

iv

Anthropology of Race  
*Genes, Biology, and Culture*

*Edited by John Hartigan*



School for Advanced Research Press  
*Santa Fe*

2013

5

Observations on the Tenacity  
of Racial Concepts in Genetics Research

Linda M. Hunt and Nicole Truesdell

The time-worn idea that the human species can be reasonably divided into biologically distinct races has long been rejected by anthropologists and many biologists. They argue that race is a social and historical fact rather than a biological reality, pointing out that there is more genetic variation within racial groups than between them. Racial identity may indeed have important biological implications affecting the health of racially labeled groups, but it is the social reality of race, rather than inherent biological group differences, that determines this (American Anthropological Association 1997; Goodman 2000; Lewontin 1972).

Anthropologists have been especially avid in teaching both the lay public and biomedical researchers that race is a social rather than biological phenomenon. For example, the American Anthropological Association has undertaken an ambitious educational campaign called "Race: Are We So Different?" that includes an interactive website and traveling museum exhibitions. In medical and health research journals, other anthropologists have been diligently publishing strong critiques of how medical and genetics research uses race (see, for example, Gravlee 2009; Hunt and Megyesi 2008a; Lee 2008; Sankar et al. 2007). But racial thinking continues to permeate US public discourse and health-related research studies.

As a medical anthropologist working closely with biomedical researchers, the first author has had many opportunities to preach the gospel of "Race Is a Social Construction" to colleagues who have been trained in disciplines in which this is a relatively novel idea. Throughout their careers, they have regularly encountered race as a matter-of-fact descriptor of patient populations and have dutifully generated and followed race-based risk profiles and treatment recommendations. In one notable situation, for three years she met regularly over coffee with a physician-colleague to discuss the concept of race in health research and health care. Gradually, the physician became an enthusiastic convert, embracing the notion that race is a social category, not a biological one. He even began to draft an article for a journal in his medical discipline, challenging the notion of biological race. But then one day he ran across a report of a new study reporting that a particular heart medication worked differently with blacks than with whites, which it attributed to differences in how blacks and whites metabolize a certain chemical. The next time they met, his conviction that race is not biological had seemingly evaporated: "Yeah, but what about these findings? There clearly are biological differences between blacks and whites." Left with an untenable choice between accepting or rejecting the existence of observed variation, he reverted to the familiar notion that racial groups are biologically meaningful. And thus, in a single stroke, the hard-won rejection of biological race was abandoned, and the physician's view of race was right back where it had started: race is a rough but useful indicator of ancestry, a convenient way to gauge genetic heritage, with important implications for disease susceptibility and drug response.

This scenario is but one instance of a much larger process that seems to be repeating itself across the broader field of biomedical and genetics research. In a cyclical fashion, the idea that the human species comprises a handful of biologically distinct racial groups emerges, is challenged, is revised, and then re-emerges in nearly its original form. The concept of biological races has been experiencing a reinvigoration in the sciences as the prodigious wave of genetics research following the Human Genome Project increasingly identifies various patterns of human genetic variation. A profusion of recent research is framed in terms presuming that the complexity of human variation can be meaningfully captured in four or five continental racial groups: Europeans, Asians, Africans, and Native Americans. For example, a recent Medline search for genetics research using these racial terms yielded nearly four thousand articles for 2011 alone.

What is there about the concept of race that makes it so tenacious as an idiom for classifying human variation? In this chapter, we review ways race

is currently being conceptualized and operationalized by genetics researchers and consider some roots and consequences of these concepts and practices. Like any science, genetics research is by no means divorced from its sociocultural context but is produced and interpreted through cultural lenses (Berger and Luckmann 1990; Keita and Kittles 1997). We argue that the idea of race persists not because it so accurately captures existing variation but because it draws upon a set of core concepts in Western culture: a Judeo-Christian notion of the primordial origins of human populations and a Eurocentric understanding of their geographic dispersion through time. These notions produce an unexamined lens through which disparate "populations" defined for diverse studies are readily mapped onto a small set of familiar racial groupings.

Efforts by social scientists to challenge the notion of biological race in such research have failed to effectively discourage these practices. By examining the basis of these widely accepted folk notions of group difference, we may come to better understand their tenacity and encourage a more productive discussion of human biological differences (Haslanger 2008) that does not revert so stubbornly to the typological thinking of traditional notions of race.

#### THE RISE AND FALL OF BIOLOGICAL RACES

C. Loring Brace observed that, prior to the time of the Renaissance, people traveled by foot or sailed close to the coastline, covering only about twenty-five miles in a day's journey, and as they moved from Europe to Africa, the differences they observed between groups of people were gradual. There was no sense that humans should be thought of as being a limited set of distinct subgroups. But, by the fifteenth century, with improvements in sailing ships, people were able to sail between great distances across the oceans, and differences between people at each end of the journey were quite striking. It was at this point that the idea of human "races" was first conceptualized (Brace 2005).

In the 1700s, at the time of vast European colonial expansion, the Swedish botanist Linnaeus undertook the monumental task of creating a taxonomy of all living creatures. In the tenth edition of his book *Systemae Naturae* (1758), Linnaeus presented a classification scheme for humans, based on physical characteristics and notions of continental boundaries prevalent at the time: Europeus, "white"; Americanus, "red"; Asiaticus, "sallow"; and Africanus, "black" (Graves 2001; Smedley 2007). Somewhat embellished over time, Linnaeus's idea that humans can be reasonably divided into these four major racial groups has endured as a fundamental

concept to this day, providing the basis for much scientific inquiry into the nature of human variation (Sauer 1993).

This schema was uncritically accepted until the 1930s, when a number of academics began to systematically challenge the notion of biological races. Boas, Montagu, and others presented serious critiques of the idea that races are biologically distinct, based on the extent and complexity of heterogeneity within supposed racial groups and on the principle of discordance, that the defining traits of racial groups are not consistently correlated with one another (Boas 1940; Brace 1964; Livingstone 1962; Montagu 1945).

Through the latter part of the twentieth century, in an era defined by post-Nazi social consciousness, the American civil rights movement, and anticolonial movements in Africa and Asia, many in the sciences were led to further reconsider the notion of race as a biological concept (Hirschman 2004; Smedley and Smedley 2005). With scientists' increasing knowledge of the complexity of human variation and its mechanisms and of human evolution and genetics, the concept of continental racial groups seemed certain to be replaced by more specific and sophisticated approaches. Some called for using breeding populations with sets of variably expressed genetic traits as the unit of study (Mukhopadhyay and Moses 1997; Richards 1997). Others argued for an emphasis on clines of variation, that is, recognizing that variation is gradual across a geographic range, rather than occurring in clearly bounded units (Livingstone 1962). Still others thought that the developments in evolutionary science would supplant the idea of the biological reality of continental races once and for all (Hiernaux 1964; Washburn 1963).

In 1972, many thought that Lewontin was putting the nails in the coffin of the concept of biological races with his oft cited study of genetic diversity in blood groups and antigens, reporting more variation within racial populations than between them. Subsequently, many have argued that when racially labeled groups are found to correlate well with genetic variation, it is primarily an artifact of selectively sampling from relatively isolated groups that are widely separated geographically. Differences are far less pronounced when sample sources are more evenly distributed geographically (Race, Ethnicity, and Genetics Working Group 2005). The completion of a draft of the human genome in 2000 further supported these earlier claims, showing virtually no genetic difference between races (McCann-Mortimer, Augoustinos, and Le Couteur 2004).

Despite the long history of scientific arguments and evidence against the notion, the claim that there are biologically distinct races has maintained a

tenacious if marginalized place in our public discourse. For example, recent works that promote this view include Jensen's 1970 studies of intelligence; Rushton's 1990s dogged rehashing of racial difference research; Rowe's psychological studies on children of mixed parentage; and Herrnstein and Murray's writings on the "Bell Curve" (Brace 2005).

### RESURRECTING RACE IN THE POST-GENOMIC ERA

Of late, the use of race/ethnicity as a biological variable has been the subject of much debate in both professional and popular media. On one side are those who defend the use of racial labels, claiming that these capture important hereditary differences between groups of people, which can usefully advance biomedical research, diagnosis, and treatment (Burchard et al. 2003; Risch et al. 2002; Rosenberg et al. 2002). On the other side are those who contend that these categories confuse more than clarify. They question not only the scientific merit of the categories themselves but also the potential dangers of promoting the erroneous notion that human racial taxonomies capture inherent biological variation (Braun 2004; Duster 2006; Feldman and Lewontin 2008; Hunt and Megyesi 2008a, 2008b).

Marks has noted that the revival of the race concept in genetics and biomedicine is particularly extraordinary, given that it requires "explicit rejection of decades of professional scholarship on the subject of human variation and the acceptance instead of common or folk knowledge" (2008:34). He argues that this trend is being fueled by the inherent conflict of interest associated with the profit motive, so intimately intertwined with current genetics research and the diagnostic and therapeutic panacea it promises. Others point to additional circumstances contributing to the reversal of the trend toward abandonment of the idea of biologically distinct races.

One circumstance is the establishment of requirements for reporting racial labels for research participants. Intending to ensure equity in the potential health benefits of publicly funded research, since 1993 the US federal government has required the inclusion of minorities in federally funded studies. The reporting format attached to this mandate uses the racial/ethnic categories of the Office of Management and Budget (OMB), which are also used in the US census (Office of Management and Budget 1997). The racial categories composing this schema are strikingly reminiscent of the ancient Linnacian taxonomy, though with some added nuances reflecting recent bureaucratic and political maneuvering: American Indian/Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; Black or African American; and White. (In addition to these racial labels,

there is a new, second layer of classification, the ethnic designations “Hispanic” and “Non-Hispanic.”)

Although the OMB categories were designed for purposes unrelated to the biological sciences, in day-to-day practice they are quickly becoming a focus of analysis in all manner of biological and genetics research, and scientists quite regularly offer biological and genetic interpretations of correlations involving these variables (Braun 2002; Outram and Ellison 2006). Kahn has pointed out that “when the federal government requires biomedical researchers and clinicians to import these social categories into explicitly biological or genetic contexts, it is creating a structural situation in which social categories of race and ethnicity may easily become confused and may be conflated with biological and genetic categories in day-to-day practice” (2006:1966).

Another circumstance contributing to the revival of biological race is the pervasive use of racial and/or continental labels in the burgeoning arena of human genetics science. The National Human Genome Research Institute’s (NHGRI) Human Genome Project, completed in 2000, was touted as having conclusively shown that racial groups are nearly identical genetically (Lee 2008). This spirit of de-racialized science was also evident in the design of the NHGRI’s DNA Polymorphism Discovery Resource (PDR), a national biobank established as part of a large-scale effort to discover medically significant SNPs (single nucleotide polymorphisms). The NHGRI designed the PDR to be “color-blind,” deliberately excluding racial and ethnic identifiers from the data. Samples were assembled with the intention of representing racial/ethnic diversity and were described with the familiar labels of European American, African American, Asian American, and Mexican American, but these labels were stripped from the data set, and researchers were strictly instructed to exclude racial and ethnic identity from their analysis (Collins, Brooks, and Chakravarti 1998). Lee argues that because the goal of genetic association studies is to find differences in the distribution of genetic markers, many found the de-racialized data frustrating and some claim that the repository has been underutilized precisely because this information has been excluded (Lee 2006, 2008).

With the development of high-throughput genotyping technology in recent years, there has been a proliferation of publicly available human genetics databases. The growth in this field is truly astounding. In less than twenty years, more than five hundred public databases have become available online (Tyshenko and Leiss 2005), and continental racial labels have found their way back into these databases in a prominent way, becoming

a hallmark of how they are organized. For example, the samples available through the widely used Coriell Cell Repositories are first organized at the continental level: North America/Caribbean, South America, Europe, Africa, Middle East, and Asia/Pacific. These are then broken down into a mixture of racial and geopolitical labels, such as Caucasian, African American, Amerindian, Mexican, Iberian, Greek, Africans north of the Sahara, Ashkenazi Jewish, and so forth (Coriell Institute 2008).

The routine use of racial/ethnic labels in this context, intermixed matter-of-factly with less controversial descriptors such as country of origin, sanctions the idea that continental racial groups are a legitimate unit of analysis in human genetics research and further aggravates the OMB problem outlined above (Kahn 2006; Lee 2006). This has the effect of not only conflating diverse criteria for classification into a single classificatory scheme (a point discussed in some detail below) but also promoting the illusion that these labels have the status of legitimate scientific categories.

#### EXAMINING RACIAL CONCEPTS AND PRACTICES

In order to more fully understand the specific ways that racial groups are being incorporated into current genetics studies, we have conducted a brief literature review of recent genetics articles using racial or continental population labels in reporting findings. Due to the preponderance of such articles, we limited our search to those published in one year, 2006, in a selection of five major genetics and medical journals (*American Journal of Human Genetics*, *Lancet*, *Journal of the American Medical Association*, *New England Journal of Medicine*, and *Nature Genetics*), yielding forty-two articles for our review.

We also conducted interviews with thirty genetics scientists regarding their understanding and use of racial/ethnic variables. These included a cross-section of human genetics researchers conducting research that used racial/ethnic variables as an integral part of their research design. This was a purposive, snowball sample of principle investigators with PhD and/or MD training. The research projects they were working on ranged from population modeling to linkage studies and focused on diseases ranging from rare inherited diseases to common chronic diseases. (For more detail on these interviews, see Hunt and Megyesi 2008b).

#### LABELS AND LABELING

Between the literature review and interviews, our analysis spans a wide variety of disciplines, such as molecular biology, endocrinology, epidemiology, biostatistics, and human genetics and a diverse sampling of types

of research projects, such as studies of human genetic evolution, disease-gene associations, and hereditary illness and studies of the genetic basis of complex diseases. Of course, researchers from different disciplines with different research goals will define study populations in fundamentally different ways. Even so, we found that researchers, in discussing and presenting their studies, quite often revert to the customary racial/ethnic labels, contrasting, for example, Asians and Europeans or Blacks and Caucasians. Using these broad group terms to describe findings from objectively disparate groups promotes the appearance that these are somehow equivalent groups without meaningful justification for such a presumption. How is it that researchers using such different approaches so readily turn to these common yet dubious terms?

Analysis of our interviews and literature review indicates that these practices are facilitated by a general acceptance, when it comes to race, of imprecise definitions and inexplicit classification practices. Stanfield has argued that this lack of conceptual and methodological care is typically a feature of studies of race. He observes that when race is the subject, folk wisdom often takes precedence over scientific rigor and the rules of procedure and evidence are readily bent or ignored (Stanfield 1993).

Indeed, throughout our analysis we have been impressed at the ambiguous and unsystematic way racial/ethnic classifications are being handled by genetics scientists. The researchers we interviewed used common racial/ethnic labels most often to describe their samples. In the context of scientific research design, it is impressive to consider how strikingly diverse these common categories are. They mix and combine an impressive array of unrelated classification types, such as skin color, language, geographic or continental regions, and religious or linguistic heritage. Table 5.1 presents the labels that interviewees mentioned most commonly to describe their samples. The table also includes types of classifications that we suggest describe the characteristics on which they are based.

Considering these labels in this light, their arbitrariness is unmistakable. The types of characteristics they refer to are strikingly diverse. Because the labels are so familiar, their vagueness and inconsistency may not appear of any real concern. However, because they are not mutually exclusive and lack clear principles for their application, they are indeed deeply problematic in terms of organizing scientific analysis.

To systematically apply categories such as these, which draw on multiple and overlapping criteria, would require careful procedures for determining which characteristics to prioritize in classifying any given case. However, the researchers we interviewed described virtually no explicit procedures

TABLE 5.1

*Racial/Ethnic Classification Terms Most Commonly Used by 30 Genetics Researchers and Types of Classification That They Appear to Represent*

Racial/Ethnic Terms	Types of Classification
Caucasian	Skin color of geographic origin
White	Skin color
Non-Hispanic White	Language and skin color
Jewish/Ashkenazi Jews	Religion and geographic origin
Asian	Continental origin
Asian American	Continental ancestral origin and geographic region
African American	Continental ancestral origin and geographic region
West African	Geographic region
Afro-Caribbean	Continental ancestral origin and geographic region
Hispanic/Latino	Language
Mexican	Country
Mexican American	Country of ancestral origin and geographic region
Native American	Ancestral group membership

or principles for determining how to classify individual cases. When they did cite a specific procedure to classify cases, most often it was the inherently idiosyncratic practice of "self-identification." (For a more complete discussion of these issues, see Hunt and Megyesi 2008b.) Because personal racial/ethnic identities are amorphous, multiple, and fluid, altering with changes in economic, geographic, and social contexts (Berry 1993; Hunt, Schneider, and Comer 2004), rather than help clarify the application of already amorphous labels, relying on self-identification would further confound the already muddled classifications.

#### POLICY AND PRACTICE IN REPORTING RACIALIZED FINDINGS

One thing nearly all the interviewees agreed upon is that using these variables in genetics research carries inherent dangers of misinterpretation and overgeneralization, which can have negative consequences for the groups in question (Hunt and Megyesi 2008a). This point has been taken up in the literature, and genetics scientists are being called upon to be careful in their choice of terminology when reporting findings (Kittles and Weiss 2003; Rosenberg et al. 2003). Concrete policies for addressing these complex issues are only just beginning to be developed (Kahn 2006;

Lee et al. 2008; Lillquist and Sullivan 2006). As yet, there appears to be little effect on the practice of reporting data divided by common racial/ethnic labels. In our review of recent medical and genetics literature, we found use of common racial terms such as "Caucasian," "Asian," or "African American" to be extremely prevalent.

The NIH's Polymorphism Discovery Resource (PDR), which included an overt effort to avoid use of racial/ethnic identifiers, was soon followed by another major NIH-sponsored genetics project, the International HapMap Project. The HapMap is an international collaborative effort to gather genetics data for comparative analysis of genetic variation between people in geographically distinct areas. The data is freely available online, and anyone with a computer connection can download data sets. The goal in making the data publicly available is to encourage biomedical researchers to identify the genetic basis of disease and drug response (Altshuler et al. 2005). Perhaps in response to researcher frustration with the lack of racial/ethnic identifiers in the PDR data, the HapMap does include group identifiers, accompanied by a careful caveat for avoiding racial labels in reporting data. This provides an interesting example of an effort to establish policies that discourage drawing and publishing findings implying biological races.

The groups from whom genetics data were gathered for the HapMap are described in the official communications of the project as "populations with African, Asian, and European ancestry" (International HapMap Consortium 2003). Protocols for the HapMap have been quite carefully designed to avoid ethical criticism for promoting notions of biological race and have included explicit efforts to avoid racialized terms in reporting findings. Researchers using the HapMap data are instructed to report their findings using specific terminology for labeling the samples in research publications: for example, "Yoruba in Ibadan, Nigeria" and "Han Chinese in Beijing, China" (International HapMap Project 2005). However, in reviewing articles reporting analyses of HapMap data, we found that although this recommended terminology is usually in the methodology section, the rest of the article reverts to common continental race labels such as "HapMap Asians" or "the European population."

In fact, one might argue that these labels are actually encouraged by the HapMap Project, because of the way that the data sets are offered. As with the Coriell data set, to download HapMap data, researchers must select populations from a drop-down menu with these options: "Utah residents with ancestry from northern and western Europe; Han Chinese in Beijing, China; Japanese in Tokyo, Japan; Asian Combined (Japanese + Chinese); and Yoruba." Kahn notes: "The resulting blocks of variation are being

identified with their source population. The population groups are already being characterized as representative of the broad continental population groups of Africa, Asia and Europe" (2006:1967). Thus, despite the formal policy to avoid racial labels in referring to this data set, continental racial concepts permeate the very design of the HapMap Project itself.

It is sobering to consider the actual samples upon which these continental labels are so matter-of-factly placed. The sample labeled "European" was actually collected in Utah in 1980 by a French research group (CEPH, or Center for the Study of Human Polymorphisms). It consists of ninety individuals from thirty familial sets of a parent and two offspring. No further selection criteria are known, so it is unknown how closely related these thirty family sets are to one another. The "African" sample also consists of ninety individuals in thirty familial sets, all of whom live in the large city of Ibadan, Nigeria, and report four Yoruban grandparents. The Yoruba is a very large West African ethnic group consisting of more than thirty million people. The "Asian" sample is ninety individuals, combining a Chinese sample and a Japanese sample of forty-five individuals each. The Chinese are also urbanites and claim at least three grandparents from a very large ethnic group, the Han, a group that includes about 92 percent of all Chinese. The Japanese are residents of Tokyo, with no further selection criteria noted (International HapMap Project 2005).

Thus, this influential international effort to document the genetic diversity of "continental populations" uses samples from people who do not necessarily inhabit the continent they are meant to represent, some of whom are known to be closely related kin, others whose kinship relationship is completely unknown, and still others whose "ancestry" is subsumed under very broad ethnic labels such as Yoruba and Han, whose kinship implications are unknown and unexamined. As a result, it clearly is not possible to distinguish between familial patterns, ethnic group linkages, and the continental origins of the observed trends to draw any conclusions about the nature of continental populations.

The fact that these racial constructions are embedded in the fabric of this data set, despite the very explicit policy and instructions of the HapMap Project, attests to the deep-seated nature of notions of continental ancestral groups and the habit of typological racial thinking. Some have called for agencies such as the NIH to develop more aggressive and explicit policies to oversee the scientific uses of racially labeled data (Stevens 2003). However, given the ease with which these notions have found their way into the HapMap Project, despite the very conscious effort to avoid them, it seems unlikely that such mandates could be very effective.

## RESEARCH POPULATIONS AND CONTINENTAL RACIAL GROUPS

The problem of ambiguity concerning the intended populations of study is not limited to the HapMap Project, by any means. Gannet (2003) has pointed out the arbitrary nature of "populations" in the context of scientific investigations. The theoretical questions driving scientific investigation are what, in fact, determine how a population is constituted, rather than any inherent characteristics associated with the group labels. Gannet further argues that it is inaccurate to assume, as continental racial labels imply, that there is a definitive collection of biologically distinct groups independent of researchers' practical needs and theoretical interests. Rather, different research questions result in different populations; naming populations creates them as discrete entities. Populations are constantly formed in various research studies concerning species genome diversity: "Genes become bounded in space and time in ways that fulfill aims, interests, and values associated with particular explanatory contexts. Population boundaries are not fixed but vary from one context of inquiry to another" (Gannet 2003:990).

The genetics studies in our literature review and interviews include a broad cross-section of types of research. As such, their study populations are radically different, and their geographic and time frames are quite distinct and, most often, only loosely defined. However, in place of careful definitions of the specific populations being studied, they routinely employ the familiar vocabulary of continental racial groups. Let us consider various research projects we have encountered and the diverse ways they construct, reproduce, and augment the concept of continental races.

One type of project is population genetics studies. These are concerned with modeling human evolution and migration. They begin not with individuals but with observed genetic frequencies. Using powerful computer programs, they examine the "fit" of those frequencies to proposed patterns of continental migrations. The "continental populations" concept is used as a propositional framework, against which the observed genetic patterns are examined and movements over time and space are hypothesized, such that the present distribution of observed genetic material can be extrapolated back to a model of geographic origins. The concept of continental races is not tested by this approach but instead may provide an analytical framework for model building.

This analysis is highly propositional, intended to model the movement of large populations across vast spans of time. Some models have captured the attention of the popular media. They have been widely interpreted as

documenting the common maternal (mitochondrial) and paternal (y-chromosome) ancestry of the human species and its emanation into distinct subgroups, which populate the continents (see, for example, Gugliotta 2008; Jones 2007; Shreeve 2006). In the public imagination, this has become a story of the beginning of continental races: groups with shared origins in Africa, a long history of isolation, and recent hybridization from these once pure stocks of human races. We will return to these concepts in some detail below.

Other population genetics studies are concerned with describing the distribution of common genetic variants in current populations. These studies also begin with genetic material, rather than individuals, and strive to identify genes that cluster in pre-identified populations. Many analyze samples from several sources, which are stratified according to the continental origins labels already associated with the samples based on their sources, as we saw in the HapMap discussion above. The notion of "continental races" precedes analysis and is used to label whatever groups the genetic clusters might be found in. This approach does not test the idea of continental populations but rather assumes it as a basis for structuring analysis.

Critics contend that the use of continental racial labels for these data sets, in the absence of any discussion of what is meant by "population" in these studies, has resulted in the samples' readily being interpreted as disguised surrogates for race (Braun and Hammonds 2008). Others argue that the haplotype trees they produce, the statistical representations frequently used to model associations found in these studies, assume primordial isolation between continental groups based on selected genetic traits (Hawks and Wolpoff 2003; Templeton 2002), without documenting concordant distribution of multiple, independent, genetically based traits, which such a conclusion would require (Keita et al. 2004).

Bolnick (2008) has offered a particularly insightful critique of a popular tool used in such studies: the population clustering program Structure. Several studies using this program have gained much attention and are often cited as evidence that continental races are biologically identifiable (Bamshad et al. 2003; Rosenberg et al. 2002). Bolnick argues that the apparent success of a program at documenting genetic differences between continental groups is due to the combined effect of dubious theoretical assumptions underlying how the program is applied to given data sets and selective sampling of isolated populations. These practices predicate analysis upon assumptions about the nature and distribution of racial groups rather than lead to discovery from the data. Thus, finding race to be a salient variable is built into the study itself (Fullwiley 2008b).

Another type of project using continental racial labels is clinical genetics studies. These are concerned with understanding the genetic basis of disease susceptibility and treatment response. Most of the studies in our literature review and interviews are of this type. Unlike population genetics studies, these begin not with genetic material but with individuals or families affected by the disease of interest. The studies may take several forms, such as epidemiological studies, sibling-pair studies, studies of large affected families, or case-control studies. Genetic analysis may involve seeking the presence or absence of a candidate gene, a gene suspected of being at play for the disease in question, or it may involve more exploratory research designs, seeking to identify genetic characteristics that cluster in affected individuals or their families. For all these studies, researchers' pre-existing logic of racial difference is imbedded into both the design and the interpretation (Fullwiley 2008b). The ways the notion of race is manifest varies, depending on the type of study.

One way that clinical genetics studies employ the race notion is by stratifying samples of comparison. A common design in case-control cohort studies is to compare affected and unaffected individuals and to report the findings divided along racial lines. Here, the lines of ancestral descent from the continental racial populations are tautologically presumed, but actual kinship data among the individual group members is not collected or examined.

Other clinical genetics studies select families or individuals belonging to a population thought to be particularly affected by the disease of interest—for example, Hispanics for diabetes or African Americans for cardiovascular disease—with the intention of increasing case findings and thereby stacking the deck toward identifying genetic factors affecting susceptibility.

Many clinical genetics studies focus on "affected families." They begin with a "proband" (that is, an individual known to have the disease of interest) and then include a number of that person's close relatives in the sample. Here we have a case in which it would seem that the unit of analysis is families. However, when the subject families are racial/ethnic minorities, the results are not discussed in terms of an affected family but instead are labeled in terms of a racial/ethnic population, for example, reporting a "major genetic determinant in Mexican Americans" or describing "a high frequency of this variant in African Americans." Interestingly, in our observation, this is not the case when the families are of the majority racial/ethnic population, such as white European families studied by researchers in the United States and Europe or Asians studied by researchers in Asia. Then the findings refer to the genetic characteristics of a family rather than a racial/ethnic population.

### SOUNDS THE SAME, BUT DIFFERENT MEANINGS

We have seen that all kinds of studies, whether population genetics or clinical genetics, presume rather than test continental ancestral lines of descent and readily identify samples, drawn in radically different ways for radically different purposes, by using the same common racial labels. These research designs divide samples into very different kinds of groups but labeled with very similar terminology. The indiscriminate use of vague and unsystematic terminology results in a semantic illusion that very different types of research are examining similar populations.

The target populations are highly varied, depending on the goals of the project: some are chosen because of their geographic isolation, others for their disease characteristics, and others for their mere availability. However, when all are labeled with the same simplistic set of terms, it seems as if a growing body of data is documenting genetic distinctions between racial populations. But, in fact, there is no reason at all to presume that samples belong to a "population" of any kind, beyond their having the same label affixed to them. In other words, the only equivalence between the "African," "European," and "Asian" samples is that they are subject to equivalent terminology.

We have seen that continental racial labels are poorly conceptualized categories and are unsystematically applied across a whole gamut of unrelated study populations. Why is it that, in these otherwise highly systematic and rigorous scientific disciplines, this particular vagueness is tolerated and replicated? Why is it that such diverse research designs so readily turn to the same frame of reference, using these common racial groupings as an unexamined way to cluster and analyze data? Evidenced by the physician-colleague described in the introduction, whose new-found conviction that race is not biological was quickly vanquished by the next study he encountered affirming group variance, there appears to be a deeply rooted logical appeal to the concept of continental races. These researchers may share a commonsense framework of assumptions about human diversity that axiomatically incorporates racial categories into everyday research practices. We propose that these assumptions are rooted in popular Western cultural concepts about human origins, geography, and history.

### RACE AND THE ORIGIN MYTH

Myths are ahistorical, traditional stories that are widely believed within a society and condense popular ideas about the natural world and history for that culture. Myths are more than just symbolic or historical stories, as Malinowski explains:

Myth...is not of the nature of fictions, such as we read today in a novel, but it is a living reality, believed to have once happened in primeval times, and continuing ever since to influence the world and human destinies. This myth...to a fully believing Christian, is the Biblical story of Creation, of the Fall, of the Redemption by Christ's Sacrifice on the Cross. As our sacred story lives in our ritual, in our morality, as it governs our faith and controls our conduct.... [Malinowski 1948:100]

Myths are commonly constructed in ways that validate power relationships, making social hierarchies appear natural and preexistent. Through myths, societies order their world. As such, myths hold an important place within larger societies, providing a mechanism through which religion and religious ideology may influence social institutions such as science and politics (Barfield 1998; Bowle 2006).

Indeed, it is increasingly recognized that the production of science does not occur in a purely objective world but instead is subject to the framing influences of social, cultural, historical, and political contexts (Berger and Luckmann 1990; Duster 2006). For Western society, the Judeo-Christian tradition provides a dominant lens through which much of the world is understood, and, as such, it would follow that this tradition may provide the dominant context in which biomedical and genetics sciences are being produced.

The Judeo-Christian myth of human origins is the story of Adam and Eve. In it, understandings of how humanity began and why it has turned out as it has in the present day are laid out as a tale of the rise and fall of Adam and Eve in the Garden of Eden (Winzeler 2008): God created the universe and the world, made Adam from the earth, and, taking one of Adam's ribs, created Eve. They lived in the perfection of the Garden of Eden, where all their needs were filled as long as they obeyed God's law. After Eve ate from the forbidden tree, God cast the couple out of the Garden. They wandered in search of a new land, and their offspring went forth and populated the earth.

Almond (1999) argues that, since the seventeenth century, the story of Adam and Eve has provided Western culture with the key for interpreting the present in terms of an ideal past, at once linking all humanity and explaining its inherent separation from the Creator through the cataclysmic event of The Fall. Although the Enlightenment ushered in scientific thought and rationality, religious influences remained central, particularly in the construction of modern human origins and debates over the existence and nature of human races.

The great debates about race in the early development of evolutionary science were tangled in efforts to reconcile scientific conclusions with biblical accounts. In this framework, competing notions of whether the human races have descended from more than one ancestral type—monogenesis versus polygenism—dominated discussions of the time. Brace (2005) contends that the concept of "race" came out of polygenism doctrine, asserting that humans of different races belonged to separate species.

In our interviews with genetics scientists, we were surprised how often their discussions of the concept of racial groups turned to popular and religious images of the origins of humankind, sometimes making specific references to Adam and Eve and their descendants. Quite commonly, as they discussed why they believed that race is important to genetics research, they would lapse into a discussion of the origins of humans, referring to a primordial set of pure types from which our current populations have descended. The story of racial origins that recurred throughout these interviews can be summarized as follows: humans originated in Africa, from where they traveled to the other continents. They established continental racial groups and then lived in relative reproductive isolation until quite recently, when they began to intermix. People today are descended from these essential populations, and the lines of ancestry are evidenced in present-day appearance, geographic location, and/or genetic profiles.

When discussing these concepts, the genetics scientists we interviewed did so in a matter-of-fact way, at times citing specific popular media, such as *National Geographic* or television documentaries about human origins. A core image in these accounts is that the people of the world are direct descendants of primordial maternal and paternal lines that originated in Africa. Humans then became firmly rooted as primary races on each of the continents, until very recent innovations increasingly resulted in admixture of these lines. One genetics epidemiologist we interviewed put it this way:

The signature of ancestry would be place. So we focus on either what [subjects] tell us about their ancestry, or their skin color. And that's the surrogate that we use to go back and be able to put people in a, I'll call it a, historical context that may be meaningful for understanding their genetic history. And it all goes back to the mutational history of life. That's why people spend so much time, even right now, trying to figure out, do we all go back to one Eve, or did human beings evolve separately in different locations?<sup>1</sup>

A molecular biologist expressed a similar perspective: "All humans eventually end up being of African origin.... All of our ancestors actually do derive from Africa.... In the mitochondria, we can see that the maternal lineage of all humans can be restricted to eight or nine distinct classes, which, you know, are eight or nine prototypical females. But it's not as though we could say, "This one was Eve living there somewhere."

The stamp of the Judeo-Christian origin myth is perhaps most blatantly obvious analyses of genetics studies tracing human origins and dispersal patterns through genetic markers associated with mitochondrial DNA and with the nonrecombining portion of the Y chromosome (Hammer et al. 1998). By studying the present-day distribution of these markers, researchers are attempting to trace humanity's lineages to the earliest common ancestors—essentially, the first man and woman, who are popularly dubbed "mitochondrial Eve and Y-chromosome Adam" (Oppenheimer 2003; Wells 2002).

Some have argued that, in reconstructing human origins using Judeo-Christian terminology, the religious doctrine associated with that terminology is likewise associated with scientific analysis and conclusions. According to Kidd (2006), Christian religious scripture has provided a primary cultural influence on the forging of the idea of race and in ideological assaults upon racism. He argues that biblical interpretations have been at the forefront of racial debates since their inception in the seventeenth century, fueling both eccentric religious doctrines of racial hatred and arguments for close interracial kinship. It is a complicated relationship, but the Christian doctrine behind scientific theories of the origins and great divisions of humankind is clearly visible. Adam and Eve leave the Garden of Eden, and their descendants spread outward to populate the earth. This is updated with current terminology from genetics science: our progenitors emerge in Africa and move outward to populate the continents, establishing a set of pure types that are only recently admixing. That this story provides a conceptual lens through which genetics sciences are being framed is evident in our interviews with genetics scientists, by the frequent references to Adam and Eve and to the idea of primordial types inhabiting the continents.

#### EUROCENTRISM AND THE CONCEPTS OF CONTINENTS

The idea of primordial pure types is logically appealing partly because of a fundamental acceptance of the idea that ancient continental populations lived in relative isolation, a certainty grounded in the seemingly concrete physical reality of continents. Continents are a core concept in

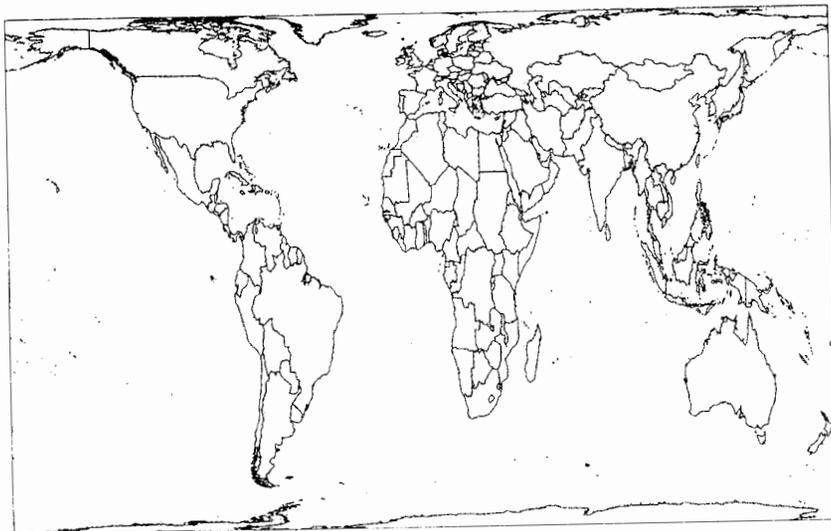
our society, permeating scientific thinking in any number of unexamined ways. The world is understood to comprise a handful of essential landmasses: Asia, Africa, Europe, Australia/Oceania, and the Americas. This is the basic framework that underlies the familiar racial taxonomy so common in biomedical and genetics research. However, this taken-for-granted understanding of inherent geographic divisions of the world, some argue, is arbitrary and misleading, firmly situated in Eurocentric spatial assumptions (Blaut 1993).

Lewis and Wigen (1997) make a strong argument that common Western understandings and visual representations of world geography are highly Eurocentric. They point out that the standard continental formulation gives Europe, which is merely a peninsula, inflated importance as a continent comparable to vastly larger landmasses such as Africa or the Americas. Europe's component parts, small nation-states such as France and Germany, are elevated to be on a par with major expanses such as China and India, which are downgraded as subcontinents. Standard world maps further reflect this Eurocentrism by inflating the size of Europe and the United States, with the rest of the world correspondingly reduced. There has been some effort in recent years to correct this misrepresentation, through redrawing maps using alternative projection techniques (Monmonier 2004; Peters 1990). How exactly to redraw the world map in a more "area-accurate" representation remains controversial, but, in most such revisions, Europe and the United States are reduced and Africa and Latin America are enlarged to more accurately capture their relative sizes (figure 5.1).

Still, the Eurocentric model of continental divisions is rarely challenged and continues to dominate Western understandings of world geography. Thus, when genetics databases label their populations along continental lines, the comparability and representativeness of their samples are misleading. Samples from the relatively tiny European peninsula are compared to samples aggregated from across great expanses such as Asia or Africa. The tendency to lump together samples from vast non-European regions contrasts markedly with the care and attention given to classifying European and US samples, distinguishing so specifically between national, regional, or familial affiliation. The cultural lens of a Eurocentric model of continental divisions is apparent in how genetics researchers produce and interpret their samples.

#### THE IDEA OF AFRICA

Another often overlooked orienting concept in "Out of Africa" origin



**FIGURE 5.1**

Example of a revised, area-accurate world map: The Peters Projection ([http://www.geography.org.uk/download/GA\\_REMapPeters.jpg](http://www.geography.org.uk/download/GA_REMapPeters.jpg)). Used with permission.

stories is the arbitrariness of the idea of “Africa” itself. Tallbear (2007) has argued that the picture of “Africa” in the popular imagination about human origins is of a place of primordialism, an ancient landmass where the human species evolved rather than a place that is populated by our contemporaries. Braun and Hammonds (2008) have examined the relationship between the birth of the idea of “Africa” as a historical construction of “nation-continent.” Drawing on Gannet’s (2001) critique of population thinking, they contend that after anthropologists named “populations” in international atlases and databases, these became “real” in the gaze of the West. This literally produced African groups as distinct and fixed entities, thus transforming heterogeneous groups into static, naturally occurring “tribes.” Imagined as such, these groups readily are made into the objects of large-scale population genetics studies, and the reality of their precolonial history as intermingled peoples with constant gene flow is erased.

When scientists speak of “African populations” or other continental populations, they formulate populations as relatively isolated, pure types with some limited exchange of genes through time. Consider the remarks of one genetics epidemiologist we interviewed: “There are researchers all over the country [who] have gone to Europe and to Africa and South

America and actually found groups of people that have never moved and, you know, constantly are breeding with the same people.”

However, the idea that these groups are or were genetically isolated and only recently have begun to mix is an erroneous formulation. The intermingling between ancient populations has constantly occurred across history; there have never been isolated, bounded, continental populations (Brown and Armelagos 2001; Cooper, Kaufman, and Ward 2003; Goodman 2000). Despite their commonsense appeal and their centrality to the familiar notion of racial groups and admixture, there is no evidence that these primordial pure types have ever existed.

In current science, the presumption of homogenous primordial groups seems clearly evidenced by common usages of the idea of Africa. We have noted an interesting tendency in the various studies we have reviewed. The specific terminology for European-descended individuals includes a long list: Caucasian, white, European, non-Hispanic white, combined with specific nation-state names within Europe. In contrast, the list of terms for African-descended individuals is impressively brief: African, African American, or, more commonly, simply “black.” On occasion, specific African groups were named as sample sources, but, without exception, these specific labels were supplanted in later remarks by the simple, broader terms “African” or “black.” Thus, despite Africa being well known to have the most genetic variation of any region, in genetics research, people of African descent are commonly treated as though they are genetically homogeneous (Ossorio and Duster 2005). This would seem further evidence of the deep roots of the continental racial concepts in the Eurocentric legacy of Western culture.

## CONCLUSION

In this chapter, we present observations about the surprising tenacity of the idea of race in biomedical and genetics sciences. We show that antique human taxonomies and theological notions of human history underlie the familiar racial groupings so common in scientific discussions of human variation. The idea of biological races is based on inaccurate assumptions and is inconsistent with observed genetic trait distributions and known human behaviors. This model assumes that racial/ethnic groups are essentially endogamous, that humans have rarely moved between continents in the past, and that gene flow between groups is rare and recent. To the contrary, exchange of mates across broad geographic areas is the norm for human populations, resulting in clinal variation rather than clearly distinct genetic stocks. Furthermore, significant intermarriage between

socially designated groups has routinely occurred throughout history, and admixture between groups has never been an exceptional event (Harry and Marks 1999; Race, Ethnicity, and Genetics Working Group 2005; Weiss 1998).

These misconceptions notwithstanding, the view that we are descended from distinct primordial human types is manifest in a fundamental way in the design and interpretation of many types of genetics studies. Race is routinely treated as a *de facto* variable in a broad cross-section of theoretically and methodologically unrelated research. Lee (2006) has described the entrenched use of race mythology in the labeling, storage, and distribution of human genetic data as a "racializing technology," lending a scientific veneer to the idea of biological race.

Interacting factors contribute to and facilitate the persistence of traditional notions of race in the biomedical and genetics sciences. The typological thinking underlying the race concept can find its way into all phases of research: study design, conceptualizing and operationalizing variables, data organization and storage, and interpretation of findings. The vague categories and inconsistent labels used in defining racial variables allow a level of imprecision that promote this outcome. The phenomenon is further facilitated by relying on these familiar constructs and labels in a broad cross-section of types of studies and time-frames and for diverse study samples. The language of continental racial groups freely supplants specific identification of the populations under study, giving the appearance of a limited set of racial groupings, as though they are somehow comparable across these studies. This further contributes to the illusion that diverse studies about highly diverse groups of people are examining characteristics of a limited set of racially named groups. Ancestral lines of descent are presumed rather than tested, and samples drawn in radically different ways for radically different purposes are readily labeled using the same common racial terminology.

We have argued that the ease with which the idea of primordial racial identities emerges in the design and interpretation of genetics science may be, at least in part, due to the deep roots of racial thinking in unexamined orientating Western cultural sensibilities. The use of Judeo-Christian mythological images in tracing human genetic origins and dispersion is one especially prominent example of how overtly cultural concepts inform and frame scientific thinking. Other cultural concepts clearly visible in this arena are Eurocentric ideas of what composes the continental landmasses themselves, with which these essential groups are associated, and the lopsided amounts of detail with which groups are described.

Anthropologists have been highly involved in raising concerns about the often off-handed way that race and ethnicity are being used in current biomedical and genetics research. We carefully review the anthropological position on race, which has become axiomatic in our field: race is a social construct, applied to highly fluid groups that do not constitute biological subgroups. We recite the evidence: race does not correspond with genetic variation; variations are nonconcordant; and variation is continuous, without regard for political boundaries, language, or religion.

However cogent and persistent these arguments may be, the flood of genetics research presuming the racial/ethnic basis of human variation continues unabated. The ineffectiveness of existing critiques may reflect an important difference between anthropological and clinical/scientific ways of thinking. In contrast to anthropologists, who embrace a tradition of relativist thought and contextualized analysis, clinical and scientific researchers are guided by a paradigm that is essentially positivistic, dealing with stable, empirical entities that exist on an objective, tangible plane and respond to the laws of nature. Categorical thinking is a hallmark of this orientation. The central problem is to correctly identify the category to which phenomena belong and to examine them following the appropriate algorithm.

Human variation *does* exist. Biomedical and genetics science will continue to observe and report patterns of variation between groups of people. The anthropological mantra of "Race Is a Social Construction" fails to interact persuasively with this fact. In the absence of a cogent solution to the problem of how best to describe and categorize these observed variations, anthropologists may find themselves flirting with irrelevance. Categorical thinkers, like the skeptical clinician described at the beginning of this chapter, rush to categorize observed variation. They reach for the handy, for the familiar, for the culturally customary, in framing their ideas about the order of the natural world.

A broadly aimed and abstract critique of the type anthropologists are apt to produce will have little effect if it fails to suggest more appropriate ways to think about the reality of variation. The arguments of anthropologists and other socially minded critics are easily mistaken for "politically correct" social justice concerns and dismissed as trumped by the "biological facts" uncovered through scientific method (Krieger 2005). A more effective approach would be to systematically engage with the conceptual and procedural assumptions inherent in specific applications of racial/ethnic categories and suggest more appropriate ways to describe and classify observed variation.

In place of promoting policies that may have little effect, a more definitive

and collaborative approach might be more successful; working together with biomedical and genetics scientists toward a more careful delineation of populations and a more accurate level of explication in discussing findings (Haslanger 2008). Achieving such a goal will require developing effective strategies in the biomedical and genetics sciences to promote a deeper awareness of the logical, methodological, and conceptual flaws that plague these practices and of the folk notions of human history that thereby permeate these otherwise rigorous fields. With a fuller awareness of the scientific inadequacy and social consequences of allowing common racial concepts to continue to be a ubiquitous filter through which scientists gaze, we will move toward developing more objective and scientifically useful notions of population variation in biomedical and genetics research.

#### **Acknowledgments**

This research was supported by the National Institute of Health National Center for Human Genome Research through grant #HG2299-05. We wish to thank the researchers we interviewed, whose kind cooperation made this research possible. Mary Megyesi made many important contributions to the development of the arguments in this chapter, helping conceptualize many of the core ideas presented here and assisting with much of the literature review and data analysis for the chapter. James Bielo and Daniel Vacanti also provided valuable assistance with a variety of data processing and analysis tasks.

#### **Note**

1. Identifying details are excluded to protect the anonymity of the interview subjects quoted in this chapter.