

UC Berkeley

UC Berkeley Previously Published Works

Title

Social Diversity in Humans: Implications and Hidden Consequences for Biological Research

Permalink

<https://escholarship.org/uc/item/0zv0p60p>

Journal

Cold Spring Harbor Perspectives in Biology, 6(5)

ISSN

1943-0264

Author

Duster, Troy

Publication Date

2014-05-01

DOI

10.1101/cshperspect.a008482

Peer reviewed

Social Diversity in Humans: Implications and Hidden Consequences for Biological Research

Troy Duster

Chancellor's Professor & Senior Fellow, Warren Institute on Law and Social Policy, Boalt School of Law, University of California Berkeley, Berkeley, California 94720

Correspondence: troy_duster@berkeley.edu

Humans are both similar and diverse in such a vast number of dimensions that for human geneticists and social scientists to decide which of these dimensions is a worthy focus of empirical investigation is a formidable challenge. For geneticists, one vital question, of course, revolves around hypothesizing which kind of social diversity might illuminate genetic variation—and vice versa (i.e., what genetic variation illuminates human social diversity). For example, are there health outcomes that can be best explained by genetic variation—or for social scientists, are health outcomes mainly a function of the social diversity of lifestyles and social circumstances of a given population? Indeed, what is a “population,” how is it bounded, and are those boundaries most appropriate or relevant for human genetic research, be they national borders, religious affiliation, ethnic or racial identification, or language group, to name but a few? For social scientists, the matter of what constitutes the relevant borders of a population is equally complex, and the answer is demarcated by the goal of the research project. Although race and caste are categories deployed in both human genetics and social science, the social meaning of race and caste as pathways to employment, health, or education demonstrably overwhelms the analytic and explanatory power of genetic markers of difference between human aggregates.

Two contradictory magnetic poles pull medical research on humans in opposite directions, producing a tension that will never be resolved. On the one hand, there is a universalizing impulse—based on a legitimate assumption that human bodies are sufficiently similar that vaccines, catheters, pasteurizing processes, and tranquilizers that work in one population will work in others. On the other hand, and unless and until research protocols establish and confirm specific similarities across populations, there is sufficient human variation that targeting

medicines for specific populations can be a legitimate—even vital—empirically driven task. The theoretical question, of course, is why a particular population or subpopulation is to be so targeted? Because of folk theories about different groups' biological difference, or because of their social and political standing? Age, gender, and race leap to the forefront. The history of research on ailments as disparate as breast and prostate cancer (Rothenberg 1997; Wailoo 2011), heart disease (Cooper et al. 2005), and syphilis (Jones 1981; Reverby 2009) provides

Editor: Aravinda Chakravarti

Additional Perspectives on Human Variation available at www.cshperspectives.org

Copyright © 2014 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a008482

Cite this article as *Cold Spring Harb Perspect Biol* 2014;6:a008482



T. Duster

strong evidence that the answer is not either/or but *both*. So, on what grounds do we choose one strategy over the other?

And it is precisely on this point that Steven Epstein (2007) raises the most fundamental question:

Out of all the ways by which people differ from one another, why should it be assumed that sex and gender, race and ethnicity, and age are the attributes of identity that are most medically meaningful? Why these markers of identity and not others? (Epstein 2007, p. 10)

The answer is profoundly social and political, economic, and cultural. The United States is the only country in the world that, as public health policy, does not operate on the assumption of the single standard human.

Moreover, by highlighting certain categories, there is the unassailable truth that other categories are thereby ignored. But more to the theoretical point, because each of the categories noted above has a potential or real biological base in either scientific or common sense understandings (Schutz 1962), when scientists report findings indicating differences, the danger is that these findings can seductively divert policymakers from seeking alternative interventions that could better address health disparities (Krieger 2011).

The goal of Epstein's monograph was to (a) better understand how ways of thinking about differences in human populations paved the way to try to "improve medical research by making it inclusive," and (b) explain how and why the strategies of exclusiveness got institutionalized:

Academic researchers receiving federal funds, and pharmaceutical manufacturers hoping to win regulatory approval for their company's products, are now enjoined to include women, racial and ethnic minorities, children, and the elderly as research subjects in many forms of clinical research . . . and question the presumption that findings derived from the study of any single group, such as middle-aged white men, might be generalized to other populations. (Epstein 2007, p. 5)

This shift has occurred only in the last two and a half decades, beginning with regulations that were developed first in 1986. Once again,

it is important to restate the relatively unique feature of this development as it applies mainly to the United States (Epstein 2007, p. 7). The rest of the world has continued to act on the presupposition of the standard human, at least until now. As we shall see, that is about to change.

THE GENOMIC REVOLUTION AND THE SEARCH FOR DIFFERENCES

At the end of the 20th century, the first draft of the Human Genome Map was completed, providing two kinds of hope for the near future. The first was quite explicitly about potential medical advances—that the completed map would spur the development of new kinds of therapies that would increase health and reduce the ravages of a wide variety of diseases. The second hope was more of a diffuse political aspiration, but it was loudly trumpeted at the famous White House news conference in June 2000. That was when President Bill Clinton (United States), Prime Minister Tony Blair (United Kingdom), and the two molecular geneticists who had led the public and private sector human genome projects all agreed that—citing findings from the first mapping and sequencing first draft—at the level of the DNA, there is no such thing as race.

However, regarding this pronouncement about the "end" of race, as Mark Twain once quipped about a newspaper article that reported that he had died, "the news of my death has been greatly exaggerated." So it has been with racial and ethnic categories. Indeed, there is substantial evidence that developments in several fields of inquiry relevant to molecular genetics (pharmacogenomics, pharmacotoxicology, clinical genetics, personalized medicine, and forensic science) have actually served to reinscribe race as a biological category (Duster 2005, 2006; Fullwiley 2007, 2008; Bolnick 2008; Kahn 2011; Roberts 2011).

Indeed, one of the most striking developments of the last few years has been the move by several governments to take strong protective "ownership" of the DNA of their own populations—a move designed to protect from possi-



ble biopiracy from the pharmaceutical industry in Western countries. Ruha Benjamin (2009) has called this “national genomic sovereignty,” and it represents the opposite of the universal notion of human DNA envisaged at the completion of the Human Genome Project.

On the surface, this policy frame asserts a deeply nationalist sentiment of self-determination in a time of increasing globalization. It implicitly “brands” national populations as biologically distinct from other populations, “naturalizing” nation-state boundaries to ensure that less powerful countries receive the economic and medical benefits that may result from population genomics. (Benjamin 2009, p. 341)

Mexico amended its General Health Law in 2008 to make “the sampling of genetic material and its transport outside of Mexico without prior approval . . . illegal” (Séguin et al. 2008, p. 6).

The Genomic Sovereignty amendment states that Mexican-derived human genome data are the property of Mexico’s government, and prohibits and penalizes its collection and utilization in research without prior government approval. It seeks to prevent other nations from analyzing Mexican genetic material, especially when results can be patented, and comes with a formidable bite in the form of prison time and lost wages. (Benjamin 2009, p. 344)

Mexico may be in the vanguard in so explicitly asserting its commitment to national “genomic sovereignty,” but the nation is hardly alone. India, China, Thailand, and South Africa have all issued policy statements or passed legislation designed to develop national genomics infrastructure to benefit their populations (Séguin et al. 2008b).

In 2009, the HUGO Pan-Asian SNP Consortium, an international research team led by Edison Liu of the Genome Institute of Singapore, mapped genetic variation and migration patterns in 73 Asian populations, with data coming from 11 Asian countries: Japan, Korea, China, Taiwan, Singapore, Thailand, Indonesia, Philippines, Malaysia, Thailand, and India. The results—which included a summary statement that “there is substantial genetic proximity of SEA [Southeast Asian] and EA [East Asian] populations”—were published in the journal

Science (HUGO Pan-Asian Consortium 2009). In the same year, the Iressa Pan-Asian study (IPASS) was carried out by researchers in Hong Kong, mainland China, Thailand, Taiwan, and Japan with the participation of 87 centers in nine countries in Asia (Mok et al. 2009). This study was the result of previous research suggesting that Asian populations have a different, more positive response to this cancer drug than do other populations.

The explicit heightened racial consciousness of data reporting in human genomic science was dramatically on display in the November 6, 2008 issue of *Nature*. That journal published two articles asserting triumphantly how, for the first time, the whole human genome of (a) “an Asian individual,” and then (b) of a Yoruban or “an African individual,” were now “revealed” (Wang et al. 2008).

Nature referred to the fact that James D. Watson, Nobel Prize winner, as codiscoverer of the DNA structure, and J. Craig Venter, head of the private sector group that cosequenced the Human Genome, each has had their full genomes sequenced. Both are white males. But why this particular taxonomic system for trying to sort out useful, important, or relevant “differences?” The answer lies in a closer examination of recent emerging scientific discourse about “ancestral populations” and the fluid and contested boundaries around what constitutes a “population.”

WHAT CONSTITUTES A “POPULATION” IN HUMAN GENETIC RESEARCH?

In 2007, *Science* magazine declared that genome-wide association studies (GWAS) were the scientific breakthrough of the year. Genome-wide association studies scan the genomes of large groups of individuals in search of markers that might be associated with specific common complex diseases (e.g., breast cancer). The frequency of a variant (single-nucleotide polymorphism or SNP) will differ across human populations.

Within a population, geneticists estimate the frequency at which a variant occurs in that population, based on a sample of individuals thought



T. Duster

to belong to that population. (Fujimura and Rajagopalan 2011)

Which brings us to the key question: “What is a population?” A central task is to identify frequency differences between case and control groups that might be indicative of increased risk for the particular disease being studied. In the last decade, scores of research papers have been published emphasizing ethnic and racial differences between “populations.” The term “populations” is in quotation marks for a compelling reason, namely, different researchers mean very different things when they use the term. From close observation of the laboratory work of geneticists who sample human groups across the globe, we now know that some use language group to mean a population, others take geographical boundaries; still others use already-collected data from previous research, in which it is unknown how the boundaries of the “population” in question were drawn, conceived, and implemented. In still other studies, a “population” is taken from the census; sometimes it is a “clinical population” as in those with a particular ailment—from cancer to hypertension, or from asthma to diabetes.

Yet another strategy is to find four grandparents whose ancestry can be traced to one of four broad continental groupings (Europe, Asia, Africa, the Americas). Although this may seem race neutral, on even superficial reflection, the social meaning of race is operating, because one does not mean ancestors who were Boers in South Africa, or grandparents who were European settlers in Quebec, or even great, great grandparents born in New Amsterdam.

This is but the tip of a numbing variety of factors that make human population strata and boundaries multilayered, porous, ephemeral, and difficult to identify (and thus):

Samples for genetic analysis are collected using operational criteria imposed by investigators and may be more representative of these operational criteria than actual breeding groups and gene pools. (Weiss and Long 2009, p. 704)

One of the most important tools now being deployed to examine human genetic variation is a computer-based program called STRUCTURE.

This program allows the researcher to identify patterns and/or clusters of DNA markers, and when an alignment of these clusters overlaps existing categories of race and ethnicity, there is the siren’s seductive call to reinscribe these categories as biologically meaningful (Bolnick 2008). As I have suggested elsewhere, any computer program so instructed could find SNP pattern differences between randomly selected residents of Chicago and residents of Los Angeles (or between any two cities in the world). To put it in ways that are incontrovertible, no one could expect that SNP patterns would be identical in choosing subjects randomly from two cities. As for Chicago versus Los Angeles, such a proposed research project would be deemed ludicrous, because the theoretical warrant for it would be hard to establish (unless there was some legitimate grounding for hypotheses about smog effects vs. subzero winter effects). But if all the Chicago residents selected were African American, and all the Los Angeles residents were Asian American, and those SNP patterns showed up, an uncritical audience, lay or scientific, could easily accept these findings as having some validity affirming biological or genetic racial differences.

When is difference just difference, and when is difference something that inexorably stratifies a population? The answer lies in immediate history, context, and setting—in particular, whether there have been social meanings attributed to that differentiation. The authors of an oft-cited piece in *Genome Biology* seem to acknowledge this when they say:

Finally, we believe that identifying genetic differences between racial and ethnic groups, be they for random genetic markers, genes that lead to disease susceptibility, or variation in drug response, is scientifically appropriate. What is not scientific is a value system attached to any such findings. Great abuse has occurred in the past with such notions as “genetic superiority” of one particular group over another. The notion of superiority is not scientific, only political, and can only be used for political purposes. (Risch et al. 2002, p. 11)

Although the sentiment is admirable, this formulation constitutes a fundamentally flawed



notion of a firewall between “science” and “politics.” All societies make sharp differentiations among their members that permit stratifying some groups over others. When humans create categories such as “caste” or “ethnic group” or “race,” those taxonomies are political, and they are stratified in the most basic meaning of hierarchy: power-based differential access to resources. These three categories routinely predate and prefigure scientific inquiry, but, as I will demonstrate, profoundly constrain that inquiry. Over time, the interaction between living at the top or bottom of a stratified hierarchy produces systematized differential access to the rawest human needs. This means that there will be a feedback loop to various health and illness outcomes to those different “populations” (i.e., so stratified). If that seems abstract, here is a poignant example of that feedback loop.

Syngenta is one of the world’s leading agribusiness companies, with more than 25,000 employees in nearly 100 countries across the globe. According to its official website, the company is dedicated to increase crop productivity through scientific advances, and to “protect the environment and improve health and quality of life.” Syngenta has a plant in St. Gabriel, Louisiana, where it manufactures a crop-enhancing product called atrazine. But atrazine has an unfortunate side effect—it “demasculinizes and feminizes” vertebrate animals who are exposed to it by inducing aromatase.¹ When humans are exposed to atrazine for sustained periods, they are at a much-increased risk for certain cancers. The production facility in St. Gabriel has a prostate cancer rate 8.4 times higher among factory workers exposed to atrazine as compared with those in surrounding communities not exposed, and it just so happens that this plant is located in a community that is >80% African American (Hayes 2010, p. 3768).

These sharply different rates of prostate cancer between Whites and Blacks can be studied scientifically by geneticists trying to understand “population differences” through a unidimen-

sional genetic prism, but with no understanding of the larger context in which humans are exposed to environmental insults—as in the first part of the formulation by Risch et al. But we can also study the systemic pattern of African Americans living close to toxic waste dumps across the whole country (Bullard 2000; Sze 2007). That is also available for systematic empirical investigation and testable formulations, otherwise known as science. Why should the decontextualized genetic inquiry of differing prostate cancer rates between Americans of European and recent African descent be characterized as apolitical “science,” whereas the rate of their increased risk to exposure to atrazine is seen as “political” science? The answer is lodged in current culturally framed notions of the hierarchy of science. Being completely ahistorical and apolitical, we could take a sample of two different populations of Whites and Blacks in the contemporary United States, and we would find differences in their rates of hypertension. Although there is some debate about the extent of the gap, Blacks do tend to have somewhat higher rates than Whites. But as Richard Cooper and his colleagues (Cooper et al. 2005) have shown, by examining hypertension prevalence rates among 85,000 subjects, cross-cultural data demonstrate that this is not evidence for a biological difference between the races. It was explicitly designed to compare racial differences, sampling Whites from eight surveys completed in Europe, the United States, and Canada—and contrasting these results with those of a sample of three surveys among Blacks from Africa, the Caribbean, and the United States. The data from Brazil, Trinidad, and Cuba show a significantly smaller racial disparity in blood pressure than found in North America, and then, most tellingly, the authors of the study conclude:

These data demonstrate that the consistent emphasis given to the genetic elements of the racial contrasts may be a distraction from the more relevant issue of defining and intervening on the preventable causes of hypertension, which are likely to have a similar impact regardless of ethnic and racial background. (Cooper et al. 2005)

¹Aromatase causes a higher estrogen-androgen ratio (Hayes 2010, p. 3768).



T. Duster

Yet the Cooper study, which involved more than 85,000 subjects across eight nations, was not taken as seriously as the study of 1056 African American subjects in a solely U.S.-based study of hypertension (Roberts 2011). Indeed, the FDA approved a drug designed by African Americans with hypertension the spring of the very same year, after the Cooper study was published (Cooper et al 2005; Kahn 2012). Because that decision was demonstrably more about economics, patenting, and politics than about science, it is naïve to think that these factors can be neatly parsed and isolated from each other.

The Transparent Conflation of Science and Politics: Genotyping Castes in India

In the introductory section, I noted that various nations around the globe have initiated their own genome projects, using national borders as the boundaries. When governments make these decisions, they are based on geopolitical considerations, not on human taxonomies generated by scientists. Nonetheless, when scientists then deploy these categories and boundaries, they are often reinscribing those very categories with scientific legitimacy and authority. It is imperative to address a fundamental misconception that when social scientists assert that some phenomenon (caste, class, race, ethnicity) is socially constructed (Haslanger 2008), the implication is that the phenomenon being examined is “not real.” To make this point, take the example of money, or more specifically, the euro. It was obviously “socially constructed,” because the German mark, the French franc, the Italian lira, and the Spanish peso (among other currencies) were converted into the euro by social, economic, and political forces and by a collaboration of decision makers. Having been thus “socially constructed,” the euro is certainly “real.” In a parallel manner, caste is socially constructed, but no less real in its consequences for life chances.

The differences in the way members of castes in India have systematically different outcomes regarding education, health, and economic well-being (or poverty level) is a direct consequence of social, cultural, economic, and

political forces. That would hardly be surprising, because endogamy rules (who can marry whom inside specific cultural categories) have precluded marriage (and to a lesser extent mating) practices for 30 centuries. Although these rules were never universally adhered to, they provided the frame in which dissent would be sometimes tolerated, sometimes sanctioned (Dirks 2001, p. 50; Kosambi 2002, p. 319). However, in a society riven by caste differences that persist to this day, what could it mean to demonstrate that there are discernible patterns of differences in the DNA of various castes (an outcome certainly to be expected after centuries of endogamy rules governing shaping marriage options)? Given the history that I am about to tell, any such differences discovered and reported regarding their respective DNA is a weapon in the hands of those who wish to explain and sustain their privilege. To wit, when some read about evidence that they have a somewhat distinctive set of microsatellite DNA markers, or DNA haplotypes, this can become grounds for suggesting that the differences in “education, health, and economic well-being (or poverty level) is a direct consequence” of genetic differences—not the other way around.

As a general phenomenon, elites of every society come to believe that their status, their high position in the social hierarchy, is both natural and just. Whether in caste, estate, or religious systems of stratification, those at the top are either universally born to privilege or frequently anointed at an early age. In class-based systems, those who themselves may have achieved a higher-class position by being mobile across class boundaries bequeath their status to their children. The oldest system of human stratification is in what in modern days we refer to as India. For much of India’s history, the population has been divided into five major castes that do not intermarry and that have been forced into particular occupations by hereditary ascription. The top three castes are the Brahmins (priestly, literate), the Kshatriyas (mainly rulers and aristocrats), and the Vaisyas (businessmen). Together these three constitute ~17% of the population. The next group is the Sudras, who do the menial labor and are by far the largest



varna, constituting about half of India's population. The last group are the Ati-Sudras, known by a variety of names, including Untouchables. Gandhi called them Harijans, or children of God, but they currently are most likely to go by the name of Dalits, or oppressed people.

The upper castes have excluded the lower castes from schools, post offices, restaurants, theaters, and barber shops. They were denied access to the courts, and, of course, with this record of exclusion, were never permitted to be employed in any professional occupation (Galanter 1984). As the priestly caste responsible for reading and interpreting the great books, the Brahmins had a monopoly on literacy. Given this history, it is hardly surprising that the Brahmins, <10% of the population, make up more than two-thirds of the students at the premier institution of higher education in the country, the University of New Delhi.

Although many Brahmins have been trumpeting the idea that India has ended the caste system and is celebrating individual meritocracy, the caste system lives on with fierce tenacity for much of India's 1.1 billion people. A series of studies on intergenerational mobility in India have produced the unsurprising finding that there is very little social mobility in the country (Dhesi and Sing 1989). Access to most good jobs is still restricted by longstanding cultural practices, and wage discrimination operates systematically for those from the Scheduled Castes (Lakshmanasamy and Madheswaran 1995).

That molecular geneticists find "allelic frequency differences" between castes should be no surprise. However, given the vast social, economic, and political gaps between castes, findings of "genetic differences" feed a newly molecularized interpretive account for those differences (health status, educational achievement, etc.). Even though these allelic frequency differences have no known function, reports of such findings constitute the basis for a molecular reinscription of caste differences. Here is some language capturing a crucial element of this trajectory:

We genotyped 132 Indian samples from 25 groups.
... we sampled 15 states and six language families.

To compare traditionally "upper" and "lower" castes after controlling for geography, we focused on castes from two states: Uttar Pradesh and Andhra Pradesh.

We genotyped all samples on an Affymetrix 6.0 array, yielding data for 560,123 autosomal SNPs... Allele frequency differentiation between groups was estimated with high accuracy (F_{ST} had an average standard error of 60.0011...). (Reich et al. 2009)

Referring to an earlier study in which researchers found differing patterns of genetic markers between different socially designated groups, one Indian Genome Project coordinator explained that researchers "had intense debates on whether to reveal the names of communities. . . . I don't think scientists are prepared yet to understand the full social ramifications if such information is made public" (Mudur 2008).

Of course, Americans are prone to dismiss any parallels between caste domination in India and race privilege in the United States. It would hardly shock an American to learn that an imprisoned Brahmin had a better chance of employment in a decent paying job in corporate India than an Irula tribesman. Yet in the United States, the work of Devah Pager (2007) makes a powerful point that is a shocking parallel. In the last few years, Pager's research has become, quite deservedly, the poster child of a social science research project that is rigorous, critical, and saturated with both theoretical and policy implications. *The New York Times* heralded her research in a full-page report in March 2004, noting her finding that "it is easier for a white person with a felony conviction to get a job than for a black person whose record is clean" (Kroeger 2004).

This social meaning of race and caste as pathways to employment, health, or education demonstrably overwhelms the analytic and explanatory power of genetic markers of difference between human aggregates. However, there is a compelling reason why the Indian Genome Project coordinator (noted above) expressed concerns about "making public" data reporting genetic differences between castes. That concern is parallel to the molecular reinscription of race (Fullwiley 2007).

T. Duster

INSPECTING “POPULATION PURITY” IN HUMAN GENOMIC RESEARCH

In a recent paper that is yet another model of how social scientists can contribute to research in human molecular genetics, Hinterberger (2010) has looked at how Canadian scientists have been trying to better understand the biological sources of complex diseases by using GWAS. The assumption is that about 8500 French settlers arrived in Canada between 1608 and 1759. They intermarried among themselves and thus produced what is called a “founder effect”:

... the Quebec “founder effect” has provided a large volume of genomic research aimed at understanding the root of common and complex disease. In 2007, a genome-wide association study... identified multiple genes underlying Crohn’s disease in the Quebec founder populations (Raelson et al. 2007). GWAS are seen to offer a powerful method for identifying disease susceptibility for common diseases such as cancer and diabetes and are at the cutting edge of genomics-based biomedicine. (Hinterberger 2010, p. 15)

But now there is an explicit lament among these scientists who express concerns that intermarriage rates are threatening the “genetic uniqueness of these groups” and thus the opportunity for this kind of research “may be lost in the next few generations” (Secko 2008). Here is where Hinterberger (2010, 2012) steps in as the social analyst to point out the problematic unexamined assumptions of the presumed genetic homogeneity of the founding population of French settlers. What genetic researchers regard as a bounded French founder population is actually not so French after all. Specifically, in the strong pressure to convert indigenous people to Christianity, the colonizing French eagerly gave these converts French surnames (Kohl-Laven 2008). An examination of Parish records provides documentation of this, and yet it is the French name that demographers and historians have used to establish the assumption that those with French names constituted the *bounded* “French founder” population. This clearly upends the otherwise taken-for-granted assumptions about the homogeneity of this population. Without an appreciation of the social diversity

of the human population of founders, the geneticists’ stated concerns about current interbreeding diluting the “pool” are misguided. This brings us to a discussion of similar concerns about the use of ancestry markers.

With very few exceptions, ancestry-informative markers are shared across all human groups. It is therefore not their presence or absence, but their rate of incidence, or frequency, that is being analyzed. When taken together, these markers appear to yield certain patterns in people and populations tested. A specific pattern of alleles on each of a set of chromosomes that have a high frequency in the “Native Americans” sampled then become established as a “Native American” ancestry signature. The problem is that millions of people around the globe will have a similar pattern, that is, they will share similar base-pair changes at the genomic points under scrutiny. This means that someone from Hungary whose ancestors go back to the 15th century could map as partly “Native American,” although no direct ancestry is responsible for the shared genetic material. Ancestry-informative markers, however, arbitrarily reduce all such possibilities of shared genotypes to “inherited direct ancestry.” In so doing, the process relies excessively on the idea of 100% purity, a condition that could never have existed in human populations.

To make claims about how a test subject’s patterns of genetic variation map to continents of origin and to populations where particular genetic variants arose, the researchers need reference populations. The public needs to understand that these reference populations comprise relatively small groups of contemporary people. Moreover, researchers must make many untested assumptions in using these contemporary groups to stand in for populations from centuries ago, representing a continent or an ethnic or tribal group. To construct tractable mathematical models and computer programs, researchers make many assumptions about ancient migrations, reproductive practices, and the demographic effects of historical events such as plagues and famines. Furthermore, in many cases, genetic variants cannot distinguish among tribes or national groups because the groups are too similar, so geneticists are on thin ice

when telling people that they do or do not have ancestors from a particular people.

Instead of asserting that someone has no Native American ancestry, the most truthful statement would be: “It is possible that although the Native American groups we sampled did not share your pattern of markers, others might, because these markers do not exclusively belong to any one group of our existing racial, ethnic, linguistic, or tribal typologies.” But computer-generated data provide an appearance of precision that is dangerously seductive.

There is a yet more ominous and troubling element of the reliance on DNA analysis to determine who we are in terms of lineage, identity, and identification. The very technology that tells us what proportion of our ancestry can be linked, proportionately, to sub-Saharan Africa (ancestry-informative markers) is the same being offered to police stations around the country to “predict” or “estimate” whether the DNA left at a crime scene belongs to a white or black person. This “ethnic estimation” using DNA relies on a social definition of the phenotype. That is, to say that someone is 85% African, we must know who is 100% African. Any molecular, population, or behavioral geneticist is obliged to disclose that this “purity” is a statistical artifact that begins not with the DNA, but with a researcher’s adopting the folk categories of race and ethnicity.

Sampling for Human Genetic Diversity— The Conundrum²

Researchers ideally would like to sample to achieve representativeness of diversity. But in order to “sample,” one must have a notion of the boundaries of the larger population base from which one is sampling. Yet those boundaries are always going to be absent when it comes to human genetic diversity, because unless or until we have a Wilson-type grid for the world’s population, we will not have a firm empirical basis for understanding who any “sample” represents. Short of an empirical basis for proceed-

ing with a sound sampling strategy, we are then left with this conundrum of talking about “sampling” when there is no bounded population delimited by some theoretical frame. Of course, that is where race and ethnicity tend to surface in these discussions, but there are bundles of unpacked assumptions built in to the idea that any five sets of people represent five races—whether biologically or socially!

There is yet a prior question of what is meant by “diversity,” and on that matter, it is vital to be really clear on the substantive meaning of genetic diversity. If the goal is to capture genetic diversity, the strategy might aim to obtain samples from people who are presumably as genetically different from each other as possible. If that is the goal, then the researcher is simply trying to capture a wide range of specifiable variation. Here is where we must get to substance, because there are numerous dimensions and levels on which people can vary from each other genetically. Thus, the idea of a high degree of variation may not be meaningful because it is not likely to capture the type of genetic variation in which the researcher is most interested. Or conversely, such a strategy might capture a lot of variation in which the researcher has little interest, for example, variation of little apparent relevance to health outcomes.

Is the researcher trying to “represent” the range of genetic variation in a specific region? In this case, one does have a sampling problem, and unless one is assuming a level of homogeneity (that is impossible to demonstrate empirically), this cannot be done with a few dozen people. Yet the report from one of the early studies, although well-intentioned, well-crafted, and designed to help better understand health differences in variable human population groups (Hinds et al. 2005), does point in that direction. The researchers were searching for, and found, patterns of SNPs differentially distributed in three population groups, formed from a total of 71 persons who were either Americans of African descent, Americans of European descent, or Han Chinese.

The title of the paper is instructive, “Whole Genome Patterns of Common DNA Variation in Three Diverse Human Populations.” Howev-

²The following discussion is indebted to Pilar Ossorio.



er, what makes these three populations diverse is the phenotype associated with a racial classification system, not a genotypic pattern of similarity that triggered the inquiry. Indeed, the authors note that the SNP patterns of genetic diversity that they found among African Americans suggest a more substantial diversity than that in the other two populations, a finding consistent with our knowledge of genetic diversity on the African continent. So, why was the question raised in this manner? The answer is a scientific catch-22. The main reason is convenience: The data were collected and marked that way in the Coriell Cell Repositories. That is an understandable rationale. However, by deploying these existing categories, any differences that emerge are likely to be “racialized,” no matter how many caveats and demurrers appear in the text of a scientific paper. Moreover, the African American group is said to be “admixed.” But in terms of the genotype, all three groups are “admixed.” So it must be the phenotype to which the authors refer with the designation of “three diverse populations.”

The clinical manifestation of a health problem might be primarily a consequence of social, cultural, and economic forces and might have little to do with “genomic diversity.” For instance, as noted above, living near a toxic waste site may increase one’s chances for developing cancer or asthma, or differing nutritional intake patterns may produce diabetes, obesity, asthma, or hypertension. Some social, cultural, and economic factors influence epigenetic processes and gene expression. Researchers will want to thoroughly characterize the iPS cells, including not only DNA sequence variation but epigenetic markings, expression profiling, metabolic profiling, etc. Socioeconomic variation that influences epigenetics or other biological phenomena may also be important to sample. Thus, diversity criteria might also include factors such as immigrant/nonimmigrant status, wealth level, educational level, and other social factors known to influence health outcomes. Those nongenetic criteria help take the focus off of race as if it were primarily a biological variable and sharply reduce attendant concerns about sampling for genetic diversity.

CONCLUSION

We return to the difficult and vexing question with which we began but for which some answers are now available. Of all the myriad ways in which humans differ, why are some categories chosen (and others ignored) in order to map human diversity for the purposes of population-specific treatment regimens, pharmaceuticals, vaccines, and even patterns of migration across the globe? The most compelling answer begins with an acknowledgement of the social aspects of the phenotype. Caste is not a biological category, it is a social category. However, when human molecular geneticists sort “populations” by these social categories and find (inevitably) different patterns of the frequency of genetic markers in those very social categories, the larger social context of those findings are arrestingly seductive as a framework for explaining differential life chances outcomes. This process constitutes the “molecularization” (Rose 2001, 2007; Fullwiley 2008) and the “geneticization” (Lippman 1991) of explanations of complex social forces.

The social, economic, and legal consequences evolve into several dimensions and have now reverberated into how patents to biotech companies are granted. Jonathan Kahn has documented how this process has unfolded at the United States Patent and Trademark Office:

... the practice of requiring patent applicants to introduce race into their biotech patents has become routinized at the USPTO. (Kahn 2011, p. 402)

So what begins with the assumption that researchers are pursuing neutral, apolitical science when they deploy folk categories of caste, race, or ethnicity, can seamlessly segue into reified practices that deliver targeted pharmaceuticals to racialized target populations (and we can assume, “caste-ized” target populations, following the Indian Genome Project) under the banner of personalized medicine. This train has already left the station (Tayo et al. 2011), and all that is left to determine is which track it will be on. A closer monitoring of the hidden assumptions will at least avoid some unfortunate collisions with the social realities (of the social diversity) of human populations.

REFERENCES

- Benjamin R. 2009. A lab of their own: Genomic sovereignty as postcolonial science policy. *Policy Soc* **28**: 341–355.
- Bolnick D. 2008. Individual ancestry inference and the reification of race as a biological phenomenon. In *Revisiting race in a genomic age* (ed. Koenig BA, Lee SS-J, Richardson SS), pp. 70–85. Rutgers University Press, New Brunswick, NJ.
- Bullard RD. 2000. *Dumping in Dixie: Race, class, and environmental quality*. Westview Press, Boulder, CO.
- Cooper RS, Wolf-Maier K, Luke A, Adeyemo A, Banegas JR, Forrester TE, Giampaoli S, Joffres M, Kastarinen M, Primatesta P, et al. 2005. An international comparison study of blood pressure in populations of European vs. African descent. *BioMed Central* **3**: 1–8.
- Deshi AS, Singh S. 1989. Education, labor market distortions and relative earnings of different religion-caste categories in India. *Can J Dev Studies* **10**: 75–89.
- Dirks NB. 2001. *Castes of mind: Colonialism and the making of modern India*. Princeton University Press, Oxford.
- Duster T. 2005. Race and reification in science. *Science* **307**: 1050–1051.
- Duster T. 2006. The molecular reinscription of race. *Patterns Prejudice* **40**: 427–441.
- Epstein S. 2007. *Inclusion: The politics of difference in medical research*. University of Chicago Press, Chicago.
- Fujimura JH, Rajagopalan R. 2011. Different differences: The use of ‘genetic ancestry’ versus race in biomedical human genetic research. *Soc Stud Sci* **41**: 5–30.
- Fullwiley D. 2007. The molecularization of race: Institutionalizing human difference in pharmacogenetics practice. *Sci Cult* **16**: 1–30.
- Fullwiley D. 2008. The biological construction of race: “Admixture” technology and the new genetic medicine. *Soc Stud Sci* **38**: 695–735.
- Galanter M. 1984. *Competing equalities: Law and the backward classes of India*. University of California Press, Berkeley, CA.
- Haslanger S. 2008. A social constructionist analysis of race. In *Revisiting race in a genomic age* (ed. Koenig BA, Lee SS, Richardson SS), pp. 56–69. Rutgers University Press, New Brunswick, NJ.
- Hayes TB. 2010. Diversifying the biological sciences: Past efforts and future challenges. *Mol Biol Cell* **21**: 3767–3769.
- Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Balinger DG, Frazer KA, Cox DR. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**: 1072–1079.
- Hinterberger A. 2010. The genomics of difference and the politics of race in Canada. In *What’s the use of race: Modern governance and the biology of difference* (ed. Whitmarsh I, Jones DS). MIT Press, Cambridge, MA.
- Hinterberger A. 2012. Categorization, census, and multiculturalism: Molecular politics and the material of nations. In *Genetics and the unsettled past: The collision of DNA, race, and history* (ed. Wailoo K, Nelson A, Lee C). Rutgers University Press, New Brunswick, NJ.
- Jones JH. 1981. *Bad blood: The Tuskegee syphilis experiment*. Free Press, New York.
- Kahn J. 2011. Mandating race: How the USPTO is forcing race into biotech patents. *Nature Biotechnology* **29**: 401–403.
- Kahn J. 2012. *Race in a bottle: The story of BiDiL and racialized medicine in a post-genomic age*. Columbia University Press, New York.
- Kohli-Laven N. 2008. Hidden history: Race and ethics at the peripheries of medical genetic research. *GeneWatch* **20**: 5–7.
- Kosambi DD. 2002. *An introduction to the study of Indian history*. Popular Prakashan, Mumbai, India.
- Krieger N. 2011. *Epidemiology and the people’s health*. Oxford University Press, New York.
- Kroeger B. 2004. When a dissertation makes a difference. *The New York Times*, March 20.
- Lakshmanasamy T, Madheswaran S. 1995. Discrimination by community: Evidence from Indian scientific and technical labor market. *Indian J Soc Sci* **8**: 59–77.
- Lippman AJ. 1991. Prenatal genetic testing and screening: Constructing needs and reinforcing inequalities. *Am J Law Med* **17**: 15–50.
- Mok TA, Yi-hong W, Thongprasert S, Chih-Chih Y. 2009. Gefitinib or Carboplatin–Paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**: 947–957.
- Mudur GS. 2008. Stamp on Tagore’s India genetic map blurs lines. *The Telegraph, Calcutta, India*, April 25.
- Pager D. 2007. *Marked: Race, crime, and finding work in an era of mass incarceration*. University of Chicago Press, Chicago.
- Raelson JW, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P, Bradley WEC, Croteau P, Nguyen-Huu Q, Segal J, et al. 2007. Genome-wide association study for Crohn’s disease in the Quebec founder population identifies multiple validated disease loci. *Proc Natl Acad Sci* **104**: 14747–14752.
- Reich DK, Thangaraj N, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* **461**: 489–495.
- Risch N, Burchard E, Ziv E, Tang H. 2002. Categorizations of humans in biological research: Genes, race and disease. *Genome Biol* **3**: comment2007.1–comment2007.12.
- Reverby S. 2009. *Examining Tuskegee: The infamous syphilis experiment and its legacy*. University of North Carolina Press, Chapel Hill, NC.
- Roberts D. 2011. *Fatal invention: How science, politics, and big business re-create race in the twenty-first century*. The New Press, New York.
- Rose N. 2007. *The politics of life itself: Biomedicine, power and subjectivity in the twenty-first century*. Princeton University Press, Princeton, NJ.
- Rose N. 2001. The politics of life itself. *Theory, Cult Soc* **18**: 1–30.
- Rothenberg K. 1997. Breast cancer, the genetic “quick fix” and the Jewish community: Ethical, legal and social challenges. *Health Matrix* **7**: 97–124.
- Schütz A, 1962. Common sense and scientific understandings. In *Collected works of Alfred Schütz*, Vol. 1. Martinus Nijhoff, The Hague, Netherlands.
- Secko D. 2008. Rare history, common disease. *The Scientist* **22** (July): 38.



T. Duster

- Séguin B, Hardy B, Singer PA, Daar AS. 2008. Genomics, public health and developing countries: The case of the Mexican National Institute of Genomic Medicine (INMEGEN). *Nat Rev Genet* **9**: S5–S9.
- Sze J. 2007. *Noxious New York: The racial politics of urban health and environmental justice*. MIT Press, Cambridge, MA.
- Tayo BO, Tell M, Tong L, Quin H, Khitrov G, Zhang W, Song Q, Gottesman O, Zhu X, Pereira Ac, et al. 2011. Genetic background of patients from a university medical center in Manhattan: Implications for personalized medicine. *PLoS ONE* **6**: e19166.
- Wailoo K. 2011. *How cancer crossed the color line*. Oxford University Press, New York.
- Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Xhang J, et al. 2008. The diploid sequence of an Asian individual. *Nature* **56**: 60–65.
- Weiss KM, Long JC. 2009. Non-Darwinian estimation: My ancestors, my genes' ancestors. *Genome Res* **19**: 703–710.