ALSO BY DOROTHY ROBERTS

Killing the Black Body: Race, Reproduction, and the Meaning of Liberty

Shattered Bonds: The Color of Child Welfare

FATAL INVENTION

How Science, Politics, and Big Business Re-create Race in the Twenty-first Century

Dorothy Roberts

THE NEW PRESS

NEW YORK
LONDON

3

Redefining Race in Genetic Terms

At a contentious 2008 meeting at the National Human Genome Research Institute (NHGRI) in Rockville, Maryland, forty scientists and bioethicists debated the best way to talk about the flood of genetic variation research emanating from laboratories across the country and the world. The discussion soon focused on how to handle the pesky term *race*. According to *Science* reporter Constance Holden, "everyone at the meeting agreed on the need for non-'fraught' terminology—'geographic ancestry,' for example, instead of 'race.'"[1] Rejecting race as a valid genetic category creates a vacuum in the scientific vocabulary: how will scientists describe genetic differences between human populations without resorting to the racial categories they have used for centuries?[2]

This reconfiguration of race for the genomic age hinges on applying two key concepts to genetic information: statistical probability and geographic ancestry. With the advent of worldwide genomic population studies, many scientists are using statistical estimates of gene frequencies that differ among geographic populations as a more objective, scientific, and politically palatable alternative to race. Instead of grouping people by race for purposes of scientific studies, why not group them by statistical genomic similarities? A second, related strategy turns to geographic ancestry. But these approaches, as we shall see, tend to merely repackage race as a genetic category rather than replace it.

Statistical Race

Some genomics researchers see race as a statistical grouping based on genetic similarity. Because genes are inherited, biologically related individuals are more likely than unrelated individuals to share genetic variants. A growing branch of population genomics treats race basically as a large family with the same "very distant relatives."[3] According to this theory, individuals belonging to the same race share more of their recent ancestry and therefore are more genetically similar to each other than to those of other races. It is important to note, however, that population biologists using a statistical approach do not actually trace the ancestry of particular individuals to place them in racial categories. Rather, they infer groupings from the statistical frequencies of particular DNA sequences sampled from distinct populations around the globe. Unlike racial typologists who classified people into natural kinds based on outward appearance, modern-day racial scientists classify people according to statistical probabilities based on huge genetic data sets.[4]

But does the statistical race concept really prove the biological nature of race any better than existing racial typologies? How can it when the entire enterprise from beginning to end—identifying populations to enter into data sets, determining which and how many genetic clusters matter, and applying the findings to our everyday lives—inescapably depends on preconceived notions of race? Genomic scientists have not discovered race in our genomes. They are taking already accepted racial categories and telling us a new way, based on computer-generated genetic differences, to verify them scientifically.

To understand how modern genomics reproduces traditional ideas of race, it is helpful to take a closer look at the assumptions behind some particularly influential scientific studies. First, another science lesson. Population genomics leverages the tiny percentage of genetic variation in the human species to identify differences in the frequency of certain alleles, or versions of genes, among groups with different geographic origins. To understand this process, we have to start with SNPs and microsatellites. Single nucleotide polymorphisms, or SNPs (pronounced "snips"), are particular points where the genomes of different individuals vary by a single DNA base pair.[5] SNPs are caused by random mutations that are then passed on to offspring and disseminate slowly away from the group in which they originally occurred.[6]

In some cases, the SNPs produce phenotypical differences; in other cases, the differences are imperceptible. Either way, scientists can infer ancestral relationships from the frequency with which these alternate spellings occur in different groups. A similar principle is at work with short segments of DNA called microsatellites that have a repeating sequence of nucleotides that varies between individuals.

Because the scale of this research is so vast (the human genome includes over 3 billion base pairs), scientists use computer software to infer a population structure from genotype data using multiple loci sampled from a number of groups. With a popular software program known as Structure, a researcher indicates how many genomic clusters the data should be grouped into. The program then allocates the individuals whose DNA was sampled into the predetermined number of clusters based on their genetic similarity. For any given number fed into the program, explains anthropologist Deborah A. Bolnick, "Structure searches for the most probable way to divide the sampled individuals into that pre-defined number of clusters based on their genotype." It uses a mathematically sophisticated algorithm to maximize the chances that the genotypes of individuals in the cluster will match.[7]

Researchers began attempting to group human populations based on their genes soon after the human genome was sequenced. A particularly high-profile project was led by Noah Rosenberg, a former high school math whiz who became a computational biologist at the University of Southern California. His team included genetic scientists from Stanford, Yale, and the University of Chicago, as well as institutes in Paris and Moscow. Unlike in previous studies, Rosenberg's team used computer software to detect clusters of genetic similarity in unidentified DNA sampled from people across the globe. Could the researchers scramble the groups' genetic signatures and put them back together with the aid of computer technology?

Rosenberg's team fed into the computer genetic information from 1,056 individuals representing fifty-two global populations. Then, following a specially designed algorithm, the computer went to work. In a landmark article, published in *Science* in 2002, the researchers announced that they had "identified six main genetic clusters, five of which correspond to major geographic regions." Although race was not mentioned, the "major geographic regions" that matched the genetic structure they discovered—Africa, Eurasia, East Asia, Oceania, and America—were quickly translated into traditional racial divisions. The researchers also concluded that "self-reported

population ancestry likely provides a suitable proxy for genetic ancestry" when evaluating individuals for disease risk.[8]

Had high-tech genomic research really confirmed eighteenth-century racial typologies? Closer inspection of the Rosenberg team's findings reveals that they do not verify five classic racial groups at all. Instead, the study's overall results confirmed the basic rule of human genetic unity: within-group genetic variation is much greater than between-group variation. Genetic differences among people within the populations they studied accounted for 93 to 95 percent of all the genetic variation the computer uncovered. Only about 5 percent of the variation found existed between groups. In fact, the distinctions between populations were so minuscule that it took a highly advanced statistical computing program surveying many accumulated differences to make reliable guesses about the geographic origin of the people sampled.

What's more, the numbers of genetic clusters they identified were arbitrary. Although the researchers emphasized six main genetic clusters in reporting their results, they actually told the computer to analyze the DNA data set using a range of numbers, not just six. Their theory was that any statistically significant clusters reflected genetic divisions of the human species and thus the natural structure of human populations. But remember, the number of genetic clusters is dictated by the computer user, not the computer program. Their article presented the results of using two to six predetermined clusters. Rosenberg later revealed that his team also analyzed the data set using six to twenty clusters, "but did not publish those results because Structure identified multiple ways to divide the sampled individuals" when the number was larger than six.[9] The larger number of clusters identified by the study could just as easily have been highlighted to demonstrate the difficulty of dividing human beings into genetic races. There is nothing in the team's findings to suggest that six clusters represent human population structure better than ten, or fifteen, or twenty.[10]

Instructed to find two clusters, the computer divided human beings into groups anchored by Africa and by the Americas. This reflects the portrait of migration that evolutionary biologists have already painted: Native Americans traveled the most genetic distance from our original ancestors in Africa. When researchers told the computer to form five clusters, they were able to divide the human species into groups that matched the indigenous peoples sampled from five continents (Africa, Eurasia, East Asia, Oceania, and the

Americas). Adding another cluster separated out an additional group made up entirely of the Kalesh, a group in the mountains of Northern Pakistan who speak an Indo-European language and whose inhabitants claim to be descendants of Greek soldiers who invaded the Indian subcontinent with Alexander the Great in 327–323 B.C.[11]

Rosenberg's study was touted in media accounts, including a front-page story by the respected *New York Times* science writer Nicholas Wade, as having proved the biological reality of race. The researchers identified "five main human populations," which, in turn, "broadly correspond with popular notions of race," Wade wrote.[12] But the study actually showed that there are many ways to slice the expansive range of human genetic variation. In a 2005 article, Rosenberg and his colleagues acknowledged that the way a genomic study is designed determines what it says about human population structure.[13] Based on their analysis of how changing key variables influenced the outcomes, they reported that the number of loci, the sample size, the number of clusters, the geographic dispersion of the samples, and assumptions about allele-frequency correlations all had an effect on clustering. Although they reiterated their earlier finding that, with a large enough worldwide dataset, "individuals can be partitioned into genetic clusters that match major geographic subdivisions of the globe," they stated this finding "should not be taken as evidence of our support of any particular concept of 'biological race'" and agreed that genetic diversity also consists of clines—differences in allele frequencies that occur gradually across regions. The original Rosenberg study itself had contained the caveat that "genetic differences among human populations derive mainly from gradations in allele frequencies rather than from distinctive 'diagnostic' genotypes."[14]

Understanding human population structure in terms of discrete genetic clusters also misrepresents the path that produced diverse human populations that diverged from shared ancestors in Africa. "Ironically, by ignoring the way population history actually works as one process from a common origin rather than as a string of creation events, structure analysis that seems to present variation in Darwinian evolutionary terms is fundamentally non-Darwinian," Penn State anthropologists Kenneth Weiss and Brian Lambert point out.[15]

One population geneticist called genomics "the computer-assisted comprehensive study of all genes."[16] New genomic technologies dumped in researchers' laps a gigantic array of unorganized genetic data to sort through.

Like the natural world that Enlightenment biologists put in order, the information derived from the human genome beckons molecular biologists to catalog it. And like their predecessors, modern-day scientists are utilizing race to make sense of new genomic discoveries. "Race has rapidly become a prominent 'search tool,'" note the editors of the anthology *Revisiting Race in a Genomic Age*.[17] Today's scientists, however, claim that their focus on "genetic clusters" has removed the political aspects of race from their research. In response to my question asking why Rosenberg's article focused on the clusters that most closely matched the five major continents and therefore our historical ideas about race, a member of his research team told me that "people who share the same continental origin are genetically similar." He went on to explain, "Race has got all these loaded connotations, so we've dropped that term from our work."[18] What this scientist failed to acknowledge was that his own acceptance of racial categories may have influenced the decision to emphasize five genetic clusters, despite his team's attempt to expunge any explicit reference to race.

Is Geographic Ancestry the New Race?

The idea of replacing a typological notion of race with a geographic category traces back to the 1937 text *Genetics and the Origin of Species* by Theodosius Dobzhansky, the founder of population genetics and one of the signatories of the 1950 UNESCO statement on race. Dobzhansky proposed that evolutionary biologists adopt as their unit of analysis "geographical race," which he defined as "populations of species that differ in the frequencies of one or more genetic variants, gene alleles or chromosomal structures."[19] Dobzhansky's population approach was a redefinition of biological race, not a rejection of it. While contrasting his concept of natural populations with Carlton Coon's racist ideas in *The Origin of Races*, Dobzhansky still maintained that "most biological species are composed of races, and *Homo sapiens* is no exception."[20] But how much genetic difference is enough to create a race? Any genetic measure requires applying some a priori concept of race in order to package human genetic variation in a limited number of biological groupings. Informed by genetics, zoologists discarded race as a useful way to divide up animals within a species because so many races were distinguished by only one or two genes: genetic testing revealed that "two animals born in the same litter could belong to different 'races,'" notes Richard Lewontin.[21]

Some genomic scientists I interviewed advocated an absolute rejection of race as applied to the human species. Charmaine Royal, a Duke University geneticist who used to work at the National Human Genome Center at Howard, says she prefers to use ancestry in her research because race simply does not apply to human beings. "So what we are talking about, call them ancestral groups, call them ethnic groups, call them something else," she told me.[22] Her former colleague at Howard, Charles Rotimi, now the director of the NIH Center for Research on Genomics and Global Health, takes a similar tack. When I spoke with him about the center's research initiatives, I noticed that he had not used the word *race* during our entire conversation. "That actually is quite deliberate," he replied. "What I was describing has nothing to do with race. I don't use race because I know the people I study are not races; they are ethnic groups."[23]

Using *ancestry* can also be a way to acknowledge that individuals inherit traits from groups whose members share genetic similarities, while reserving *race* to designate a political category. "People are born with ancestry that comes from their parents but are assigned a race" is how Camara Jones, a research director at the Centers for Disease Control (CDC), explains it.[24] Ancestry is a far more accurate tool than race for describing human genotypes because it can reflect the true nature of individual and group heterogeneity. An individual can have ancestors from multiple geographic regions (as opposed to belonging to one race), and the regions can be defined in multiple ways, from small local areas to entire continents.[25] In terms of genetics, ancestry gives a better account than race, for example, of someone whose Irish-descended mother was born in Wichita, Kansas, and whose father came from the Luo group in Nyangoma-Kogelo, Kenya.

But while some scientists reject race altogether and others distinguish between (biological) ancestry and (social) race, an increasingly prominent trend is to redefine race *as* genetic ancestry. Concerned about how the Human Genome Project should deal with the subject of race, Robert Cook-Deegan, policy advisor to HGP director James Watson, wrote a letter to population geneticist Luca Cavalli-Sforza in 1989 asking him "what genetic research revealed about the reality of race and whether human genomics could lead to a new racism."[26] In a long and detailed reply, Cavalli-Sforza agreed that it was difficult to distinguish between groups based on genetic traits: "[W]hy classify races if the result is arbitrary and uncertain?" But he stopped short of refuting the existence of races at the genetic level or denying

the validity of using the concept of race in genomic science. What he objected to were the "social" constructions of race popular with the "man on the street" based on visible physical traits, such as skin color and hair texture. Far from rejecting its scientific validity altogether, Cavalli-Sforza replaced the popular conception of race with a more scientific one—a "genetic definition of race" discovered at the molecular level using advanced scientific methods.

In 2002, genetic scientists Neil Risch and Esteban Burchard went further, erasing any distinction between race and ancestry.[27] Instead of distinguishing between a "man on the street" concept and a genetic concept of race, as Cavalli-Sforza had, they defined race "on the basis of the primary continent of origin." Ancestry, on the other hand, "refers to the race/ethnicity of an individual's ancestors, whatever the individual's current affiliation."[28] In other words, race is where one's ancestors come from, and ancestry is the race of one's ancestors. Their concession that "migrations have blurred the strict continental boundaries" did not dissuade them from associating race with continental origin, nor did their observation that Ethiopians and Somalis of East Africa, as well as North Africans, are "intermediate between sub-Saharan Africans and Caucasians" owing to their genetic resemblance to Caucasians. "The existence of such intermediate groups should not, however, overshadow the fact that the greatest genetic structure that exists in the human population occurs at the racial level," they stated, never explaining why we should simply ignore the blurred boundaries and intermediate groupings to uphold an equation between race and ancestry.[29] The definition of race in a 2010 article in the *Pharmacogenomics Journal* as "population clusters based on genetic differences due to evolutionary pressure" that is "often used to imply geographic or genetic ancestry" is increasingly common.[30] These scientists are treating *social* categories, determined by law, custom, and political affiliation, as if they are *biological* ones.

Studies that seek to discover natural groupings of human beings are only as informative as the populations they sample. For their part, Risch and Burchard relied on genetic material from two or three indigenous groups from each of the five continents, which together were supposed to represent the entire human race.[31] In another study, three sub-Saharan populations—two pygmy groups and the Lisongo—stood in for Africa; while Chinese, Japanese, and Cambodians represented East Asia; and Northern Europeans and Northern Italians were the Caucasians. Sampling a handful of ethnic groups

to symbolize an entire continent mimics a basic tenet of racial thinking: that because races are composed of uniform individuals, any one can represent the whole group.[32] "Even our view of the Big Few might change were it not for our curious convenience of overlooking places such as India," wrote medical geneticist Rick Kittles and biological anthropologist Kenneth M. Weiss.[33] People from India don't fall neatly into an "Asian" genetic cluster. A 2003 study of fifty-eight DNA markers from many Indian populations, for example, traced their ancestral lineages to Africa, Central Asia, southern China, and Europe.[34]

This flawed sampling method is now built into the infrastructure of genomics research. A major initiative to document human genetic variation, the Human Genome Diversity Project (HGDP), relied on samples drawn from groups assumed to be geographically separate and isolated.[35] Based on the fear that many indigenous tribes were on the brink of extinction, HGDP scientists collected DNA from such groups around the globe to preserve this "precious genetic information" before it vanished. Led by Cavalli-Sforza, who launched the project in 1991, research teams descended on more than seven hundred indigenous communities worldwide to take blood from dozens of their members. In addition to "immortalizing" indigenous genes, the HGDP analyzed the genetic data to compare variation among indigenous groups in order to "facilitate studies of the genetic geography and history of our species," Cavalli-Sforza wrote.[36] The HGDP met an ignominious and unexpected end when tribal leaders accused the scientists of biocolonialism, for exploiting native genetic information in the same way that European colonizers had exploited their ancestors' natural resources.[37] The cell lines derived from the samples live on, however. Most of the research on human population structure, including the Rosenberg study cited above, used the HGDP samples for their data sets.

The relatively small number of indigenous populations sampled for the HGDP archives do not represent humankind's genetic diversity, nor do they paint an accurate portrait of the migrations and intermixing that contributed to most contemporary groups. Similarly, geographic areas with high levels of intermixture—North Africa, Spain, the Middle East, and the Balkans, for example—are rarely included in genomic studies.[38] Northern and eastern Africans are never selected to represent the continent because they do not fit the profile of "black" Africans—they have mixed too much with Europeans, Arabs, and other non-Africans. Even assuming the isolated

indigenous groups sampled by the HGDP are genetically "pure," their unusual purity is all the more reason they cannot stand in for all the other populations of the world that are marked by intermixture from migration, commerce, and conquest.

A more accurate study of human genetic variation would use an objective sampling method. It would select the populations randomly and systematically across the globe, including those that reflect historic intermingling, instead of cherry-picking groups that best fit a priori racial classifications.[39] If researchers collected DNA samples continuously from region to region throughout the world, they would find it impossible to infer neat boundaries between large geographical groups that look like races.

When they do not attempt to fit findings into predetermined boxes, genomic population studies have discovered that (1) many of the individuals sampled fit in more than one cluster, and (2) the clusters nonetheless leave out whole groups of people who do not fit anywhere. Almost all of the Mozabites from Algeria, for example, belong to both Eurasian and African clusters. A 2008 *Science* study conducted by Cavalli-Sforza and Marcus Feldman analyzed a data set of 938 individuals from fifty-one populations at 650,000 common SNPs with an enhanced computer program called Frappe. Although the scientists were able to segregate the populations into five continental groups, their more significant findings challenged this simple breakdown, revealing mixtures among many groups such as Palestinians, Druze, and Bedouins, who have ancestral contributions from the Middle East, Europe, and South/Central Asia.

The expanded analysis also detected "finer substructures" when individual regions were examined separately. The East Asian populations divided into a "north–south genetic gradient," Europeans separated into eight populations, and those from the Middle East divided into four. Similarly, when Michael Bamshad of the University of Washington ran Structure using DNA from several continents, he found two separate sub-Saharan clusters: one consisted of the Mbuti, one of the indigenous pygmy groups in the Congo, plus three stray individuals; the other consisted of all sub-Saharan Africans *except* the Mbuti and the three other individuals.[40]

While the computer-generated findings from all of these studies offer greater insight into the genetic unity and diversity of the human species, as well as its ancient migratory history, none support dividing the species into discrete, genetically determined racial categories.[41] In 1994, at the outset of

genetic-clustering research, Cavalli-Sforza predicted that classifying clusters as races would prove a "futile exercise" because "every level of clustering would determine a different population and there is no biological reason to prefer a particular one." Unfortunately, this did not stop Cavalli-Sforza from illustrating his work with color-coded world maps showing "four major ethnic regions": African in yellow, Mongoloids in blue, Caucasoids in green, and aboriginal Australians in red.[42]

The way some population geneticists treat the mixing of different groups also illustrates how genomic research is organized by race. If there are no pure races, we should not conceive of people with mixed ancestry as being a combination of two or more pure races. But this is exactly how many genomic scientists describe what they call racial "admixture."[43] An article titled "Reconstructing Genetic Ancestry Blocks in Admixed Individuals," co-authored by Neil Risch, states, "If the admixing occurred recently, we can imagine that each chromosome was assembled by stitching together long segments of DNA from a particular ancestral population." The authors refer to these imagined chromosomal segments that are identified with the component populations as "ancestry blocks."[44] But remember, this is how they *imagine* what their theory of admixing pure populations would look like.

A scientist from the Rosenberg team whose talk I attended used this same building-block imagery to explain how African Americans fit into the continental clusters the team identified. Noting that a large portion of African Americans have both European and African ancestry, he urged the audience to think of African American genomes "as a series of pieces that come from one or the other population." To illustrate this point, he showed a picture of a string of yellow and green blocks, representing an African American individual's chromosome. "As I go along the chromosome," he told the audience, "I can actually estimate which bits come from European ancestry, and which bits come from African." On the screen was projected a color-coded genome, with yellow and green blocks symbolizing "European" and "African" genes.[45] This graphic left the impression that, at the molecular level, African Americans are composed of distinguishable pieces of pure European and pure African ancestry neatly strung together. A 2010 article on admixture in the *Pharmacogenomics Journal* included a similar graphic with blue and red chromosomes representing different ancestral populations and a caption explaining that "the admixed individual's genomes are a mosaic of the two initial ancestral chromosomes."[46] These pictures of

color-coded genomes represent what genomic scientists imagine about race, supported by statistical calculations based on racial assumptions, not on discoveries of scientific fact.

From Segregated Gene Banks to Color-Coded Genomes

Another problem with replacing a social conception of race with race based on geographic ancestry is that the distinction gets completely blurred when genetic scientists use race as a variable in laboratory research. These researchers typically abandon the usual rigor applied to scientific studies in order to classify their DNA samples, analyze their data, and report their findings according to race. A young medical anthropologist from Harvard, Duana Fullwiley, trailed scientists in two biopharmaceutical labs to investigate firsthand how they categorized the genetic samples used in their research. During a six-month fieldwork stay at the lab run by Esteban Burchard at University of California at San Francisco's department of biopharmaceutical sciences, Fullwiley interviewed researchers investigating the pharmacogenetics of cell membrane transporters, molecules that are vital to drug delivery. She discovered that race served as an unquestioned organizing principle for the collection, analysis, and reporting of genetic data.[47] Far from carefully scrutinizing the scientific validity of racial classifications, the researchers simply inserted race into their studies.

Fullwiley found that the laboratories practiced an extreme form of racial segregation at the genomic level. To obtain molecular data for their research, the scientists purchased DNA from the Human Genetic Cell Repository located at Coriell Institute for Medical Research, a nonprofit company in Camden, New Jersey, which houses the world's largest collection of human cell lines available for scientific research. Coriell labels samples according to the self-reported race of the donors, so the genetic material arrived at the lab already classified by race. Unsatisfied with Coriell's racial labeling, Burchard applied for a grant to build a genetic database specifically for his research that collected more "racially pure" DNA by "excluding anyone who reported racial mixing in their genealogies for the past three generations."[48] Burchard believed that this sampling method would allow him to segregate the DNA in his lab according to a more accurate test for race. Yet even his own, supposedly more rigorous, criteria still incorporated the DNA donors' own social definitions of race. Thus the concept of biological race is stamped

on the very raw materials that go into pharmacogenetic studies, starting researchers on a trajectory that shapes the scientific conclusions they reach.

Once, Fullwiley noticed that two young investigators assigned to analyze the racial breakdown of a particular gene spent days "playing with" the data by applying various statistical tests using two different software programs. When Fullwiley asked about this exercise, one of the researchers confided that they were trying to manipulate the data to make the racial associations appear stronger. "These genotypes are specific to Caucasians, and we know that they are different in minority groups," the researcher explained. "So we want to make that difference stand out, which needs to be done, or else science will never change. People will just keep looking at Caucasian genes." The researcher apparently believed that showing nonwhite, race-specific variations would make minorities more worthy of study.[49]

The lab scientists not only assumed that African American and Caucasian DNA samples had significantly different allele frequencies, but they also perceived each as the other's "*opposite* race." They predicted that black and white DNA would always produce dramatically disparate findings. When researchers found results that were inconsistent with their perception of racial categorization, instead of rethinking their presumptions about racial sameness and difference, they usually "reacted against the data," writes Fullwiley. So when African American allele frequencies turned out to be more similar to Caucasian ones than expected, one scientist concluded the racially labeled samples must have been contaminated.[50]

The idea that blacks and whites represent opposite races is patently unscientific. Aside from the flaws inherent in treating Caucasians and African Americans as biological races in the first place, the pattern of human populations migrating out of their African homeland starting around eighty thousand years ago does not place Europeans and Africans the farthest apart. DNA studies of human evolution, buttressed by fossil and archeological evidence, show that human groups that journeyed out of Africa reached Europe about forty thousand years ago. Evolutionary biologists posit that geographic distance is a good predictor of genetic distance, and parts of Africa and Europe are swimming distance from one another. The intimate intertwining of Europeans and Africans in the ensuing centuries through trade, conquest, enslavement, and migration make it absurd to consider them opposites from a genetic standpoint.

So where does the notion of "opposite race" come from? It is part of an

ideology about race that pits blacks against whites as the moral and social antithesis of each other.[51] This fictional opposition expresses the fundamental contrast between the innate character of each group essential to U.S. racial hierarchy. This ideology manifests itself in familiar (though often implicit) racial stereotypes that paint blacks as having a negative trait for every positive trait possessed by whites: blacks are lazy, while whites are industrious; blacks are ugly, while whites are beautiful; blacks are ignorant, while whites are smart; blacks are criminal, while whites are law-abiding; and so on.[52] Taken to its extreme, "opposite race" signifies not only difference but also enmity. "The black race is believed to be the perennial enemy of the white race, against whom all whites must unite," writes theologian George Kelsey. "'Opposite race' thus means 'race in opposition.'"[53] In racist circles, the worst thing you can call a white man is "nigger lover." But the idea of enemy races can also be seen in old claims that blacks are taking white people's jobs, or that Latinos are invading U.S. borders.

It is this ideological opposition, so ingrained in our racial culture—not any genetic evidence—that makes some scientists automatically think of black and white genotypes as being opposites. Of course, the researchers did not have these stereotypes in mind. To the contrary, they seemed to be accentuating assumed racial differences at the genetic level in a misconceived effort to fill a gap in research on minorities. Whatever the motivation, however, their statistical analyses of the genetic data sets were heavily influenced by social ideas about race.

Two anthropologists from Michigan State University, Linda Hunt and Mary Megyesi, found the same lack of scientific rigor when they interviewed thirty genetic scientists about their use of racial classifications in their research. The scientists, who held medical degrees or PhDs in fields ranging from human genetics and molecular biology to biostatistics, were all principal investigators in research projects in which race was a central variable. Most of the researchers categorized the DNA samples they worked with according to the familiar racial categories adopted by the federal government for the census and other administrative purposes: American Indian, Asian/Pacific Islander, black, white, and Hispanic. They simply lumped together people from different ethnic groups and geographic locations into these large social classifications, without any valid biological justification. "For example, samples collected from a relatively isolated village in rural China were described as 'Asian,' as were those taken from individuals of partial Japanese

heritage living in suburban Detroit," wrote Hunt and Megyesi in their 2008 article. "Or a person labeled as 'black' in a study in San Francisco was classified as belonging to the same ancestral group as individuals being sampled in Nigeria."[54]

What's more, the researchers did not use any scientific criteria or specialized language to describe the racial categories. Instead, they dredged up the familiar colloquial labels that the average person on the street would use in identifying someone's race: Caucasian, white, Jewish, Hispanic, Mexican, African American, and so forth. Many of the classifications were nonequivalent (juxtaposing skin color with national origin, for example) as well as overlapping, so researchers had to make a subjective decision about where to place some subjects. A medical doctor studying the genetic basis for chronic disease in African Americans developed his own idiosyncratic technique for handling mixed ancestry among his research subjects. "The way we classify people sort of minimizes the admixture of whites," he explained. "You don't get considered 'white' if you look too much black or you look too much phenotypically nonwhite or if you have certain type of hair—you don't get called just plain 'white.'"[55]

An endocrinologist used a computerized list of eight thousand Spanish surnames to classify research subjects as Hispanic. "We took the view that if you have a Hispanic surname, you're Hispanic until proven otherwise," he stated.[56] Who knows how many people with Spanish names who had no Hispanic ancestry (whatever that means) were admitted to the study or how many Hispanics without Spanish names were left out—but treating people with certain last names as a genetic grouping is no sillier or more arbitrary than the other methods scientists use to make race seem like a biological classification.

One wonders how genetic scientists using widely varying, inconsistent, arbitrary, and ambiguous definitions of racial categories can possibly rely on or replicate the results from studies dependent on such classifications (or get them published in respectable journals). Because the definition of race varies across countries, it is even more hazardous to link data from race-based genetic research conducted globally. Added to this confusion is the fact that the research subjects who contribute the genetic data typically identify their own race, without any uniform criteria. Researchers have no way of knowing whether or not (or how) the participants are applying identical racial identity tests, and it is very unlikely they all define race precisely the same

way. Many scientists use inconsistent definitions of race even within the same article, shifting from self-identified race to describe their research subjects to a classification of genotypes when discussing their findings.

Published reports of biomedical and genetic studies rarely describe how race was determined or the rationale for analyzing the data on the basis of race. "The lack of disciplinary clarity or consensus with respect to a central term of analysis . . . was not a barrier to publication of thousands of articles evaluating racial differences in a host of medical conditions," reported a survey on the use of race variables in genetic studies.[57] Some medical and scientific journals have addressed these methodological errors by adopting editorial policies that require more rigorous scrutiny of racial variables. In 2000, the editors of *Nature Genetics* declared that the journal would start requiring that "authors explain why they make use of particular ethnic groups or populations, and how classification was achieved," writing, "We hope that this will raise awareness and inspire more rigorous design of genetic and epidemiological studies."[58] Despite these improvements, unscientific gaps remain. As Margaret Winker, deputy editor of the *Journal of the American Medical Association* (*JAMA*), notes, "Still lacking are careful consideration of what has actually been measured when race/ethnicity is described, consistent terminology, hypothesis-driven justification for analyzing race/ethnicity, and a consistent and generalizable measurement of socioeconomic status."[59]

In no other field do scientists routinely use such a poorly defined variable as a critical component of their research. In the field of genetics itself, "genetic and disease variables are carefully defined and systematically classified," Hunt and Megyesi point out.[60] And yet the findings produced by scientists' faulty use of race as a research variable are taken to confirm the very racial categories that are being employed in such a sloppy manner. The public and major media outlets assume that researchers claiming to show racial disparities at the genetic level must have used rigorous scientific methods to define racial classifications, identify the race of research subjects, and group them with others in the same category. Nothing could be further from the truth.

When social scientists or legal scholars point out these flaws, we are often accused of meddling where we have no place. It appears to be a common belief that genomic and biomedical researchers should be left alone to investigate race objectively at the molecular level, while sociologists and their ilk should stick to understanding how race functions in society. "In an unadulterated scientific environment," wrote Rick Carlson, then a clinical professor at the University of Washington School of Public Health, "racial variables would be weighted by the same measures applied to other variables, such as temperature." He bemoaned that, in "addressing questions about race and genetics, social sciences have achieved parity with the 'harder' life sciences." The influence of social approaches to race creates, he argued, "a real peril that lowbrow theories wrapped in tendentious and oily slogans will get the public's ear and gain even footing with scientific proof as worthy of belief," comparing the debate about intelligent design and human evolution to the debate about social science and biological perspectives on race. Here again, we see the refrain that biological studies of race are ipso facto scientifically valid while the mountain of evidence that race is a political category amounts to "social science posturing."[61] This closed-minded faith that racial science must be true helps to shield scientists' flawed methods from public scrutiny.

Although genetic researchers routinely force genetic samples into preexisting racial categories, they often have trouble articulating what race means or even identifying what their own races are. One of the oddest discoveries Fullwiley and Hunt and Megyesi made was that some of the scientists they interviewed could not apply to themselves the racial classifications they applied to the DNA in their labs. When asked to describe his racial/ethnic background, one researcher, who was born in Mozambique, explained that he had one grandparent from Cape Verde and others from Portugal. "But if you go back a few generations, I've got people from all over the place," he elaborated. "So, I usually go for 'Other.'" The researcher revealed that he also considered identifying as Hispanic, but decided against it after noticing that the official definition doesn't include Mozambique or Portugal. Despite his confusion about his own racial category, this researcher was somehow able to categorize his clinical subjects by race.[62]

Fullwiley found a similar disconnect between the presumed reality of genetic race and the fuzziness of social race in the minds of the scientists she studied. One researcher defended the racial categories she used in the lab because she believed "there are ethnic-specific SNPs," but then conceded that she could not apply these same categories to herself because her father was Indian and her mother was part Czech. She concluded that she approached race in two divergent ways: "When I'm doing my genetic type research, I want things very well defined, and in a social setting I don't even want to think about it."[63]

These scientists seem oblivious to their reliance on social assumptions about race in conducting their genetic research. Without realizing it, they import social classifications into their work as if these classifications had biological validity. The problem ultimately lies not only in scientists' shoddy classification methods, but also in the impossible task of classifying people into a few clearly demarcated biological groups called races, even with the most advanced genomic knowledge and statistical computing at their disposal. Yet many of today's genomic scientists have faced this challenge by redefining race to fit twenty-first-century theories and technologies.

Where Does Geography Get Us?

While the scientists I have discussed so far use traditional or redefined racial categories in their research, there is another group of scientists who are trying to eliminate notions of race from their research altogether. Some biomedical geneticists who conduct genome-wide association studies (GWAS) to identify genetic contributions to common complex diseases, such as heart disease, type 2 diabetes, and cancer, are replacing race with the concept of "genetic ancestry" to differentiate populations. They use a population genetics software technology called Eigenstrat that divides DNA samples into clusters on the basis of SNP variation scores. Unlike the Structure program, Eigenstrat doesn't require prespecifying the number of expected clusters; nor does it depend on any presorting of samples using race, ethnicity, ancestry, or a theory of human evolution. Thus, it enables GWAS researchers to "create *categories of genetic difference* that are *not* categories of race," write University of Wisconsin sociologists Joan Fujimura and Ramya Rajagopalan.[64]

Yet Fujimura and Rajagopalan discovered that these medical geneticists tend to translate the Eigenstrat clusters of genetic similarity in terms of "genome geography": the scientists believe that individuals with similar SNP variation scores are likely to have "shared ancestry" that traces to a specific geographic location. Similarity is assumed to mean relatedness, which is assumed to mean a common geographic origin, such Europe, Africa, or Asia. In other words, genetic ancestry is equated with geographic ancestry. Just as happened with the Rosenberg study, the media, the public, and other researchers often read the genetically similar populations identified by a computer program as racial categories.[65]

Geographic ancestry does not solve the problem of race. If you look at a map of the world, you will see that parts of Africa are very close to Europe and the Middle East and other parts are very far from these regions. Because they are closer to the Arab Peninsula, African Somalis are genetically more similar to people in Saudi Arabia than they are to people in western or southern Africa. Likewise, the Saudis are more similar to the Somalis than to Norwegians, who are geographically more distant.[66] Yet molecular geneticists routinely refer to African ancestry as if everyone on the continent is more similar to each other than they are to people of other continents, who may be closer both geographically and genetically.

The same is true for Europe and Asia. We speak about the two "continents" as if they are very far apart. But Europe occupies the same land mass as Asia. England is much closer to Turkey, the nation seen as bridging the two continents, than it is to the eastern edge of Russia. Most of Russia is much closer to China than it is to Germany. The Rosenberg cluster study actually identified Eurasia—comprised of Europe, the Middle East, and Central Asia—as one of the five main continental groups. Yet newspaper coverage of the Rosenberg study conveniently ignored how the Eurasian cluster contravened, rather than confirmed, everyday racial categories.

Many geographic boundary lines are not dictated by natural barriers; they are drawn by political deal-making in the wake of wars, colonialism, and negotiated treaties. Consider the Middle East or the Arab world. The region comprising twenty-two countries in northern Africa and the Arab peninsula stretches across the continents of Africa and Eurasia, from the Atlantic to the Persian Gulf. Sudan, the African country devastated by civil war and the humanitarian crisis in Darfur, is a member of the Arab League despite its large population of black groups in the south. Is Turkey, which is predominantly Muslim and geographically closer to the Middle East than Sudan, part of the Arab subgrouping, or is it part of Asia? Or Europe? Or both?

The misperception of continental populations as natural groupings is grounded in a broader concept of populations as natural, isolated, and static. Populations came to be seen as "bounded units amenable to scientific sampling, analysis, and classification" as a result of Western linguistic and anthropological studies of indigenous peoples in the late nineteenth century and the first half of the twentieth century, historians of science Lundy Braun and Evelynn Hammonds show. They trace the scientific framing of African populations to European missionaries at the end of the nineteenth century who condensed diverse tongues of multiple groups into a single written

language in order to introduce these groups to the Bible and convert them to Christianity. Later, colonial administrators foisted unified tribal identities on people who were geographically dispersed, spoke a variety of dialects, and had not previously felt any political allegiance to each other. When social anthropologists conducting extensive fieldwork in Africa in the 1930s and '40s treated tribes as self-contained units of study, they hardened the view of populations as bounded, fixed, and natural entities.[67]

"Once named and entered into international atlases and databases by anthropologists in the U.S.," write Braun and Hammonds, "the existence of populations as distinct, naturally occurring and static formations became self-evident, thus setting the stage for their use in large-scale population genetic studies—and for the reinvigoration of broad claims of human difference based on population identity."[68] In designing the Human Genome Diversity Project, Cavalli-Sforza relied heavily on the comprehensive atlas of the world's people compiled by the Yale anthropologist George Peter Murdock in 1967. These seemingly objective groupings based on geographical ancestry, in turn, are taken as verification of the racial classifications that they appear to mirror.

Population genomics trades the fallacy that races are naturally bounded by biology with the fallacy that populations that map onto races are natural formations that became biologically cohesive. While it was once thought that races are created from a biological essence, it is now thought that populations create a biological essence that mirrors race. But we should challenge genomic scientists who take it for granted that human beings are naturally organized into definable, genetically cohesive populations. If we pause for a moment to examine the political, cultural, and even arbitrary borders that delimit populations and consider how mutable, porous, and continually changing these boundaries are, the scaffolding of population genomics that seems to be supporting race begins to look very wobbly.

It is true that ancestry explains why some groups are more genetically similar than others better than race. Ancestry, at least, is a biological mechanism, whereas race is a political relationship. While categorizing someone by race requires an "other," tracing someone's ancestry is more politically neutral. But scientists run into trouble when they simply substitute geographic ancestry for race in order to dodge controversy or to give race biological legitimacy. Cramming findings about genetic ancestry into predetermined racial categories—or geographic populations that map onto them—turns

serious objections to the biological concept of race into a battle over semantics and political correctness, ignoring the grave political consequences of dividing people into races. It appears that many scientists do not even believe this distinction makes a difference: they have concocted a thinly disguised euphemism for race they hope will not stir up as much controversy. Geographic ancestry has not replaced race—it has modernized it.

The Enduring Faith in Race

How should we view the progression of racial science that brings us to the doorstep of a new genomic concept of race? Are the errors of the past—the legend of Ham, Cuvier's racial typology, craniometry, eugenics—the product of flawed research methods that today's scientists have corrected with advanced genomic theories, state-of-the art computing, and giant DNA databases? Did Blumenbach happen to devise an accurate classification of human races in 1795 despite using a faulty technique? The answer: there are no biological races in the human species. Period. That conclusion was confirmed by the most ambitious research project on human biology yet undertaken, the Human Genome Project. A mountain of evidence assembled by historians, anthropologists, and biologists proves that race is not and cannot possibly be a natural division of human beings. Think about the origins of the concept of race, the way racial groupings have been reconfigured over time, or their differing meanings around the world, as I described in part 1. Race *must* be a political category.

Why, then, do most Americans cling to a false belief that biological races really do exist? Why do they latch on to whatever trivial proof they can muster to confirm their misconceptions about race? I am not referring to redneck white supremacists who spout vitriol about the inferiority of colored people. Many of my left-leaning colleagues, for example, balked at my book project. "Of course we should be working toward racial equality," they said, "but what if scientists *are* able to identify races genetically?"

"Racism is a faith." This was the summation of George D. Kelsey, the prominent black theologian who mentored Dr. Martin Luther King Jr. Kelsey argued that racism initially arose as an ideological justification for colonialism and slavery but gradually "heightened and deepened in meaning and value so that it pointed beyond the historical structures of relation, in which it emerged, to human existence itself."[69] The same can be said of race itself,

the system of human classification that facilitates racism. Race started as a crude device for parceling people into servant and master classes, whose historical roots and scientific rationales we now reject. Yet race has outlasted its original historical context because it developed into a deeply held belief about the nature of human beings, a belief that continues to be useful in ordering our contemporary society.

Most Americans do not deduce that biological races exist from sound scientific evidence and reasoning. They are inculcated with this belief in the same way a child is raised in a religion. Children in the United States learn to divide all people into racial groups and come to have faith in race as a self-evident truth, like a traditional creation story that explains how the world works. Anthropologists describe the common meaning of race that defies scientific facts as a "folk concept." This is why Ashley Montagu called race "the witchcraft of our time." In 1942, he wrote, "It is the contemporary myth. Man's most dangerous myth."[70] Perhaps the best proof of its power is Montagu's own inability to renounce race definitively in the 1950 UNESCO statement.

According to folklorist Judith Neulander, for a folk story to persist it must contain "elements that can be modified without changing what the tale is about," enabling it "to dodge later discreditation."[71] Science has been responsible for giving racial folklore its superficial plausibility by updating its definitions, measurements, and rationales without changing what the tale is about: once upon a time human beings all over the world were divided into large biological groups called races.

Believing in race can be compared to believing in astrology. People who have faith in astrology find constant confirmation that horoscope predictions are reliable and that astrological signs determine personality types. For the faithful, the twelve divisions of the zodiac are as accurate as Blumenbach's five divisions of human beings. The funny thing is, biostatisticians can find significant medical differences according to astrological signs. In the 1990s, a major randomized clinical trial compared the effectiveness of an intravenous drug, an oral aspirin, and a placebo to treat 17,000 patients who were hospitalized with signs of a heart attack. The study found a huge overall statistical benefit for patients who got the aspirin over the placebo. To test the strength of the outcome, the researchers divided the patients into twelve subgroups by their astrological signs. They found that the zodiac made a difference: their statistical analysis showed that patients born under

Gemini or Libra suffered an adverse effect from aspirin.[72] Unsurprisingly, physicians laughed off this finding because it was more scientifically plausible to interpret the results as an insignificant coincidence. But an astrology enthusiast would take it as proof that zodiac signs determine people's health and drug response.[73]

If race is a faith, like astrology, how can science resolve the debate over its meaning? If scientists could settle the question whether or not race is biological, we would not still be debating it. As we have seen, a generation of scholars has definitively refuted prior versions of racial science by revealing their errors and biases.[74] The work of scientists is invaluable for dispelling misconceptions about the biological definition of race, and scientists should continue to educate the public about its scientific invalidity. I, too, felt it was crucial to lay an empirical groundwork in part 1, to use genetic science to challenge the racial lies still being told about humanity. Many Americans' belief in race depends on sheer ignorance of the scientific evidence about human genetic unity and diversity.

But the resilience of racial science raises serious doubts about the efficacy of fresh efforts to debunk it based on more accurate and less prejudiced scientific methods. Race is a political system that will not be brought down with scientific evidence alone. Race persists neither because it is scientifically valid nor because its invalidity remains to be proven. Race persists because it continues to be *politically useful*. It is therefore imperative to evaluate the political function of race at the present time and wage a political assault against it. I realize this talk of politics will be called unscientific by those invested in preserving the biological view of race. But as many scientists have shown, the science of race has long been riddled with unscientific flaws. And as both history and current practice clearly demonstrate, racial science and politics are inseparable.

Believing in biological races, not only racism, is an irrational moral conviction that scientific evidence alone has been unable to overcome. In fact, over the course of U.S. history, scientists themselves have worked as much to uphold this moral conviction as to defeat it. In November 2010, University of Chicago business professor Richard Thaler, coauthor (with Cass Sunstein) of *Nudge: Improving Decisions About Health, Wealth, and Happiness*, asked contributors to the Internet forum Edge to name their favorite examples of "wrong scientific beliefs that were held for long periods of time." The responses listed dozens of false but durable scientific theories, ranging from

"life generates spontaneously" to "stomach ulcers are caused by stress" to "genes are made of protein."[75] I hope one day the scientific theory of biological races will make the list. Scientists' defining mission to test accepted beliefs utilizing empirical evidence remains an important aspect of challenging antiquated views about race. It is the faith in race—the religion of separating human beings into racial groups—that makes it difficult for Americans to think like scientists.

4

Medical Stereotyping

Imagine if every single day a jumbo jet loaded with 230 African American passengers took off into the sky, reached a cruising altitude, then crashed to the ground, killing all aboard. According to former surgeon general David Satcher, this is exactly the impact caused by racial health disparities in the United States. In a 2005 article, he and several other health experts reported that there had been 83,570 "excess" black deaths in 2002. That represents the number of African Americans who would still have been alive that year if their life expectancy were the same as that of whites.[1] The number of excess deaths is closer to 100,000 today.[2] In one generation, between 1940 and 1999, more than 4 million African Americans died prematurely relative to whites.[3] Overall life expectancy is actually declining in some counties where there is a high proportion of African Americans.[4]

In my hometown of Chicago, one third of all black deaths are excess in terms of the black–white mortality gap. In other words, one out of every three black people who died in 2000 would have survived if black and white death rates were equal. Chicago is a very segregated city, so longevity varies geographically. Of Chicago's seventy-seven community areas, the twenty-two with the lowest life expectancies are more than 90 percent black. There is a difference of sixteen years between the white neighborhood with the highest life expectancy and the black neighborhood with the lowest. Blacks are more likely to die prematurely (before the age of sixty-five) from most major illnesses: cancer, stroke, diabetes, kidney disease, AIDS, and coronary heart disease, to name a few. Race matters at the beginning of life as well.