey of college students who self-identified as "white" in a northeastern U.S. university, ~30% were estimated to have <90% European ancestry (Shriver et al. 2003).

In the United States, social and legal conventions developed over time that forced individuals of mixed ancestry into simplified racial categories (Gossett 1997). An example is the "one-drop rule" implemented in some state laws that treated anyone with a single known African American ancestor as black (Davis 2001). The decennial censuses conducted since 1790 in the United States also created an incentive to establish racial categories and fit people into those categories (Nobles 2000). In other countries in the Americas where mixing among groups was more extensive, social categories have tended to be more numerous and fluid, with people moving into or out of categories on the basis of a combination of socioeconomic status, social class, ancestry, and appearance (Mörner 1967).

Efforts to sort the increasingly mixed population of the United States into discrete categories generated many difficulties (Spickard 1992). By the standards used in past censuses, many millions of children born in the United States have belonged to a different race than have one of their biological parents. Efforts to track mixing between groups led to a proliferation of categories (such as "mulatto" and "octoroon") and "blood quantum" distinctions that became increasingly untethered from self-reported ancestry. A person's racial identity can change over time, and self-ascribed race can differ from assigned race (Kressin et al. 2003). Until the 2000 census, Latinos were required to identify with a single race despite the long history of mixing in Latin America; partly as a result of the confusion generated by the distinction, 42% of Latino respondents in the 2000 census ignored the specified racial categories and checked "some other race" (Mays et al. 2003).

Ethnicity as a Way of Categorizing People

As the problems surrounding the word "race" became increasingly apparent during the 20th century, the word "ethnicity" was promoted as a way of characterizing the differences between groups (Huxley and Haddon 1936; Hutchinson and Smith 1996). Ethnicity typically emphasizes the cultural, socioeconomic, religious, and political qualities of human groups rather than their genetic ancestry. It may encompass language, diet, religion, dress,

customs, kinship systems, or historical or territorial identity (Cornell and Hartmann 1998).

However, as a way of understanding human groups, ethnicity also suffers from several shortcomings. First, ascribing an ethnic identity to a group can imply a much greater degree of uniformity than is actually the case. In the United States, the ethnic group "Hispanic or Latino" contains such subgroups as Cuban Americans, Mexican Americans, Puerto Ricans, and recent immigrants from Central America (Hayes-Bautista and Chapa 1987). Combining these groups into a single category may serve useful bureaucratic or political ends but does not necessarily result in a better understanding of these groups.

Also, ethnicity, like race, is a malleable concept that can change dramatically in different times or circumstances (Waters 1990; Smelser et al. 2001). Ethnic groups may come into existence and then dissipate as a result of broad historical or social trends. Individuals might change ethnic groups over the course of their lives or identify with more than one group. A researcher, clinician, or government official might assign an ethnicity to an individual quite different from the one that person would acknowledge (Kressin et al. 2003).

Finally, despite attempts to distinguish "ethnicity" from "race," the two terms often are used interchangeably (Oppenheimer 2001). Ethnic groups can share a belief in a common ancestral origin (Cornell and Hartmann 1998), which also can be a defining characteristic of a racial group. Furthermore, ethnic groups tend to promote marriage within the group, which creates an expectation of biological cohesion regardless of whether that cohesion existed in the past.

Ancestry as a Way of Categorizing People

An alternative to the use of racial or ethnic categories in genetics research is to categorize individuals in terms of ancestry. Ancestry may be defined geographically (e.g., Asian, sub-Saharan African, or northern European), geopolitically (e.g., Vietnamese, Zambian, or Norwegian), or culturally (e.g., Brahmin, Lemba, or Apache). The definition of ancestry may recognize a single predominant source or multiple sources. Ancestry can be ascribed to an individual by an observer, as was the case with the U.S. census prior to 1960; it can be identified by an individual from a list of possibilities or with use of terms drawn from that person's experience; or it can be calculated from genetic data by use of loci with allele frequencies that differ geographically, as described above. At least among those individuals who participate in biomedical research, genetic estimates of biogeographical ancestry generally agree with self-assessed ancestry (Tang et al. 2005), but in an unknown percentage of cases, they do not (Brodwin 2002; Kaplan 2003).

Despite its seemingly objective nature, ancestry also has limitations as a way of categorizing people (Elliott and Brodwin 2002). When asked about the ancestry of their parents and grandparents, many people cannot provide accurate answers. In one series of focus groups in the state of Georgia, 40% of ~100 respondents said they did not know one or more of their four grandparents well enough to be certain how that person(s) would identify racially (Condit et al. 2003). Misattributed paternity or adoption can separate biogeographical ancestry from socially defined ancestry. Furthermore, the exponentially increasing number of our ancestors makes ancestry a quantitative rather than qualitative trait—5 centuries (or 20 generations) ago, each person had a maximum of >1 million ancestors (Ohno 1996). To complicate matters further, recent analyses suggest that everyone living today has exactly the same set of genealogical ancestors who lived as recently as a few thousand years in the past, although we have received our genetic inheritance in different proportions from those ancestors (Rohde et al. 2004).

In the end, the terms "race," "ethnicity," and "ancestry" all describe just a small part of the complex web of biological and social connections that link individuals and groups to each other.

Racial, Ethnic, and Ancestral Categories in Genetics Research

The Effects of Racial and Ethnic Identities on Health

Racial and ethnic groups can exhibit substantial average differences in disease incidence, disease severity, disease progression, and response to treatment (LaVeist 2002). In the United States, African Americans have higher rates of mortality than does any other racial or ethnic group for 8 of the top 10 causes of death (Hummer et al. 2004). U.S. Latinos have higher rates of death from diabetes, liver disease, and infectious diseases than do non-Latinos (Vega and Amaro 1994). Native Americans suffer from higher rates of diabetes, tuberculosis, pneumonia, influenza, and alcoholism than does the rest of the U.S. population (Mahoney and Michalek 1998). European Americans die more often from heart disease and cancer than do Native Americans, Asian Americans, or Hispanics (Hummer et al. 2004).

Considerable evidence indicates that the racial and ethnic health disparities observed in the United States arise mostly through the effects of discrimination, differences in treatment, poverty, lack of access to health care, health-related behaviors, racism, stress, and other socially mediated forces. The infant mortality rate for African Americans is approximately twice the rate for European Americans, but, in a study that looked at members of these two groups who belonged to the military and received care through the same medical system, their

infant mortality rates were essentially equivalent (Rawlings and Weir 1992). Recent immigrants to the United States from Mexico have better indicators on some measures of health than do Mexican Americans who are more assimilated into American culture (Franzini et al. 2001). Diabetes and obesity are more common among Native Americans living on U.S. reservations than among those living outside reservations (Cooper et al. 1997). Rates of heart disease among African Americans are associated with the segregation patterns in the neighborhoods where they live (Fang et al. 1998). Furthermore, the risks for many diseases are elevated for socially, economically, and politically disadvantaged groups in the United States, suggesting that socioeconomic inequities are the root causes of most of the differences (Cooper et al. 2003; Cooper 2004).

However, differences in allele frequencies certainly contribute to group differences in the incidence of some monogenic diseases, and they may contribute to differences in the incidence of some common diseases (Risch et al. 2002; Burchard et al. 2003; Tate and Goldstein 2004). For the monogenic diseases, the frequency of causative alleles usually correlates best with ancestry, whether familial (for example, Ellis–van Creveld syndrome among the Pennsylvania Amish), ethnic (Tay-Sachs disease among Ashkenazi Jewish populations), or geographical (hemoglobinopathies among people with ancestors who lived in malarial regions). To the extent that ancestry corresponds with racial or ethnic groups or subgroups, the incidence of monogenic diseases can differ between groups categorized by race or ethnicity, and health-care professionals typically take these patterns into account in making diagnoses.

Even with common diseases involving numerous genetic variants and environmental factors, investigators point to evidence suggesting the involvement of differentially distributed alleles with small to moderate effects. Frequently cited examples include hypertension (Douglas et al. 1996), diabetes (Gower et al. 2003), obesity (Fernandez et al. 2003), and prostate cancer (Platz et al. 2000). However, in none of these cases has allelic variation in a susceptibility gene been shown to account for a significant fraction of the difference in disease prevalence among groups, and the role of genetic factors in generating these differences remains uncertain (Mountain and Risch 2004).

The Allelic Architecture of Disease

The genetic architecture of common diseases is an important factor in determining the extent to which patterns of genetic variation influence group differences in health outcomes (Reich and Lander 2001; Pritchard and Cox 2002; Smith and Lusis 2002). According to the common disease/common variant hypothesis, common variants present in the ancestral population before the dispersal of modern humans from Africa play an important role in human diseases (Goldstein and Chikhi 2002). Genetic variants associated with Alzheimer disease, deep venous thrombosis, Crohn disease, and type 2 diabetes appear to adhere to this model (Lohmueller et al. 2003). However, the generality of the model has not yet been established and, in some cases, is in doubt (Weiss and Terwilliger 2000; Pritchard and Cox 2002; Cardon and Abecasis 2003). Some diseases, such as many common cancers, appear not to be well described by the common disease/common variant model (Kittles and Weiss 2003; Wiencke 2004).

Another possibility is that common diseases arise in part through the action of combinations of variants that are individually rare (Pritchard 2001; Cohen et al. 2004). Most of the disease-associated alleles discovered to date have been rare, and rare variants are more likely than common variants to be differentially distributed among groups distinguished by ancestry (Risch et al. 2002; Kittles and Weiss 2003). However, groups could harbor different, though perhaps overlapping, sets of rare variants, which would reduce contrasts between groups in the incidence of the disease.

The number of variants contributing to a disease and the interactions among those variants also could influence the distribution of diseases among groups. The difficulty that has been encountered in finding contributory alleles for complex diseases and in replicating positive associations suggests that many complex diseases involve numerous variants rather than a moderate number of alleles, and the influence of any given variant may depend in critical ways on the genetic and environmental background (Risch 2000; Weiss and Terwilliger 2000; Altmüller et al. 2001; Hirschhorn et al. 2002). If many alleles are required to increase susceptibility to a disease, the odds are low that the necessary combination of alleles would become concentrated in a particular group purely through drift (Cooper 2004).

Population Substructure in Genetics Research

One area in which racial and ethnic categories can be important considerations in genetics research is in controlling for confounding between population substructure, environmental exposures, and health outcomes. Association studies can produce spurious results if cases and controls have differing allele frequencies for genes that are not related to the disease being studied (Cardon and Palmer 2003; Marchini et al. 2004), although the magnitude of this problem in genetic association studies is subject to debate (Thomas and Witte 2002; Wacholder et al. 2002). Various methods have been developed to detect and account for population substructure (Morton and Collins 1998; Hoggart et al. 2003), but these meth-

ods can be difficult to apply in practice (Freedman et al. 2004).

Population substructure also can be used to advantage in genetic association studies. For example, populations that represent recent mixtures of geographically separated ancestral groups can exhibit longer-range linkage disequilibrium between susceptibility alleles and genetic markers than is the case for other populations (Hoggart et al. 2004; Patterson et al. 2004; Smith et al. 2004; McKeigue 2005). Genetic studies can use this admixture linkage disequilibrium to search for disease alleles with fewer markers than would be needed otherwise. Association studies also can take advantage of the contrasting experiences of racial or ethnic groups, including migrant groups, to search for interactions between particular alleles and environmental factors that might influence health (Chaturvedi 2001; Collins et al. 2003).

Conclusions

When deciding whether and how to use racial, ethnic, and ancestral categories in research, geneticists face conflicting demands. On the one hand, many observers have made powerful arguments in favor of reducing or even eliminating the use of racial or ethnic categories in genetics research (Fullilove 1998; Goodman 2000; Lee et al. 2001; Braun 2002; Duster 2003, 2005; Stevens 2003; Kahn 2004; Sankar et al. 2004; Ossorio and Duster 2005). These observers point out that the use of these categories reinforces the widespread impression that health inequities arise through the action of genetic differences and independent of socially mediated mechanisms. In this way, genetics research that involves making population comparisons can inaccurately stereotype racial and ethnic groups, both by implying that such groups are clearly delineated and by associating health outcomes with all individuals in those groups rather than with only those individuals who exhibit the outcome. Furthermore, according to critics, an overemphasis on the genetic component of health differences shifts attention and resources away from established contributors to health disparities—in particular, the differences in treatment and socioeconomic disadvantages that disproportionately affect minority groups (Sankar et al. 2004). Genetics research offers no evidence that any one group is superior or inferior to any other, although some individuals continue to try to distort genetic findings to buttress prejudiced outlooks. Biomedical research that accentuates genetic differences among groups, say critics of this research, is as conceptually flawed as the race science of the 19th century (Bhopal 1997).

On the other hand, race and ethnicity are such prominent aspects of many societies that it is difficult, and often inadvisable, to ignore them in genetics research. The members of these groups can have widely disparate economic, social, and psychological experiences and can be exposed to very different environments as a consequence of their membership in a particular group. These differential experiences and environmental exposures can be used to investigate the biological mechanisms that contribute to health disparities among groups (LaVeist 1996). In addition, self-identified race, ethnicity, or ancestry can provide measures of population substructure that help avoid false-positive results in association studies.

One way for geneticists to ease the dilemma they face is to try to move beyond racial, ethnic, or ancestral categories in their work (Ota Wang and Sue 2005; Shields et al. 2005). Rather than using racial, ethnic, or ancestral labels as proxies for much more detailed social, economic, environmental, biological, or genetic factors, researchers can try to measure these factors directly. For example, controlling for socioeconomic status by use of census tract data can substantially reduce the excess mortality risk observed in disadvantaged minority populations (Krieger et al. 2005). Similarly, genotyping to estimate biogeographical ancestry can be a better control for population substructure than self-identified race, ethnicity, or ancestry (Shields et al. 2005).

When the use of racial or ethnic categories in research is deemed necessary, researchers can avoid overgeneralization by using labels that are as specific as possible. Today many genetic investigations label populations with the same loose terms used by the public (Sankar and Cho 2002; Clayton 2003; Collins 2004; Comstock et al. 2004). But labels such as "Hispanic," "Black," "Mexican American," "White," "Asian," "European," or "African" can have ambiguous or contradictory meanings among researchers, research subjects, and the general public. Use of such broad labels without careful definitions can impair scientific understanding and imply that distinctions between socially defined populations are genetically well established. Genetics researchers often rely on the categories specified in the U.S. census—encouraged by regulations that urge diversity of study populations—but these categories are used today mainly for administrative and social purposes and were not designed for genetics research. Even when the census categories are used to select research subjects to ensure diversity, researchers can analyze their results using more-specific labels that are closely tied to the scientific questions being asked (Kaplan and Bennett 2003). For example, labels based on biogeographical ancestry may be suited for many genetics studies, socially based labels may be more appropriate for health disparities and clinical research, and both types of information may be valuable for studies of some gene-environment interactions.

Individuals can be assigned to specific population categories in a number of ways, with the most appropriate

way, again, depending on the research question being investigated (Foster and Sharp 2004; International HapMap Consortium 2004). Research subjects can be asked to identify themselves with geographical or cultural populations, which may be defined by the researcher or by the local communities within which the research is being conducted. Communities and researchers can choose categories together through a consultative or engagement process between researchers and the community (Foster et al. 1999; Condit et al. 2002). Categorical systems also can include the possibility of simultaneous, multiple-group memberships in groups at higher or lower levels of organization.

A number of journals, including *Nature Genetics* (Anonymous 2000), *Archives of Pediatrics & Adolescent Medicine* (Rivara and Finberg 2001), and the *British Medical Journal* (Anonymous 1996), have separately issued guidelines stating that researchers should carefully define the terms they use for populations, and some journals have asked researchers to justify their use of racial or ethnic groups in research. But enforcement of these guidelines has been uneven, and compliance will continue to be spotty without greater awareness among researchers of the difficulties and risks involved in defining populations (Sankar and Cho 2002; Anonymous 2004).

Efforts to move past the use of racial and ethnic categories in genetics research often will require consideration of a very broad range of additional variables (Chakravarti and Little 2003). These variables will differ from study to study, but even a partial list includes racism and discrimination, socioeconomic status, social class, personal or family wealth, environmental exposures, insurance status, age, diet and nutrition, health beliefs and practices, educational level, language spoken, religion, tribal affiliation, country of birth, parents' country of birth, length of time in the country of residence, and place of residence along with genetic variation (Kaplan and Bennett 2003). Research that successfully integrates such a wide range of variables will require the collaboration of individuals with many different disciplinary backgrounds (Bonham et al. 2005).

A particular challenge for interdisciplinary teams will be designing their studies and reporting their results in ways that convey to the public the complexities of biological systems (Weiss 1998; Clark 2002; Chakravarti and Little 2003). Within the highly interconnected network of factors involved in complex diseases, the influence of any given allele likely will depend on past and current biological and environmental contexts, which often will make it difficult to demonstrate that a given variant directly "causes" a particular condition (Weatherall 1999; Page et al. 2003). Growing appreciation of the ways in which gestational influences (Sallout and Walker 2003), childhood illnesses (Gluckman and Hanson 2004), obesity (Calle and Kaaks 2004), exposure to toxins (Whyatt et al. 2004), stress (Wallace et al. 2004), and other factors influence later illnesses highlights the multiple interconnections among biological mechanisms, environmental influences, and chance events (Shostak 2003).

Despite this complexity, genetics researchers have a unique opportunity to reduce at least some of the confusion and controversy surrounding the issues of race, ethnicity, ancestry, and health. They can demonstrate the irrelevance of racial and ethnic labels for pursuing many research questions and health improvement objectives—for example, by clarifying the many ways in which environmental factors that extend across groups interact with biological processes to produce common diseases (Lin and Kelsey 2000; Rotimi 2004). By emphasizing the close genetic affinities between members of different groups, researchers can reduce the widespread misconception that substantial genetic differences separate groups (Wilson et al. 2001; Olson 2002; Jorde and Wooding 2004). As the complex origins of human traits, behaviors, and diseases slowly are unraveled, how genetics research is conducted could influence whether racial and ethnic discrimination increases or decreases over time.

Acknowledgments

The Race, Ethnicity, and Genetics Working Group of the National Human Genome Research Institute appreciates the valuable input received on earlier drafts of this paper from Michael Bamshad, Wylie Burke, Mildred Cho, Troy Duster, Sara Hull, Lynn Jorde, Jeff Long, Jeff Murray, and Ken Weiss.

References

Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet 69:936–950

American Association of Physical Anthropologists (1996) AAPA statement on biological aspects of race. Am J Phys Anthropol 101:569–570

Anonymous (1996) Style matters: ethnicity, race, and culture: guidelines for research, audit, and publication. BMJ 312: 1094

——— (2000) Census, race, and science. Nat Genet 24:97–98

——— (2004) The unexamined "Caucasian." Nat Genet 36: 541

Aoki K (2002) Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. Ann Hum Biol 29: 589–608

Bamshad M, Wooding S, Salisbury BA, Stephens JC (2004) Deconstructing the relationship between genetics and race. Nat Rev Genet 5:598–609

Bamshad M, Wooding SP (2003) Signature of natural selection in the human genome. Nat Rev Genet 4:99–111

Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA,

Jorde LB (2003) Human population genetic structure and inference of group membership. Am J Hum Genet 72:578–589
Banton M (1977) The idea of race. Westview Press, Boulder
Bhopal R (1997) Is research into ethnicity and health racist, unsound, or important science? BMJ 314:1751–1756
Bonham V, Warshauer-Baker E, Collins FS (2005) The complexity of the constructs. Am Psychol 60:9–15
Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, Spooner NA (2003) New ages of human occupation and climatic change at Lake Mungo, Australia. Nature 421:837–840
Braun L (2002) Race, ethnicity, and health: can genetics explain disparities? Perspect Biol Med 45:159–174
Brodwin P (2002) Genetics, identity, and the anthropology of essentialism. Anthropol Quart 75:323–330
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
Calafell F (2003) Classifying humans. Nat Genet 33:435–436
Calle EE, Kaaks R (2004) Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer 4:579–591
Cardon LR, Abecasis GR (2003) Using haplotype blocks to map human complex trait loci. Trends Genet 19:135–140
Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598–604
Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. Nat Genet Suppl 33:266–275
Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
Chakravarti A, Little P (2003) Nature, nurture and human disease. Nature 421:412–424
Chaturvedi N (2001) Ethnicity as an epidemiological determinant—crudely racist or crucially important? Int J Epidemiol 30:925–927
Clark ME (2002) In search of human nature. Routledge, New York
Clayton EW (2003) The complex relationship of genetics, groups, and health: what it means for public health. J Law Med Ethics 30:290–297
Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305:869–872
Collins FS (2004) What we do and don't know about "race," "ethnicity," genetics and health at the dawn of the genome era. Nat Genet 36:S13–S15
Collins FS, Green ED, Guttmacher AE, Guyer MS, for the US National Human Genome Research Institute (2003) A vision for the future of genomics research. Nature 422:835–847
Comstock RD, Castillo EM, Lindsay SP (2004) Four-year review of the use of race and ethnicity in epidemiologic and public health research. Am J Epidemiol 159:611–619
Condit C, Templeton A, Bates BR, Bevan JL, Harris AT (2003) Attitudinal barriers to delivery of race-targeted pharmacogenomics among informed lay persons. Genet Med 5:385–392
Condit CM, Parrott R, Harris TM (2002) Lay understandings of the relationship between race and genetics: development of a collectivized knowledge through shared discourse. Public Understand Sci 11:373–387
Cooper RS (2004) Genetic factors in ethnic disparities in health. In: Anderson NB, Bulatao RA, Cohen B (eds) Critical perspectives on racial and ethnic differences in health in later life. National Academy Press, Washington, DC, pp 267–309
Cooper RS, Kaufman JS, Ward R (2003) Race and genomics. N Engl J Med 348:1166–1170
Cooper RS, Rotimi CN, Kaufman JS, Owoaje EE, Fraser H, Forrester T, Wilks R, Riste LK, Cruickshank JK (1997) Prevalence of NIDDM among populations of the African diaspora. Diabetes Care 20:343–348
Cornell S, Hartmann D (1998) Ethnicity and race: making identities in a changing world. Pine Forge Press, Thousand Oaks, CA
Davis FJ (2001) Who is black? One nation's definition. Pennsylvania State University Press, University Park
Dikötter F (1992) The discourse of race in modern China. Stanford University Press, Stanford
Douglas JG, Thibonnier M, Wright JT (1996) Essential hypertension: racial/ethnic differences in pathophysiology. J Assoc Acad Minor Phys 7:16–21
Duster T (2003) Backdoor to eugenics, 2nd ed. Routledge, New York
——— (2005) Race and reification in science. Science 307: 1050–1051
Ebersberger I, Metzler D, Schwarz C, Pääbo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 70:1490–1497
Elliott C, Brodwin P (2002) Identity and genetic ancestry tracing. BMJ 325:1469–1471
Eswaran V (2002) A diffusion wave out of Africa: the mechanism of the modern human revolution? Curr Anthropol 43: 749–774
Fang J, Madhavan S, Bosworth W, Alderman MH (1998) Residential segregation and mortality in New York City. Soc Sci Med 47:469–476
Fernandez JR, Shriver MD, Beasley TM, Rafla-Demetrious N, Parra E, Albu J, Nicklas B, Ryan AS, McKeigue PM, Hoggart CL, Weinsier RL, Allison DB (2003) Association of African genetic admixture with resting metabolic rate and obesity among women. Obes Res 11:904–911
Fischer A, Wiebe V, Pääbo S, Przeworski M (2004) Evidence for a complex demographic history of chimpanzees. Mol Biol Evol 21:799–808
Foster MW, Sharp RR (2004) Beyond race: towards a whole-genome perspective on human populations and genetic variation. Nat Rev Genet 5:790–796
Foster MW, Sharp RR, Freeman WL, Chino M, Bernsten D, Carter TH (1999) The role of community review in evaluating the risks of human genetic variation research. Am J Hum Genet 64:1719–1727
Franzini L, Ribble JC, Keddie AM (2001) Understanding the Hispanic paradox. Ethn Dis 11:496–518
Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault

AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet 36:388–393

Fullilove M (1998) Abandoning "race" as a variable in public health research—an idea whose time has come. Am J Public Health 88:1297–1298

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Gluckman PD, Hanson MA (2004) Living with the past: evolution, development, and patterns of disease. Science 305: 1733–1736

Goldenberg DM (2003) The curse of ham: race and slavery in early Judaism, Christianity, and Islam. Princeton University Press, Princeton

Goldstein DB, Chikhi L (2002) Human migrations and population structure: what we know and why it matters. Ann Rev Genomics Hum Genet 3:129–152

Goodman AH (2000) Why genes don't count (for racial differences in health). Am J Public Health 90:1699–1702

Gossett TF (1997) Race: the history of an idea in America, 2nd ed. Oxford University Press, New York

Gower BA, Fernandez JR, Beasley TM, Shriver MD, Goran MI (2003) Using genetic admixture to explain racial differences in insulin-related phenotypes. Diabetes 52:1047–1051

Guthrie RD (1996) The mammoth steppe and the origin of mongoloids and their dispersal. In: Akazawa T, Szathmary E (eds) Prehistoric Mongoloid dispersals. Oxford University Press, New York, pp 172–186

Hannaford I (1996) Race: the history of an idea in the West. Johns Hopkins University Press, Baltimore

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

Harpending H, Rogers A (2000) Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet 1:361–385

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Hawks J, Hunley K, Lee SH, Wolpoff M (2000) Population bottlenecks and Pleistocene human evolution. Mol Biol Evol 17:2–22

Hayes-Bautista DE, Chapa J (1987) Latino terminology: conceptual bases for standardized terminology. Am J Public Health 77:61–68

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4:45–61

Hoerder D (2002) Cultures in contact: world migrations in the second millennium. Duke University Press, Durham, NC

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM (2003) Control of confounding of genetic associations in stratified populations. Am J Hum Genet 72:1492–1504

Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM (2004) Design and analysis of admixture mapping studies. Am J Hum Genet 74:965–978

Horowitz DL (2001) The deadly ethnic riot. University of California Press, Berkeley

Hummer RA, Benjamins MR, Rogers RG (2004) Racial and ethnic disparities in health and mortality among the U.S. elderly population. In: Anderson NB, Bulatao RA, Cohen B (eds) Critical perspectives on racial and ethnic differences in health in later life. National Academy Press, Washington, DC, pp 53–94

Hutchinson J, Smith AD (eds) (1996) Ethnicity. Oxford University Press, New York

Huxley J, Haddon AC (1936) We Europeans: a survey of racial problems. Harper, New York

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

International HapMap Consortium (2003) The International HapMap Project. Nature 426:789–796

——— (2004) Integrating ethics and science in the International HapMap Project. Nat Rev Genet 5:467–475

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Isaac B (2004) The invention of racism in classical antiquity. Princeton University Press, Princeton

Jablonski NG (2004) The evolution of human skin and skin color. Annu Rev Anthropol 33:585–623

Jorde LB, Bamshad M, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays 20:126–136

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000*a*) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988

Jorde LB, Watkins WS, Kere J, Nyman D, Eriksson AW (2000*b*) Gene mapping in isolated populations: new roles for old friends? Hum Hered 50:57–65

Jorde LB, Wooding SP (2004) Genetic variation, classification, and "race." Nat Genet Rev Suppl 36:S28–S33

Kaessmann H, Heissig F, von Haeseler A, Pääbo S (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78–81

Kaessmann H, Wiebe V, Weiss G, Pääbo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. Nat Genet 27:155–156

Kahn J (2004) How a drug becomes "ethnic": law, commerce, and the production of racial categories in medicine. Yale J Health Policy Law Ethics 4:1–46

Kaplan EA (2003) Black like I thought I was. LA Weekly, October 7 (http://www.alternet.org/story/16917; accessed August 8, 2005)

Kaplan JB, Bennett T (2003) Use of race and ethnicity in biomedical publication. JAMA 289:2709–2716
Keita SOY, Kittles RA (1997) The persistence of racial thinking and the myth of racial divergence. Am Anthropol 99:534–544
Kevles DJ (1985) In the name of eugenics. Knopf, New York
King M-C, Motulsky AG (2002) Mapping human history. Science 298:2342–2343
Kittles RA, Weiss KM (2003) Race, ancestry, and genes: implications for defining disease risk. Annu Rev Genomics Hum Genet 4:33–67
Klein RG (1999) The human career: human biological and cultural origins, 2nd ed. University of Chicago Press, Chicago
Kressin NR, Chang BH, Hendricks A, Kazis LE (2003) Agreement between administrative data and patients' self-reports of race/ethnicity. Am J Public Health 93:1734–1739
Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV (2005) Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. Am J Public Health 95:312–323
Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90:19–30
Lahr MM (1996) The evolution of modern human diversity: a study of cranial variation. Cambridge University Press, Cambridge, United Kingdom
Lahr MM, Foley RA (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol Suppl 27:137–176
LaVeist TA (1996) Why we should continue to study race…but do a better job: an essay on race, racism and health. Ethn Dis 6:21–29
——— (ed) (2002) Race, ethnicity, and health. Jossey-Bass, San Francisco
Lee SS, Mountain J, Koenig BA (2001) The meanings of "race" in the new genomics: implications for health disparities research. Yale J Health Policy Law Ethics 1:33–75
Lewis B (1990) Race and slavery in the Middle East. Oxford University Press, New York
Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398
Li WH, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129:513–523
Lieberman DE, McBratney BM, Krovitz G (2002) The evolution and development of cranial form in *Homo sapiens*. Proc Natl Acad Sci USA 99:1134–1139
Lieberman L (2001) How "Caucasoids" got such big crania and why they shrank: from Morton to Rushton. Curr Anthropol 42:69–95
Lin SS, Kelsey JL (2000) Use of race and ethnicity in epidemiologic research: concepts, methodological issues, and suggestions for research. Epidemiol Rev 22:187–202
Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182
Long JC, Kittles RA (2003) Human genetic diversity and the nonexistence of biological races. Hum Biol 75:449–471
Mahoney MC, Michalek AM (1998) Health status of American Indians/Alaska natives: general patterns of mortality. Fam Med 30:190–195
Marchini J, Cardon LR, Phillips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. Nat Genet 36:512–517
Marks J (1995) Human biodiversity: genes, race, and history. Aldine de Gruyter, New York
Mays VM, Ponce NA, Washington DL, Cochran SD (2003) Classification of race and ethnicity: implications for public health. Annu Rev Public Health 24:83–110
McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433:733–736
McKeigue PM (2005) Prospects for admixture mapping of complex traits. Am J Hum Genet 76:1–7
Meltzer M (1993) Slavery: a world history, rev ed. DaCapo Press, Cambridge, MA
Miller GH, Magee JW, Johnson BJ, Fogel ML, Spooner NA, McCulloch MT, Ayliffe LK (1999) Pleistocene extinction of *Genyornis newtoni*: human impact on Australian megafauna. Science 283:205–208
Mörner M (1967) Race mixture in the history of Latin America. Little, Brown, Boston
Morton NE, Collins A (1998) Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci USA 95:11389–11393
Mosse GL (1985) Toward the final solution, 2nd ed. University of Wisconsin Press, Madison
Mountain JL, Risch N (2004) Assessing genetic contributions to phenotypic differences among "racial" and "ethnic" groups. Nat Genet Suppl 36:S48–S53
Nobles M (2000) Shades of citizenship: race and the census in modern politics. Stanford University Press, Stanford
Nordborg M (1998) On the probability of Neanderthal ancestry. Am J Hum Genet 63:1237–1240
Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90
Ohno S (1996) The Malthusian parameter of ascents: what prevents the exponential increase of one's ancestors? Proc Natl Acad Sci USA 93:15276–15278
Olson S (2002) Mapping human history. Houghton Mifflin, Boston
Oppenheimer GM (2001) Paradigm lost: race, ethnicity, and the search for a new population taxonomy. Am J Public Health 91:1049–1055
Ossorio P, Duster T (2005) Controversies in biomedical, behavioral, and forensic sciences. Am Psychol 60:115–128
Ota Wang V, Sue S (2005) In the eye of the storm: race and genomics in research and practice. Am Psychol 60:37–45
Pääbo S (2003) The mosaic that is our genome. Nature 421: 409–412
Page GP, George V, Go RC, Page PZ, Allison DB (2003) "Are we there yet?": Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am J Hum Genet 73:711–719
Parra EJ, Kittles RA, Shriver MD (2004) Implications of correlations between skin color and genetic ancestry for biomedical research. Nat Genet 36:S54–S60
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 SD (2003) Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 100:177–182

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D (2004) Methods for high-density admixture mapping of disease genes. Am J Hum Genet 74:979–1000

Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC (2004) Information on ancestry from genetic markers. Genet Epidemiol 26:305–315

Platz EZ, Rimm EB, Willett WC, Kantoff PW, Giovannucci E (2000) Racial variation in prostate cancer incidence and in hormonal system markers among male health professionals. J Natl Cancer Inst 92:2009–2017

Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124–137

Pritchard JK, Cox NJ (2002) The allelic architecture of human disease genes: common disease-common variant…or not? Hum Mol Genet 11:2417–2423

Provine WB (1986) Geneticists and race. Am Zoologist 26: 857–887

Rawlings JS, Weir MR (1992) Race- and rank-specific infant mortality in a US military population. Am J Dis Child 146: 313–316

Rees JL (2003) Genetics of hair and skin color. Annu Rev Genet 37:67–90

Reich DA, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17:502–510

Relethford JH (2002) Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393–398

Risch N (2000) Searching for the genetic determinants in a new millennium. Nature 405:847–856

Risch N, Burchard E, Ziv E, Tang H (2002) Categorization of humans in biomedical research: genes, race and disease. Genome Biol 3 (http://genomebiology.com/2002/3/7/comment/2007) (electronically published July 1, 2002; accessed August 25, 2005)

Rivara F, Finberg L (2001) Use of the terms race and ethnicity. Arch Pediatr Adolesc Med 155:119

Rohde D, Olson S, Chang J (2004) Modeling the recent common ancestry of all living humans. Nature 431:562–566

Roseman CC (2004) Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. Proc Natl Acad Sci USA 101:12824–12829

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Rotimi CN (2004) Are medical and nonmedical uses of large-scale genomic markers conflating genetics and "race"? Nat Genet 36:S43–S47

Sallout B, Walker M (2003) The fetal origin of adult diseases. J Obstet Gynaecol 23:555–560

Sankar P, Cho MK (2002) Toward a new vocabulary of human genetic variation. Science 298:1337–1338

Sankar P, Cho MK, Condit DM, Hunt LM, Koenig B, Marshall P, Lee SS, Spicer P (2004) Genetic research and health disparities. JAMA 291:2985–2989

Satta Y, Takahata N (2002) Out of Africa with regional interbreeding? Modern human origins. Bioessays 24:871–875

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G G, Pääbo S (2004) No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol 2:313–317

Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. Genome Res 14:1679–1685

Shields AE, Fortun M, Hammonds EM, King PA, Lerman C, Rapp R, Sullivan PF (2005) The use of race variables in genetic studies of complex traits and the goal of reducing health disparities: a transdisciplinary perspective. Am Psychol 60:77–103

Shipler D (1997) A country of strangers: blacks and whites in America. Knopf, New York

Shostak S (2003) Locating gene-environment interaction: at the intersections of genetics and public health. Soc Sci Med 56:2327–2342

Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA (2003) Skin pigmentation, biogeographical ancestry, and admixture mapping. Hum Genet 112:387–399

Smedley A (1999) Race in North America: origin and evolution of a worldview, 2nd ed. Westview Press, Boulder

Smelser N, Wilson WJ, Mitchell F (eds) (2001) America becoming: racial trends and their consequences. Vol 2. National Academy Press, Washington, DC

Smith DJ, Lusis AJ (2002) The allelic structure of common disease. Hum Mol Genet 11:2455–2461

Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, et al (2004) A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet 74:1001–1013

Snowden FM (1983) Before color prejudice: the ancient view of blacks. Harvard University Press, Cambridge, MA

Spickard PR (1992) The illogic of American racial categories. In: Root MPP (ed) Racially mixed people in America. Sage, Newbury Park, CA, pp 12–23

Stanton W (1960) The leopard's spots: scientific attitudes toward race in America, 1815–1859. University of Chicago Press, Chicago

Stevens J (2003) Racial meanings and scientific methods: changing policies for NIH-sponsored publications reporting human variation. J Health Polit Policy Law 28:1033–1087

Stringer C (2002) Modern human origins: progress and prospects. Philos Trans R Soc Lond B Biol Sci 357:563–579

Sturm RA, Teasdale RD, Box NF (2001) Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene 277:49–62

Swisher CC 3rd, Curtis GH, Jacob T, Getty AG, Suprijo A, Widiasmoro (1994) Age of the earliest known hominids in Java, Indonesia. Science 263:1118–1121

Takahata N, Lee S, Satta Y (2001) Testing multiregionality of modern human origins. Mol Biol Evol 18:172–183

Takaki R (1993) A different mirror: a history of multicultural America. Little, Brown, Boston

Tang H, Quertermous T, Rodriguez B, Kardia SLR, Zhu X, Brown A, Pankow JS, Province MA, Hunt SC, Boerwinkle E, Schork NJ, Risch NJ (2005) Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. Am J Hum Genet 76:268–275

Tate SK, Goldstein DB (2004) Will tomorrow's medicines work for everyone? Nat Genet 36:S34–S42

Templeton AR (1998) Human races: a genetic and evolutionary perspective. Am Anthropol 100:632–650

——— (2002) Out of Africa again and again. Nature 416: 45–51

Thomas DC, Witte JS (2002) Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 11:505–512

Tishkoff SA, Kidd KK (2004) Implications of biogeography of human populations for "race" and medicine. Nat Genet 36: S21–S27

Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu R-b, Deinard AS, Sirugo G, Jenkins T, Kidd KK, Clark AG (2000) Short tandem-repeat polymorphism/*Alu* haplotype variation at the PLAT locus: implications for modern human origins. Am J Hum Genet 67:901–925

Tishkoff SA, Verrelli B (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340

Tishkoff SA, Williams SM (2002) Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 3:611–621

Todorov T (1993) On human diversity. Harvard University Press, Cambridge, MA

Trinkaus E, Milota S, Rodrigo R, Mircea G, Moldovan O (2003) An early modern human from the Pestera cu Oase, Romania. Proc Natl Acad Sci USA 100:11231–11236

Underhill PA, Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Vega WA, Amaro H (1994) Latino outlook: good health, uncertain prognosis. Annu Rev Public Health 15:39–67

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al (2001) The sequence of the human genome. Science 291:1304–1351

Wacholder S, Rothman N, Caporaso N (2002) Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 11:513–520

Wall JD (2000) Detecting ancient admixture in humans using sequence polymorphism data. Genetics 154:1271–1279

Wallace R, Wallace D, Wallace RG (2004) Coronary heart disease, chronic inflammation, and pathogenic social hierarchy: a biological limit to possible reductions in morbidity and mortality. J Natl Med Assoc 96:609–619

Waters M (1990) Ethnic options: choosing identities in America. University of California Press, Berkeley

Weatherall D (1999) From genotype to phenotype: genetics and medical practice in the new millennium. Philos Trans R Soc Lond B Biol Sci 354:1995–2010

Weiss KM (1998) Coming to terms with human variation. Annu Rev Anthropol 27:273–300

Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? Nat Genet 26:151–157

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC (2003) Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. Nature 423:742–747

Whyatt RM, Rauh V, Barr DB, Camann DE, Andrews HF, Garfinkel R, Hoepner LA, Diaz D, Dietrich J, Reyes A, Tang D, Kinney PL, Perera FP (2004) Prenatal insecticide exposures and birth weight and length among an urban minority cohort. Environ Health Perspect 112:1125–1132

Wiencke JK (2004) Impact of race/ethnicity on molecular pathways in human cancer. Nat Rev Cancer 4:79–84

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

Wolpoff M, Caspari R (1997) Race and human evolution: a fatal attraction. Simon & Schuster, New York

Wolpoff M, Hawks J, Frayer DW, Hunley K (2001) Modern human ancestry at the peripheries: a test of the replacement theory. Science 291:293–297

Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH (2002) Larger genetic differences within Africans than between Africans and Eurasians. Genetics 161:269–274

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH (2003) Low nucleotide diversity in chimpanzees and bonobos. Genetics 164:1511–1518

Ziętkiewicz E, Yotova V, Gehl D, Wambach T, Arrieta I, Batzer M, Cole DEC, Hechtman P, Kaplan F, Modiano D, Moisan J-P, Michalski R, Labuda D (2003) Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am J Hum Genet 73:994–1015