Clines Without Classes: How to Make Sense of Human Variation

Joan H. Fujimura, Deborah A. Bolnick, Ramya Rajagopalan, Jay S. Kaufman, Richard C. Lewontin, Troy Duster, Pilar Ossorio and Jonathan Marks

Sociological Theory 2014 32: 208
DOI: 10.1177/0735275114551611

Article

AMERICAN SOCIOLOGICAL ASSOCIATION

Sociological Theory
2014, Vol. 32(3) 208–227

DOI: 10.1177/0735275114551611
stx.sagepub.com
SAGE

Clines Without Classes: How to Make Sense of Human Variation

Joan H. Fujimura[1], Deborah A. Bolnick[2], Ramya Rajagopalan[1], Jay S. Kaufman[3], Richard C. Lewontin[4], Troy Duster[5], Pilar Ossorio[1], and Jonathan Marks[6]

Abstract

This article examines Shiao, Bode, Beyer, and Selvig's (2012) arguments in their article "The Genomic Challenge to the Social Construction of Race" and finds that their claims are based on fundamentally flawed interpretations of current genetic research. We discuss current genomic and genetic knowledge about human biological variation to demonstrate why and how Shiao et al.'s recommendations for future sociological studies and social policy, based on their inadequate understanding of genomic methods and evidence, are similarly flawed and will lead sociology astray.

Keywords

race, social construction of race, human genetic variation, theory, genetics

In their 2012 article, "The Genomic Challenge to the Social Construction of Race," Shiao et al. (2012:73) claim that research on human genetic variation indicates the existence of "clinal classes" of humans that "provide a biological basis for [U.S.] racial and ethnic categories." They "argue that the recent research in genetics demonstrates that certain racial, and also ethnic, categories have a biological basis in statistically discernible clusters of alleles rather than in the traditional notions of human races" (pp. 68-69). Referring to the work of some geneticists, they note that "membership in these statistically identified genetic clusters appears to be concordant with at least one individual characteristic: racial/ethnic self-identification." They then offer "a reformulation of racial constructionism that accepts that

[1]University of Wisconsin-Madison, Madison, WI, USA
[2]University of Texas at Austin, Austin, TX, USA
[3]McGill University, Montreal, QC, Canada
[4]Harvard University, Cambridge, MA, USA
[5]University of California, Berkeley, CA, USA
[6]University of North Carolina-Charlotte, Charlotte, NC, USA

Corresponding Author:
Joan H. Fujimura, Department of Sociology, University of Wisconsin-Madison, 8128 Social Science Building, 1180 Observatory Drive, Madison, WI 53706, USA.
Email: fujimura@ssc.wisc.edu

recent genetic research has identified a biological basis for race/ethnicity that exceeds the more realistic threshold of statistically identifiable clusters" (p. 71).

Shiao et al. (2012) criticize social scientists for ignoring new genomic data when they posit that human racial groups are socially constructed. New statistical analyses of genomic data based on collections of human DNA, Shiao et al. argue, indicate that human biological variation is patterned into clusters that mirror the sociocultural groupings commonly designated as races and ethnicities. That is, they hold that "races" are indeed biological categories, albeit with blurred edges. Since the authors surmise that social constructivists are uncomfortable with any framing of races as biological categories, they propose that the "biological race" concept be replaced by the concept of "clinal classes." However, there is so much overlap between the two concepts that Shiao et al. are really proposing a replacement of terms rather than a fundamentally new concept. Shiao et al. also propose that these biological races or clinal classes are reflected in "the social perception of ancestry," which they define as how people classify themselves and others based on physical characteristics. The authors thus argue that while social perceptions of race may be socially constructed, such perceptions nevertheless coincide with biological races, as defined by what they think are genetic markers of biological ancestry.

Despite their claims to the contrary, Shiao et al. (2012) are using recent human genetic research to reintroduce the age-old construct of biological races, just renamed as "clinal classes." Shiao et al.'s interpretation of genetic clustering research opens the possibility that readers will misunderstand the clustering technologies and the genetic evidence.[1]

In this article, we show that the genetic research used by Shiao et al. (2012) to construct their clinal classes does not constitute evidence for the existence of genetically distinct racial and ethnic groups. We demonstrate that commonly used U.S. racial and ethnic categories do *not* represent what they call clinal classes and are *not* equivalent to them. Nor are the genetic clusters identified by the *Structure* or principal components analysis (PCA) computer software programs homologous to racial or ethnic categories, as Shiao et al. claim. We also show that sets of biological features usually thought to delineate racial groups—such as skin color, hair texture, and facial features—do not covary. Instead, they vary across space, irrespective of racial group.

Further, Shiao et al. (2012) misrepresent the complexity of concepts and categories of self-identified race and ethnicity and their complex histories of production over historical time. A wealth of sociocultural and historical evidence has demonstrated that human race and ethnic categories are socio-historically constructed, changing over time and differing by locale (e.g., Boas 1913; Du Bois 1940; Duster 2003; Hacking 2005; Haraway 1997; Stepan 1986). Sociologists have demonstrated that racial and ethnic categories and identities are relational, processual, dynamic, and disaggregated (Blumer 1958). They have studied racialized and ethnic formations as cultural idioms, cognitive schemas, discursive frames, contingent practices, or political projects (e.g., Brubaker, Loveman, and Stamatov 2004; Davis 2001; López 1996; Omi and Winant 1994). Studies have examined how race classification schemes differ between countries or cultures and even within a particular nation over time (e.g., Kim 2008; Loveman and Muniz 2008; Telles 2004). Others have examined the many inconsistencies between official and practical interpretations of race classifications—in apartheid South Africa, for example (Bowker and Star 1999). Despite occasional attempts in the late twentieth century to make the case for "innate biological" racial differences in, for example, intelligence, the consensus of social and historical scholarship is that racial and ethnic categories are formed through—and should be studied as—political, social, cultural, and psychological processes. Many scholars have examined how such socially constructed racial beliefs and attitudes have influenced studies proposing a biological basis to race

(Benjamin 2009; Cho and Sankar 2004; Duster 2006; Fausto-Sterling 2008; Goodman 2006; Morning 2009). More recent studies have examined the "one-drop" rule in the United States that played the primary role in structuring "African American" or "Black" racial membership in legal and political arenas (e.g., Washington 2011).[2] The suggestion that all this sociohistorical variation has a biological basis makes little sense.

CLINAL GENETIC VARIATION IN HUMANS

In 1938, evolutionary biologist Julian Huxley coined the term *cline* to describe the pattern of biological variation in some species. A *cline* is "a gradation in measurable characters."[3] Because traits change gradually and continuously across geographic space in a cline, one cannot objectively divide a cline into discrete categories; the placement of any dividing line will be arbitrary.

Many biologists writing in the eighteenth, nineteenth, and twentieth centuries—and even a few in the twenty-first century—believed that biology could be used to classify humans into different classes or subspecies. Arguing against this position, biological anthropologist Frank Livingstone (1962) proposed that the genetic continuity of humans occurred in the form of "clinal variation." He argued that human genetic variation was the result of constant movements and migrations among people, that there were no major reproductive barriers between peoples, and therefore that genetic variation was continuous across large geographic areas. As a result of these processes, Livingstone noted that each character or feature varied "nonconcordantly," that is, that characters or features did not vary together for each individual. This, he argued, was evidence that humans could not be divided into distinct "classes" or "subspecies" or "races."

Over the past 50 years, anthropologists and human geneticists have produced more evidence that most variation in the human species is clinal. Both gene frequencies and phenotypic traits, such as skin color or cranial morphology, vary clinally, with few sharp discontinuities (Barbujani and Colonna 2010; Batai and Kittles 2013; Marks 2010; Relethford 2009; Serre and Pääbo 2004). Furthermore, both genetic and phenotypic traits vary *nonconcordantly*, with different traits exhibiting distinct patterns of variation across geographic space (Goodman 2000; Jorde and Wooding 2004). Skin color, for example, exhibits a continuous latitudinal gradient: individuals who live at lower latitudes (near the equator) and who are exposed to greater ultraviolet (UV) radiation intensity exhibit darker skin than individuals living at higher latitudes who experience lower UV radiation intensity (Jablonski and Chaplin 2010). In contrast, certain genetic variants show a southwest to northeast gradient in western Eurasia (Rosser et al. 2000) while others exhibit an east to west gradient across Eurasia (Xiao et al. 2004). Other traits also show distinct clinal gradients across the world. Thus, sets of biological features usually thought to delineate racial groups—such as skin color, hair texture, and facial features—do not covary. Instead, they vary independently across space, irrespective of racial group (Goodman 2000; Marks 2010).

Because of such nonconcordant clinal variation, our species is not subdivided into discrete, genetically distinct, and biologically homogeneous racial or continental groups. Instead, human populations tend to be most genetically similar to others who live nearby, irrespective of continental boundaries, and the degree of genetic similarity between populations is inversely correlated with geographic distance (Barbujani and Colonna 2010; Li et al. 2008; Ramachandran et al. 2005; Relethford 2004).

These patterns have been shaped by three key factors over the course of human evolution: (1) Migration and gene flow (genetic exchange) between populations is spatially patterned. Extensive interbreeding almost always occurs between neighboring groups, which are connected by complex social, economic, and marital relationships regardless of racial or cultural

affiliation labels (Marks 2010). Geographic distance also limits migration, so restricted gene flow has produced a pattern of "isolation by distance" among human populations and has contributed to the gradients of genetic change that we see in our species (Cavalli-Sforza, Menozzi, and Piazza 1994; Jay et al. 2013; Lawson-Handley et al. 2007; Relethford 2004). (2) Clinal variation in humans also reflects the successive long-distance migrations and founder effects that occurred as humans migrated out of Africa and dispersed throughout the world (Campbell and Tishkoff 2008; DeGiorgio, Jakobsson, and Rosenberg 2009; Ramachandran et al. 2005). (3) Clinal variation in some genetic and phenotypic traits reflects natural selection and adaptation in response to environmental gradients (Barbujani and Colonna 2010; Jablonski and Chaplin 2010; Relethford 2004).

Thus, because of the cumulative effects of these evolutionary forces, clinal variation is the predominant characteristic of human biological variation. With data on a sufficiently large number of genetic variants, though, it is possible to detect small and subtle discontinuities in some clines. There are two evolutionary reasons for this. First, sequential founder effects and random genetic changes (*genetic drift*) occurred as humans spread throughout the world, producing gene frequency differences between some groups (e.g., Asians and Native Americans). Second, some environmental and geophysical factors (e.g., mountain ranges or large bodies of water) have reduced gene flow between geographically proximate populations. It is therefore possible to detect some genetic clusters, or groups of individuals who are more genetically similar to one another than they are to other individuals.[4] Because of nonconcordant clinal variation, however, different clustering patterns emerge from studies of different genetic variants (see the following), and genetic clusters are usually comprised of people who live geographically close together. A genetic cluster may therefore contain people from the same continent (e.g., Africa) or people from different continents who live fairly close to the continental boundary (e.g., northeastern Africans and Middle Easterners, or northeastern Siberians and native Alaskans). The genetic clusters that have been detected do not always correlate with racial or ethnic categories.

CLINAL CLASSES: AN OXYMORON AND NOT A DESCRIPTION OF RACIAL/ETHNIC CATEGORIES

Having explained the pattern of human clinal variation and the forces that have shaped that pattern, we now turn to Shiao et al.'s (2012) notion of "clinal classes," explain why it is an oxymoron, and discuss three reasons why their idea is problematic.

Shiao et al. argue that the genetic clustering patterns revealed by recent genomic data demonstrate that significant biological differences do exist between socially defined racial and ethnic groups. They suggest that "roughly 60 to 150 randomly selected loci are needed to separate individuals into genetic clusters that are homologous to continentally based racial categories"[5] (Shiao et al. 2012:71), and they argue that those clusters are equivalent to the racial categories used on the U.S. census and in the contemporary United States. They coin the term *clinal classes* to describe these groupings, defining a *clinal class* as a statistically constructed cluster of individuals who share biological ancestry over the past 50,000 years, so that "clinal classes assume a common evolutionary history, possess extensive genetic similarities, and coexist with clinal variations both within each class and across classes" (Shiao et al. 2012:72).

However, by definition, a cline is not a class; therefore, a clinal class is an oxymoron. Beyond that, there are at least three key problems with their assertions:

1. What Shiao et al. refer to as the "concept of clinal classes" is not actually a new concept. It is largely identical to an older, discredited race concept that creates typologies of people and imposes those typologies upon the human species.

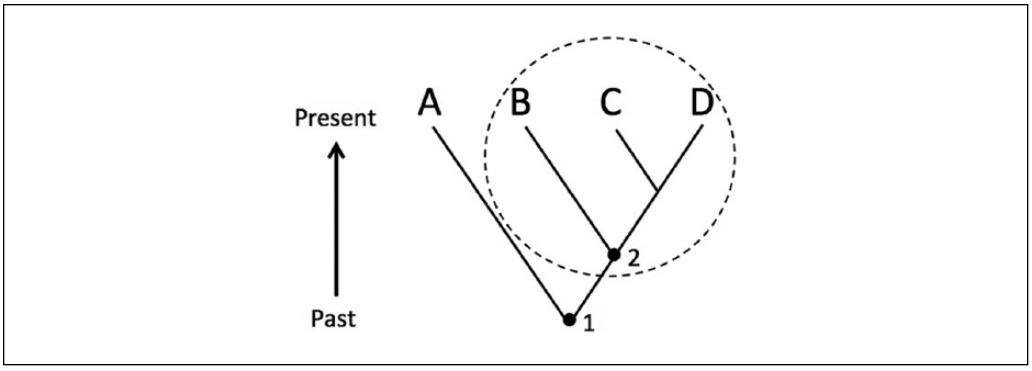

Figure 1. An example of a "phylogenetic tree" diagram showing four individuals or populations (A-D) and the ancestral relationships among them. Point 1 represents the most recent ancestor of A-B-C-D; point 2 represents the most recent common ancestor of B-C-D. Thus, B, C, and D share a more recent common ancestor (point 2) than any does with A, so B is more closely related to C and D than it is to A. Evolutionary biologists refer to B-C-D (circled by the dashed line) as a "clade" (a group containing an ancestor and all descendents). Shiao et al.'s (2012) "genetic watershed" is another term for this kind of grouping. In contrast, a group containing just A and B is not a clade or a genetic watershed. As discussed later in the text, a group containing Africans or black Americans is equivalent to an A-B group, not a B-C-D group.

2. The observed genetic clustering patterns do not actually indicate that there is a small number of genetically differentiated human groups that exhibit clear and extensive biological differences.
3. The racial and ethnic categories that we recognize in the United States today cannot be considered clinal classes because they do not meet Shiao et al.'s definitional criteria. In particular, racial/ethnic categories do not accurately represent patterns of biological ancestry and genetic relatedness.

We discuss each of these problems in turn.

CLINAL CLASSES: REINVENTING THE EVOLUTIONARY LINEAGE CONCEPT OF RACE

Shiao et al. (2012) assert that their concept of clinal classes provides a new theoretical framework, but what they describe—statistically constructed clusters of individuals who they claim share biological ancestry and a common evolutionary history and who exhibit extensive genetic similarities—is essentially the evolutionary lineage concept of race. This biological race concept has been previously discussed and critiqued by Templeton (1999, 2013), Marks (2008), and others, and it defines races as "distinct evolutionary lineage[s] within a species" (Templeton 1999:632). As Templeton noted, this definition assumes that races are "genetically differentiated due to barriers to genetic exchange that have persisted for long periods of time," leading to a situation in which "many traits should reflect the common evolutionary history of the [race]" (Templeton 1999:632). The extensive similarities between Shiao et al.'s clinal class concept and the evolutionary lineage concept of race are obvious. Furthermore, Shiao et al.'s (p. 73) concept of "genetic watersheds" is also not new. A "genetic watershed" is analogous to what evolutionary biologists call *clades* when drawing "phylogenetic tree" diagrams of the ancestral relationships between individuals or populations (Figure 1). A *clade* is defined as a group containing an ancestor and all of its

descendants (Figure 1), and Shiao et al.'s "genetic watersheds" describe essentially the same thing. Thus, Shiao et al. are really just proposing a replacement of terms, not any fundamentally new concepts.

ISSUES WITH THE PRODUCTION AND INTERPRETATION OF GENETIC CLUSTERS: *STRUCTURE*-BASED ANALYSES

The second problem with Shiao et al.'s (2012) assertions is that genetic clustering studies do not indicate that our species is divided into a small number of genetically differentiated groups. Shiao et al. cite two statistical approaches—the Bayesian clustering program *Structure* (Pritchard, Stephens, and Donnelly 2000) and principal components analysis (Paschou et al. 2007; Patterson, Price, and Reich 2006)—which they say yield "genetic clusters that are homologous to racial and ethnic categories" (Shiao et al. 2012:71-72) and "provide a biological basis for racial and ethnic categories" (p. 73). However, just because the *Structure* program sometimes yields what appear to be continentally differentiated clusters is not proof that genetically distinct races exist. To explain this point, we need to briefly describe how the *Structure* program works. (See Bolnick [2008] and Kalinowski [2011] for more in-depth analyses of the *Structure* program and its limitations.)

Structure uses multilocus genotype data to estimate the number (K) of genetic clusters in a given data set and to assign individuals probabilistically to those K clusters, assuming that (a) all genetic markers are unlinked and inherited independently of one another and (b) the clusters (populations) are in Hardy-Weinberg equilibrium (i.e., not experiencing any genetic changes from generation to generation at the population level). *Structure* produces whatever number K of clusters is requested by the user. To determine which value of K (and which set of genetic clusters) best represents the variation present in a given data set, the user needs to calculate the likelihood score for each value of K (i.e., the likelihood of observing this data set given that value of K) and select the value of K that maximizes that likelihood (Pritchard et al. 2000).[6]

As Shiao et al. (2012) note, Rosenberg et al.'s (2002) analysis of 377 genetic markers in 52 populations from around the world did identify five genetic clusters that correspond to "major geographic regions" (continental groupings) when K was set to 5. However, their data set also produced various other clusters when K was set to other numbers between 2 and 20, and comparison of the likelihood scores did *not* suggest that $K = 5$ best represented the genetic variation in this dataset (Bolnick 2008). Some other models (with different values for K and different genetic clusters) had higher likelihood scores, indicating that they better fit the data, but no single value of K clearly maximized the probability of observing the particular genotypes seen in Rosenberg et al.'s (2002) data set. In other words, just because Rosenberg et al. (2002) could identify subtle genetic discontinuities between continental groups of samples when they arbitrarily specified 5 clusters a priori does not mean that those clusters better represent or account for human genetic variation than another model specification that does not conform to continental boundaries.

Furthermore, it should be noted that more recent analyses using greater numbers of genetic markers and different types of markers have yielded different and even less clear-cut clusters. Rosenberg et al. (2005) analyzed 993 genetic markers in the same population samples and identified a slightly different set of genetic clusters when $K = 6$ (Native Americans were divided into two clusters and the Kalash of Central/South Asia did not form a separate cluster). When Jakobsson et al. (2008) examined 525,910 single nucleotide polymorphisms and 396 copy-number variants in a representative subset of these same samples, the *Structure* analysis yielded somewhat different clusters and clearly showed clines across geographic space.[7]

Finally, it should be noted that *Structure* results are highly dependent on whether the method's assumptions and underlying model (i.e., that the sampled individuals are drawn from *K* separate populations) hold for any given data set. If those assumptions are violated—as is the case when the data set has been shaped by restricted gene flow with isolation by distance (Pritchard and Wen 2004)—then the inferred clusters may be incorrect or "rather arbitrary" (Bolnick 2008; Pritchard and Wen 2004:14; Weiss and Long 2009). The sampling scheme used also matters: variation in sample size, as well as nonrandom or discontinuous sampling, is known to yield incorrect inferences about population genetic structure and can lead to the appearance of multiple discontinuous clusters when in fact none exist (Kalinowski 2011; Schwartz and McKelvey 2009; Serre and Pääbo 2004). Because human genetic variation has been shaped by restricted gene flow with isolation by distance and because discontinuous sampling has been used in studies of human populations, *Structure*-based inferences about human population genetic structure may be misleading.

Shiao et al. (2012) do acknowledge the circularity of analysis in some *Structure*-based studies, which began by predefining groups and identifying specific genetic markers with frequency differences that differentiate those groups (ancestry informative markers, or AIMs) and then used those markers to recreate the same clusters of individuals.[8] They call markers that are used to differentiate predefined groups (loaded AIMs) and place less weight on the findings of such studies. However, Shiao et al. also misinterpret *Structure* results from studies that did not use loaded AIMs (e.g., Rosenberg et al. 2002).

ISSUES WITH THE PRODUCTION AND INTERPRETATION OF GENETIC CLUSTERS: PCA-BASED ANALYSES

Shiao et al. (2012) also argue that principal components analysis of human genetic data divides our species into genetically differentiated groups, similar to the clusters produced by *Structure*, and that PCA therefore provides evidence for their clinal classes. They maintain that since studies that use PCA do not use predefined groups, the resulting markers are not loaded and the methods are unbiased. Shiao et al. then interpret the clusters produced using PCA technologies as demonstrating that racial groups are biological groups or clinal classes.

However, Shiao et al. (2012) are not correct. Some researchers who use newer PCA technologies like Eigenstrat (Patterson et al. 2006) to cluster individuals based on their genetic variation still use predefined populations, usually groups designated by ethnicity or nation, to help inform their analyses (Fujimura and Rajagopalan 2011).[9] In some such studies, researchers collect DNA from different predefined racial/ethnic/national groups based on self-identification and then use PCA software to make two-dimensional plots based on a subset of the DNA collected from these individuals (Patterson et al. 2006; Price et al. 2008; Yamaguchi- Kabata et al. 2008). If the Eigenstrat plots show clusters that correspond to how individuals self-identified at the time of DNA collection, then they identify the single nucleotide polymorphisms (SNP)[10] markers that they consider to contribute most to the top two dimensions and call them AIMs.

For another example, some of the researchers who use PCA technologies use designations of predefined nationality, ethnicity, or population labels attached to the samples to decide which samples to include in the analysis in the first place. In some cases, when looking for patterns of variation that differentiate many groups, they will combine all samples from many groups before analysis. In others, when looking for specific differences between only two or three groups, they will include only the samples from those groups in the statistical analysis. The choice of which samples from which groups to include can change the outcome since PCA is designed to identify the differences that separate only those samples that

were entered in the program at the outset. When samples from a particular group are entered, the form that cluster takes on the plot will depend on which other samples from which other groups are also simultaneously entered. This is a different use of predefined populations than using the program to find differences between predefined populations.

Our point is that while PCA and *Structure* programs cluster samples "blindly" to group labels, researchers' decisions about which samples to compare together, and their interpretations of the results, are often informed by these labels. Previously determined information about individuals' nationality, ethnicity, or linguistic group may be used to decide which groups to compare, how to draw the clusters when their boundaries are non-obvious, which configurations of the clusters to pay attention to (especially in *Structure*), and which SNPs should be called AIMs. There is necessarily a back and forth between the "inside" of the data, graphs, and plots and the "outside" of the ways the samples are collected. This sometimes leads researchers to infer that observed genetic differences indicate different group memberships (where groups are denoted by national, regional, racial, or ethnic designations), even when that is not the only or most likely explanation for the observed differences.

Another key point about the process of constructing genetic clusters is that human geneticists make decisions about which subset of individuals to use to "represent" a "race" or "national group" in their sampling procedures and in their cluster analysis. The subsets they use are obviously extremely small compared to the number of individuals who identify with that race or nationality label. They thus extrapolate their results from a small number of individuals to make inferences about a vastly larger number of individuals who self-identify with the same race or nationality label and whose genetics have not been studied. For these reasons, SNP clusters are not solely the product of the DNA data themselves.

Human geneticists make many decisions about whom to include and how to include them in their analyses. If sociologists want to use the work of geneticists, it is *crucial* that we understand how the genetic technologies and research outcomes are constructed and do not oversimplify the results. Unfortunately, Shiao et al. (2012) do just that, as we explain in the following section.

RACIAL/ETHNIC CATEGORIES ARE NOT CLINAL CLASSES

Perhaps the most problematic of Shiao et al.'s (2012) assertions is the suggestion that socially recognized racial and ethnic categories are homologous to clinal classes. Given Shiao et al.'s definition of clinal class, commonly used racial and ethnic categories do not represent clinal classes and are not equivalent to them. Nor are the genetic clusters produced by *Structure* or PCA homologous to racial or ethnic categories. This is because the members of each racial and ethnic group are quite biologically diverse (Barbujani, Ghirotto, and Tassi 2013). Not all individuals who identify as members of a racial/ethnic group share a common evolutionary history or possess extensive genetic similarities due to shared biological ancestry.

This disconnect between our socially recognized racial categories and the notion of clinal classes (or between racial categories and *Structure* or PCA clusters) occurs because our socially produced racial categories are not based on the evolutionary history of our species. As Templeton (1999:635) noted, "when a biological race is defined as a distinct evolutionary lineage within a species, the question of race can only be answered in the context of the recent evolutionary history of the species." Thus, it is important to know that our species evolved in Africa before dispersing throughout the rest of the world (Campbell and Tishkoff 2008), so African (and recently African-derived) populations have had more time to accumulate genetic diversity through mutation. When humans dispersed out of (east) Africa, the

individuals who left Africa carried only a subset of the genetic variants found in the ancestral African population. Because of this history, the most divergent genetic lineages are found within Africa, and non-Africans exhibit significantly less genetic diversity than African (and African Diaspora) populations (Campbell and Tishkoff 2008; Long, Li, and Healy 2009; Tishkoff et al. 2009). Some Africans (especially in East Africa) also share more genetic similarities and evolutionary history with non-Africans than with other Africans. For example, when Kalinowski (2011) examined the genetic data collected by Rosenberg et al. (2005), he found that some African populations (the Kenyan Bantu, Mandenka, and Yoruba) share more genetic variants with Europeans than with other African populations, such as the San or Mbuti.

Similarly, when we look at the distribution of mitochondrial DNA (mtDNA) lineages throughout the world, we see that the oldest and most divergent lineages, named L0-L6, are all found in Africa, whereas the only lineage found outside of Africa (until very recently) is L3—the most recently evolved of the L lineages (Underhill and Kivisild 2007). Thus, both Africans and non-Africans belong to the L3 lineage (or clade), so their mtDNAs are more similar and they share more recent common ancestry (and more evolutionary history) with one another than with non-L3 Africans. In other words, the L3 clade (or, in Shiao et al.'s term, "genetic watershed") contains some but not all Africans, just as the other L clades ("genetic watersheds") contain some but not all Africans. Because, at least in the United States, we generally lump all African and African Diaspora peoples into the same racial group, racial categories are not clades, clinal classes, or genetic watersheds.

Thus, while racial/ethnic categorization may partly reflect social perceptions of biological ancestry, as Shiao et al. (2012) suggest, it is critical to note that these perceptions do not reflect *actual* patterns of biological ancestry and relatedness. Instead, racial and ethnic categories provide poor representations of ancestry and relatedness. African Americans, for example, are extremely variable in how much West African ancestry they have (1 percent to 100 percent, as estimated genetically by Bryc et al. [2010]), and some black Americans actually share more biological ancestry with white Americans than with other black Americans (Bryc et al. 2010). Social rules like the "one-drop" rule, not biological ancestry, play the primary role in structuring racial membership (Washington 2011).

Finally, it is important to note that contrary to what Shiao et al. (2012) claim, most people do not actually "'know,' that is, self-identify with, their genetic ancestry, whether their ancestors lived in sub-Saharan Africa, western Eurasia, the Pacific Islands, eastern Eurasia, or the Americas 50,000 to 2,000 years ago" (Shiao et al. 2012:79). Humans have moved around a lot over the past 50,000 years. The ancestors of Native Americans lived in Africa or Eurasia before about 20,000 years ago (O'Rourke and Raff 2010), but Native Americans do not identify as African or Eurasian. Furthermore, as the recent publication of a 24,000-year-old Siberian genome sequence shows (Raghavan et al. 2014), the current distribution of genetic lineages does not provide an accurate or complete record of where those lineages were found thousands of years ago. Genetic variants that we characterize as "European" today clearly had a different or wider geographic distribution in the past (Raghavan et al. 2014), suggesting that we do not actually know where our ancestors lived 50,000 years ago or where genetic lineages were found then.

SHIAO ET AL.'S PRESENTATION OF FICTITIOUS DATA

In addition to the criticisms discussed previously, we also want to raise a concern about Shiao et al.'s (2012) graph on page 71 of their paper. The authors have reprinted a graph from philosopher Neven Sesardic's (2010) article. Sesardic constructed these points on the

plot to emphasize his prediction that one could in theory cleanly separate groups if one analyzes correlations between many genetic markers. Sesardic used these fictitious points on the graph to argue that races can be considered biological categories. We emphasize that this graph represents an idealized case and is not based on any real data. Sesardic does not define the axes of the plot, and Shiao et al. do not make it clear that this is a fictitious example. This is important to point out because real data may not look like this, as we discuss in the following section.

We also note that it is not possible to divide this scatterplot of points into the two clusters using *only* the genetic information represented by the points. In his idealized example, Sesardic (2010) had to first designate the squares as different from the triangles. That is, a priori information was required to differentiate the points. *This initial differentiation of the points lies outside of and prior to any assessment based on the data presented in this plot.* One must have an independent criterion for designating squares as squares and triangles as triangles in the first place. Without such a criterion, all points would have been drawn using the same symbol, and no divisions would have been evident. As Bolnick (2008) and others have shown, while subtle discontinuities can be detected in some clines, the exact placement and interpretation of dividing lines in clinal data is arbitrary or influenced by external preconceived ideas about the boundaries between groups. Furthermore, putting the dividing line between the squares and triangles, as Sesardic did, is not necessarily more biologically meaningful that putting it elsewhere, at least not based just on the information contained in the points presented in the graph. Making a division between the points represented by squares and the points represented by triangles is only meaningful and useful if we have a third variable defining squares and triangles as different; that is, only if we have a priori information that squares and triangles are distinct. One cannot define clusters based just on the presented data points alone.

The graph is also misleading about the kind of information such genetic analyses can provide. Their graph implies that the clusters that are discerned in such plots are unequivocal and uncontroversially defined, whereas in practice there are always debates and negotiations among geneticists about how to cluster, whom to include, and whom to exclude (Fujimura and Rajagopalan 2011). Shiao et al. (2012:79) argue that "genetic clusters" and "clinal classes" provide a precise "measurement of ancestry" because they are not socially constructed. In fact, though, these "genetic clusters" are just as socially constructed.

THE LEWONTIN APPROACH: PROPORTION OF VARIATION

Another issue we wish to discuss is Shiao et al.'s (2012) attempt to discredit population geneticist Richard Lewontin's (1972) findings, which have been used by social scientists and biologists to argue against human races as biological categories. Lewontin found that for the ethnic groups he studied, 85 percent of the total genetic variation was due to individual differences within populations and only 15 percent to differences between populations. In his 2008 paper with human geneticist Marcus Feldman, Lewontin reconfirmed his earlier finding, citing the papers by Rosenberg et al. (2002, 2005) that used microsatellite DNA markers to show the same general patterns.

At first, Shiao et al. (2012:69) appear to agree with Lewontin's findings. However, later in the article (p. 70), they present A. W. F. Edwards's (2003) assertion that Lewontin's conclusion is a "fallacy" because he relied on the "average degree of variation of individual genes instead of also considering their correlations" (Shiao et al. 2012). Shiao et al. claim that new approaches examining multiple markers and their correlations show that one can classify individuals into groups that mirror race categories. They base their claim on

Rosenberg et al.'s (2005) article showing that one can develop clusters based on associations between genetic markers of individuals and geographic regions of origin.

However, Lewontin's finding is not a fallacy, because Shiao et al. (2012) and Lewontin are discussing entirely different questions. Feldman and Lewontin (2008) agree with Rosenberg et al.'s (2002, 2005) findings using multiple genetic markers and their correlations to identify groups that correspond to geographic region of origin. The production of clusters is distinct from questions about what proportion of genetic variation is found within or between groups. First, we stress that most of these markers are not in protein encoding genes, so it is unclear whether they are important for any phenotypic outcomes. Second, the geographic patterns that Rosenberg et al. identified stem from small differences in the frequencies of many genetic markers. That is, many genetic markers are present in all of the major geographic clusters that Rosenberg et al. (2002, 2005) constructed using *Structure*, but the *frequencies* of those markers vary, with some being more common in some populations than others. "There does not exist any gene for which one major geographical cluster includes 100 percent of one genotype while another major geographical cluster has 100 percent of another genotype" (Feldman and Lewontin 2008:93).

The key point here is that Lewontin argues that the geographic clustering approach is not appropriate for answering the question of whether group differences are meaningful for anything *other than* geographic ancestries. He considers this approach to be a taxonomy that can be used with some success to hypothesize associations between individuals and regions of geographic origin. For Lewontin, if one wants to measure the *meaningfulness* of differences between groups with different geographic ancestries, one needs to use a "proportion of variation" approach. This approach compares genetic variation among individuals within groups to the genetic variation between these groups. Ironically, Edwards's (2003:799) reported findings confirm Lewontin's (1972) findings: Edwards's results on within (84 percent) and between (16 percent) genetic variation are almost identical to those of Lewontin's (1972) results (85 percent within, 15 percent between populations).

WHAT DO DIFFERENCES BETWEEN GENETIC CLUSTERS MEAN?

Shiao et al. (2012) confuse the ability to *construct* genetic clusters with the importance or meaning of those clusters (i.e., their ability to explain traits of interest or to represent major or biologically significant patterns). PCA methods identify markers that seem to differ among the individuals included in the analysis. Some researchers cited by Shiao et al. (e.g., Risch et al. 2002) highlight these differences and bring them to the fore. But what makes these points in the DNA any more important than the hundreds of thousands of others in human genomes? Shiao et al. believe they can see race or ethnic groups mirrored in the clusters produced by small allele frequency difference patterns aggregated over many loci. But these small allele frequency differences have *not* been found to be associated with features that are coded as racial or ethnic.

Is there a relationship between genetic clusters, the genetic markers used by some researchers to differentiate these clusters, and phenotypic features like skin color? Shiao et al. (2012) leave unaddressed the question of how the studied genetic markers are related to the phenotypic features we code as race. They *imply* a relationship as if it is common sense and clear-cut. But, in fact, biology is much more complicated. For example, pigmentation phenotypes are complex, and there is no single, simple "gene for skin color" or other body pigmentation. Geneticists have so far identified dozens of genes that, along with environmental stimuli, contribute to pigmentation (Sturm and Duffy 2012). However, there are likely to be hundreds of genes and environmental stimuli that influence skin color or any

racially coded phenotype. The relationship between genotype and phenotype is multifaceted, and phenotype is a product of many interactions over time and place, of biology, of DNA and other molecules that regulate gene expression, of geography and environment, of histories of people mixing in complicated ways. Assessing the DNA *alone* is too simplistic an approach to find the complex web of factors that together produce many of our physical characteristics.

Thus, the proposed "clinal class" concept does not take into account the complexity of biology. It does not capture and convey the heterogeneous webs of relationships in which DNA sequence is mediated by biological, environmental, and social factors and that DNA sequence on its own plays just one part in producing the final phenotypes we see, particularly for most complex traits including those coded as "racial." Biologists are still trying to understand the various mechanisms that help to control how and when DNA is expressed (Fujimura 2005; Krieger 2012; Oyama 2000).

CONCLUSION

In a latter day version of eighteenth-century attempts to biologically classify humans into racial categories, Shiao et al. (2012) use their interpretation of twenty-first–century genetic research to make claims about the existence of "clinal classes" that mirror racial categories as commonly understood in the United States. Their claim is in direct contrast to social scientific and historical studies that show that U.S. race categories have changed over time and place and in deep dialogue with political, cultural, and institutional movements. Their claim also does not accord with the genetic data they cite.

The most troubling of Shiao et al.'s (2012:83) assertions is their claim that "recent genetics research has identified a biological basis for race and ethnicity" and that clusters of human genetic variation, which they call "clinal classes," are "homologous to [U.S.] racial and ethnic classifications." Both of these claims are unfounded. Our article demonstrates that commonly used U.S. racial and ethnic categories do not represent what they call clinal classes and are not equivalent to them. Nor are the clusters identified by *Structure* or PCA homologous to racial or ethnic categories.

Shiao et al.'s (2012) "clinal classes," like earlier claims that race categories are delineated by biology or genetics, fail to accurately represent the patterns of human genetic variation or the makeup of human racial groups. This is in part because they misinterpret the *Structure* and PCA studies on which they rely. Because of the cumulative effects of evolutionary forces, clinal variation is the predominant characteristic of human biological variation, and the small discontinuities in DNA marker frequencies that some studies have reported do not represent major biological differences between groups of people.

While racial/ethnic categorization may partly reflect social perceptions of biological ancestry, as Shiao et al. (2012) suggest, it is critical to remember that these perceptions do not reflect *actual* patterns of biological ancestry and relatedness. Even if we ignore the fact that people change their self-designated racial labels by situations, the members of each racial and ethnic group are quite biologically diverse, and not all members of a racial/ethnic group share a common evolutionary history or extensive genetic similarities. Because social rules, not biological ancestry, have played the primary role in structuring racial membership in the United States, racial and ethnic categories provide poor representations of ancestry and relatedness. In analyses of genetic variation, many individuals who self-identify as a particular race do not fit within the simplistic representation of geographical areas that Shiao et al. draw.

Our article primarily addressed Shiao et al.'s (2012) claim that group genetic differences mirror racial groups because all of their other arguments rest on the existence of this

relationship. But it is also critical to point out that while one may be able to construct or identify clusters using various technologies, those clusters do not say anything *meaningful* about the biology of the groups. Shiao et al. confuse the ability to *construct* difference with its importance or meaning, that is, its ability to explain traits of interest or to represent major or biologically significant patterns. The ability to construct clusters is not significant in the context of the traits and behaviors of interest to sociologists. There is no evidence that the studied markers or observed clusters have any bearing on these traits or behaviors.

Unfortunately, Shiao et al. (2012) argue exactly the opposite: They suggest that "group biological differences" have meaning for the conduct of future sociological research on behavior. They hope to inspire others to conduct research relating genetic frequency differences by clinal class or race *to* social behaviors because they think there may be inherited differences in these behaviors that assort by race and have a genetic basis.

But sociologists and social epidemiologists have documented for decades that intergenerational effects in human behaviors have primarily been due to environmental and socioeconomic factors that are experienced in childhood, adolescence, and adulthood (reviewed in Adler and Rehkopf 2008; Brulle and Pellow 2006; Kahn, Wilson, and Wise 2005; Power, Kuh, and Morton 2013; Walsemann, Geronimus, and Gee 2008; Willson, Shuey, and Elder 2007). These explanations incorporate experiences of bias, prejudice, and discrimination that individuals experience based on their socially perceived race, and they account for most observed behavioral variation. In contrast, existing genetic research suggests that it will be exceedingly difficult to identify any specific genetic factors that have a large effect on behavioral traits (Beckwith and Morris-Singer 2012; Chaufan and Joseph 2013; Joseph 2011; Turkheimer 2011).

Shiao et al.'s (2012) arguments suggest—with no robust evidence to back up the claim—that physical traits like skin, hair, and eye color may correlate with abilities, talents, behaviors, and political leanings. They also speculate that these unsubstantiated correlations have a genetic basis, although Shiao et al. (p. 77) themselves acknowledge the lack of evidence supporting the view that clinal classes/race groups differ in their individual outcomes because of inherited genetic differences, including outcomes of interest to social scientists. What, then, is the use of clinal classes? Even the sociological research inquiries into race or ethnicity that Shiao et al. propose in the second half of their paper neither require nor would benefit from genetic or biological explanations of race. In our view, Shiao et al.'s misunderstandings of genetics ultimately pose an obstacle to studying how discrimination, racism, prejudice, and bias produce and reinforce socioeconomic inequalities and other disadvantages for racially marked individuals in society.

For decades, social scientists have been studying perceptions of racial and ethnic physical features (e.g., skin color) and their effects on attitudes of racism, prejudice, and discrimination. That these features may have some biological inputs is completely tangential to the more pressing questions of why these particular features continue to be mobilized as salient dividers in American politics and culture.

We close with a question and argument about the role of science in adjudicating questions like "is there a biological basis for social categories of race." Some sociologists have argued that sociologists can and should study the social construction of race and not attend to claims about the role of biology or genetics in that constitution. We fully support the first part of this statement, but we disagree with the second part. The publication of Shiao et al.'s (2012) article is testament to the fact that genetic research is being used by sociologists to argue for the role of genetics in the constitution of race categories and the role of genetics in shaping social behavior. Given this re-introduction of genetic arguments for differences in behaviors by race, sociologists cannot ignore genetic arguments about racial differences. To address

those arguments, sociologists need to learn to evaluate claims based on genetic data, and it is critical that social scientists and geneticists work together to counter misinformed claims about genetic racial categories and misinformed claims about the role of racialized genetics in explaining human behaviors.

ACKNOWLEDGEMENTS

Fujimura thanks Kjell Doksum for his comments of many drafts of this article. Bolnick thanks Jennifer Raff and Rick Smith for comments. Rajagopalan acknowledges the Sociology Department at University of Wisconsin-Madison and the Life Science Foundation. Kaufman thanks the Canada Research Chairs program. Ossorio thanks the Morgridge Institute for Research. Marks acknowledges the Templeton Foundation and the Notre Dame Institute for Advanced Study.

FUNDING

Fujimura received financial support from the NIH grant 5 R03 HG006571-02.

NOTES

1. Relatedly, Pamela Sankar (2008) called for social scientists to study the social construction of race in contemporary statistical genetics research, which is more nuanced than typological race discussions of previous centuries. Some of us have written on this topic (Bolnick 2008; Fujimura and Rajagopalan 2011; Ossorio and Duster 2005), and our analyses differ markedly from Shiao et al. (2012), who interpret the findings of these new algorithmic technologies in genetics as largely substantiating eighteenth-century race concepts.
2. See Fujimura, Duster, and Rajagopalan (2008) for more discussion of these issues. See also Morning (this issue).
3. To expand, "a gradual change in a character or feature across the distributional range of a species or population, usually correlated with an environmental or geographic transition" (*The American Heritage Dictionary* 2014).
4. Statistical or population geneticists construct genetic clusters by analyzing different types of genetic variants, including single nucleotide polymorphisms (SNPs), microsatellites, and copy number variants (CNVs). Many of these variants are located in noncoding regions of the genome, although some are found in gene regions.
5. We note the "60 to 150 randomly selected loci" to which Shiao et al. (2012) refer are *not* randomly selected. Human geneticists select the loci that most separate individuals. In fact, if human geneticists used 60 to 150 randomly selected loci, they would most likely find no consistent differences whatsoever.
6. Researchers have also used a variety of statistical methods to compare the likelihood scores for different values of K, including Wilcoxon two-sample tests (Rosenberg et al. 2001) and Evanno, Regnaut, and Goudet's (2005) ad hoc statistic, ΔK, or the second order rate of change of the likelihood function with respect to K.
7. It should also be noted that principal components analysis (PCA) of the same data showed gradations of change from one geographic region to the next, with some members of one geographic region or continent being more genetically similar to members of a different region/continent than to some members of their own region/continent (Jakobsson et al. 2008).
8. We note that the *Structure*-based studies described earlier in this section (Jakobsson et al. 2008; Rosenberg et al. 2002; Rosenberg et al. 2005) did not take the approach that Shiao et al. criticized.
9. PCA tools do not require unlinked loci, as does *Structure*. Genetic positions along the chromosomes ("loci") are considered "linked" if they are close enough on the chromosome to not be shuffled during the production of gametes (eggs and sperm); linked loci are inherited together, such that the sequence at one locus is not independent of sequences at other linked loci.
10. In the case of Eigenstrat PCA technologies, the genomic data often consists of SNPs, thus we call the clusters "SNP clusters." The human genome is constituted of approximately 3 billion DNA

subunits called nucleotides. Humans have identical nucleotide sequences along most of their genomes. However, researchers estimate that human genomes differ at about 1 in every 1,000 nucleotides. These sites where differences have been identified are called SNPs or single nucleotide polymorphisms. The Eigenstrat technology uses the variation in SNPs to define clusters. SNPs occur less frequently in regions that code for genes and more often in the regions of the genome that do not code for genes.

REFERENCES

Adler, Nancy E. and David H. Rehkopf. 2008. "U.S. Disparities in Health: Descriptions, Causes, and Mechanisms." *Annual Review of Public Health* 29:235-52.

American Heritage Dictionary. 2014. "Cline." Retrieved July 10 2014 (http://www.ahdictionary.com/word/search.html?q=cline).

Barbujani, Guido and Vincenza Colonna. 2010. "Human Genome Diversity: Frequently Asked Questions." *Trends in Genetics* 26(7):285-95.

Barbujani, Guido, Silvia Ghirotto, and Francesca Tassi. 2013. "Nine Things to Remember about Human Genome Diversity." *Tissue Antigens* 82(3):155-64.

Batai, Ken and Rick A. Kittles. 2013. "Race, Genetic Ancestry, and Health." *Race and Social Problems* 5(2):81-87.

Beckwith, Jon and Corey A. Morris-Singer. 2012. "Missing Heritability: Hidden Environment in Genetic Studies of Human Behavior." Pp. 122-36 in *NOMOS LII: Evolution and Morality*, edited by J. Fleming and S. Levinson. New York: New York University Press.

Benjamin, Ruha. 2009. "A Lab of Their Own: Genomic Sovereignty as Postcolonial Science Policy." *Policy and Society* 28(4):341-55.

Blumer, Herbert. 1958. "Race Prejudice as a Sense of Group Position." *The Pacific Sociological Review* 1(1):3-7.

Boas, Franz. 1913. "Changes in Bodily Form of Descendents of Immigrants." Pp. 60-75 in *Race, Language and Culture*, edited by F. Boas. New York: The Free Press/Macmillan.

Bolnick Deborah, A. 2008. "Individual Ancestry Inference and the Reification of Race as a Biological Phenomenon." Pp 77-85 in *Revisiting Race in a Genomic Age*, edited by B. Koenig, S. Lee, and S. Richardson. New Brunswick, NJ: Rutgers University Press.

Bowker, Geoffrey C. and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Brubaker, Rogers, Mara Loveman, and Peter Stamatov. 2004. "Ethnicity as Cognition." *Theory and Society* 33(1):31-64.

Brulle, Robert J. and David N. Pellow. 2006. "Environmental Justice: Human Health and Environmental Inequalities." *Annual Review of Public Health* 27:103-24.

Bryc, Katarzyna, Adam Auton, Matthew R. Nelson, Jorge R. Oksenberg, Stephen L. Hauser, Scott Williams, Alain Froment, Jean-Marie Bodo, Charles Wambebe, Sarah A. Tishkoff, and Carlos D. Bustamante. 2010. "Genome-wide Patterns of Population Structure and Admixture in West Africans and African Americans." *Proceedings of the National Academy of Sciences U.S.A*. 107(2):786-91.

Campbell, Michael C. and Sarah A. Tishkoff. 2008. "African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping." *Annual Reviews of Genomics and Human Genetics* 9:403-33.

Cavalli-Sforza, L. Luca, Paolo Menozzi, and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Chaufan, Claudia and Jay Joseph. 2013. "The 'Missing Heritability' of Common Disorders: Should Health Researchers Care?" *International Journal of Health Services* 43(2):281-303.

Cho, Mildred and Pamela Sankar. 2004. "Forensic Genetics and Ethical, Legal, and Social Implications beyond the Clinic." *Nature Genetics Supplement* 36:S8–S12.

Davis, F. James. 2001. *Who Is Black? One Nation's Definition*. University Park, PA: Pennsylvania State University Press.

DeGiorgio, Michael, Mattias Jakobsson, and Noah A. Rosenberg. 2009. "Explaining Worldwide Patterns of Human Genetic Variation Using a Coalescent-based Serial Founder Model of Migration Outward from Africa." *Proceedings of the National Academy of Sciences U.S.A*. 106(38):16057-62.

Du Bois, W.E.B. 1940. *Dusk of Dawn: An Essay toward an Autobiography of a Race Concept*. New York: Harcourt.

Duster, Troy. 2003. *Backdoor to Eugenics*. New York: Routledge.

Duster, Troy. 2006. "Comparative Perspectives and Competing Explanations: Taking on the Newly Configured Reductionist Challenge to Sociology." *American Sociological Review* 71(1):1-15.

Edwards, A. W. F. 2003. "Human Genetic Diversity: Lewontin's Fallacy." *BioEssays* 25:798-801.

Evanno, Guillaume, Sebastien Regnaut, and Jérôme Goudet. 2005. "Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study." *Molecular Ecology* 14(8):2611-20.

Fausto-Sterling, Anne. 2008. "The Bare Bones of Race." *Social Studies of Science* 38(5):657-94.

Feldman, Marcus and Richard C. Lewontin. 2008. "Race, Ancestry, and Medicine." Pp 89-101 in *Revisiting Race in a Genomic Age*, edited by B. Koenig, S. Lee, and S. Richardson. New Brunswick, NJ: Rutgers University Press.

Fujimura, Joan H. 2005. "Postgenomic Futures: Translations across the Machine-Nature Border in Systems Biology." *New Genetics and Society* 24(2):195-225.

Fujimura, Joan H. and Ramya Rajagopalan. 2011. "Different Differences: The Use of Ancestry versus Race in Biomedical Human Genetic Research." *Social Studies of Science* 41(1):5-30.

Fujimura, Joan H., Troy Duster, and Ramya Rajagopalan. 2008. "Race, Genetics, and Disease: Questions of Evidence, Matters of Consequence." *Social Studies of Science* 38(5):643-56.

Goodman, Alan H. 2000. "Why Genes Don't Count (for Racial Differences in Health)." *American Journal of Public Health* 90(11):1699-702.

Goodman, Alan H. 2006. "Two Questions about Race." Is Race 'Real?' A Web Forum Organized by the Social Science Research Council. Social Science Research Council. Retrieved July 10 2014 (http://raceandgenomics.ssrc.org/Goodman/).

Hacking, Ian. 2005. "Why Race Still Matters." *Daedalus* 134(1):102-16.

Haraway, Donna J. 1997. "Race: Universal Donors in a Vampire Culture." Pp. 213-66 in *Modest_Witness@Second_Millennium.FemaleMan©_Meets_OncoMouse⇔: Feminism and Technoscience*. New York: Routledge.

Huxley, Julian. 1938. "Clines: An Auxiliary Taxonomic Principle." *Nature* 142:219-20.

Jablonski, Nina G. and George Chaplin. 2010. "Human Skin Pigmentation as an Adaptation to UV Radiation." *Proceedings of the National Academy of Sciences U.S.A*. 107(suppl 2):8962-68.

Jakobsson, Mattias, Sonja W. Scholz, Paul Scheet, J. Raphael Gibbs, Jenna M. VanLiere, Hon-Chung Fung, Zachary A. Szpiech, James H. Degnan, Kai Wang, Rita Guerreiro, Jose M. Bras, Jennifer C. Schymick, Dena G. Hernandez, Bryan J. Traynor, Javier Simon-Sanchez, Mar Matarin, Angela Britton, Joyce van de Leemput, Ian Rafferty, Maja Bucan, Howard M. Cann, John A. Hardy, Noah A. Rosenberg, and Andrew B. Singleton. 2008. "Genotype, Haplotype and Copy-Number Variation in Worldwide Human Populations. *Nature* 451(7181):998-1003.

Jay, Flora, Per Sjōdin, Mattias Jakobsson, and Michael G. B. Blum. 2013. "Anisotropic Isolation by Distance: The Main Orientations of Human Genetic Differentiation." *Molecular Biology and Evolution* 30(3):513-25.

Jorde, Lynn B. and Stephen P. Wooding. 2004. "Genetic Variation, Classification and 'Race.'" *Nature Genetics* 36:S28-S33.

Joseph, Jay. 2011. "The Crumbling Pillars of Behavioral Genetics." *GeneWatch* 24:4-7.

Kahn, Robert S., Kathryn Wilson, and Paul H. Wise. 2005. "Intergenerational Health Disparities: Socioeconomic Status, Women's Health Conditions, and Child Behavior Problems." *Public Health Reports* 120(4):399-408.

Kalinowski, Steven T. 2011. "The Computer Program STRUCTURE Does Not Reliably Identify the Main Genetic Clusters within Species: Simulations and Implications for Human Population Structure." *Heredity* 106(4):625-32.

Kim, Nadia Y. 2008. *Imperial Citizens: Koreans and Race from Seoul to LA*. Stanford, CA: Stanford University Press.

Krieger, Nancy. 2012. "Methods for the Scientific Study of Discrimination and Health: From Societal Injustice to Embodied Inequality—An Ecosocial Approach." *American Journal of Public Health* 102(5):936-44.

Lawson Handley, Lori J., Andrea Manica, Jérôme Goudet, and François Balloux. 2007. "Going the Distance: Human Population Genetics in a Clinical World." *Trends in Genetics* 23(9):432-39.

Lewontin, Richard C. 1972. "The Apportionment of Human Diversity." *Evolutionary Biology* 6:381-98.
Li, Jun Z. Li, Devin M. Absher, Hua Tang, Audrey M. Southwick, Amanda M. Casto, Sohini Ramachandran, Howard M. Cann, Gregory S. Barsh, Marcus W. Feldman, Luigi L. Cavalli-Sforza, and Richard M. Myers. 2008. "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation." *Science* 319(5866):1100-04.
Livingstone, Frank. B. 1962. "On the Non-Existence of Human Races." *Current Anthropology* 3:279-81.
Long, Jeffrey C., Jie Li, and Meghan E Healy. 2009. "Human DNA Sequences: More Variation and Less Race." *American Journal of Physical Anthropology* 139:23-34.
López, Ian Haney. 1996. *White by Law: The Legal Construction of Race*. New York: New York University Press.
Loveman, Mara and Jeronimo M. Muniz. 2008. "How Puerto Rico Became White: Boundary Dynamics and Intercensus Racial Reclassification." *American Sociological Review* 72(6):915-39.
Marks Jonathan. 2008. "Race: Past, Present, and Future". Pp. 21-38 in *Revisiting Race in a Genomic Age*, edited by B. Koenig, S. Lee, and S. Richardson. New Brunswick, NJ: Rutgers University Press.
Marks Jonathan. 2010. "Ten Facts about Human Variation." Pp. 265-76 in *Human Evolutionary Biology*, edited by M. Muehlenbein. New York: Cambridge University Press.
Morning, Ann. 2009. "Toward a Sociology of Racial Conceptualization for the 21st Century." *Social Forces* 87(3):1167-92.
Omi, Michael and Howard Winant. 1994. *Racial Formation in the United States: From the 1960s to the 1990s*. New York: Routledge.
O'Rourke, Dennis H. and Jennifer A. Raff. 2010. "The Genetic History of the Americas: The Final Frontier." *Current Biology* 20:R202-R207.
Ossorio, Pilar and Troy Duster. 2005. "Race and Genetics: Controversies in Biomedical, Behavioral, and Forensic Sciences." *American Psychologist* 60:115-28.
Oyama, Susan. 2000. *The Ontogeny of Information: Developmental Systems and Evolution*. 2nd ed. Durham, NC: Duke University Press.
Paschou, Peristera, Elad Ziv, Esteban G. Burchard, Shweta Choudhry, William Rodriguez-Cintron, Michael W. Mahoney, and Petros Drineas. 2007. "PCA-Correlated SNPs for Structure Identification in Worldwide Human Populations." *PLoS Genetics* 3(9):1672-86.
Patterson, Nick, Alkes L. Price, and David Reich. 2006. "Population Structure and Eigenanalysis." *PLoS Genetics* 2(12):2074-93.
Power, Chris, Diana Kuh, and Susan Morton. 2013. "From Developmental Origins of Adult Disease to Life Course Research on Adult Disease and Aging: Insights from Birth Cohort Studies." *Annual Review of Public Health* 34:7-28.
Price, Alkes L., Johannah Butler, Nick Patterson, Cristian Capelli, Vincenzo L. Pascali, Francesca Scarnicci, Andres Ruiz-Linares, Leif Groop, Angelica A. Saetta, Penelope Korkolopoulou, Uri Seligsohn, Alicja Waliszewska, Christine Schirmer, Kristin Ardlie, Alexis Ramos, James Nemesh, Lori Arbeitman, David B. Goldstein, David Reich and Joel N. Hirschhorn. 2008. "Discerning the Ancestry of European Americans in Genetic Association Studies." *PLoS Genetics* 4(1):9-17.
Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155(2):945-59.
Pritchard, Jonathan K. and W. Wen. 2004. *Documentation for structure software: Version 2*. Chicago, IL.
Raghavan, Maanasa, Pontus Skoglund, Kelly E. Graf, Mait Metspalu, Anders Albrechtsen, Ida Moltke, Simon Rasmussen, Thomas W. Stafford, Jr. Ludovic Orlando, Ene Metspalu, Monika Karmin, Kristiina Tambets, Siiri Rootsi, Reedik Mägi, Paula F. Campos, Elena Balanovska, Oleg Balanovsky, Elza Khusnutdinova, Sergey Litvinov, Ludmila P. Osipova, Sardana A. Fedorova, Mikhail I. Voevoda, Michael DeGiorgio, Thomas Sicheritz-Ponten, Søren Brunak, Svetlana Demeshchenko, Toomas Kivisild, Richard Villems, Rasmus Nielsen, Mattias Jakobsson, and Eske Willerslev. 2014. "Upper Palaeolithic Siberian Genome Reveals Dual Ancestry of Native Americans." *Nature* 505:87-91.
Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza. 2005. "Support for the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa." *Proceedings of the National Academy of Sciences U.S.A*. 102(44):15942-47.

Relethford, John H. 2004. "Global Patterns of Isolation by Distance Based on Genetic and Morphological Data." *Human Biology* 76(4):499-513.

Relethford, John H. 2009. "Race and Global Patterns of Phenotypic Variation." *American Journal of Physical Anthropology* 139(1):16-22.

Risch, Neil, Esteban Burchard, Elad Ziv, and Hua Tang. 2002. "Categorization of Humans in Biomedical Research: Genes, Race and Disease." *Genome Biology* 3(7):1-12.

Rosenberg, Noah A., Terry Burke, Kari Elo, Marcus W. Feldman, Paul J. Freidlin, Martien A. M. Groenen, Jossi Hillel, Asko Mäki-Tanila, Michèle Tixier-Boichard, Alain Vignal, Klaus Wimmers, and Steffen Weigend. 2001. "Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes from 20 Chicken Breeds." *Genetics* 159:699-713.

Rosenberg, Noah A., Jonathan K. Pritchard, James L. Weber, Howard M. Cann, Kenneth K. Kidd, Lev A. Zhivotovsky, and Marcus W. Feldman. 2002. "Genetic Structure of Human Populations." *Science* 298(5602):2381-85.

Rosenberg, Noan A., Saurabh Mahajan, Sohini Ramachandran, Chengfeng Zhao, Jonathan K. Pritchard, and Marcus W. Feldman. 2005. "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure." *PLoS Genetics* 1(6):660-71.

Rosser, Zeë H., Tatiana Zerjal, Matthew E. Hurles, Maarja Adojaan, Dragan Alavantic, António Amorim, William Amos, Manuel Armenteros, Eduardo Arroyo, Guido Barbujani, Gunhild Beckman, Lars Beckman, Jaume Bertranpetit, Elena Bosch, Daniel G. Bradley, Gaute Brede, Gillian Cooper, Helena B. S. M. Côrte-Real, Peter de Knijff, Ronny Decorte, Yuri E. Dubrova, Oleg Evgrafov, Anja Gilissen, Sanja Glisic, Mukaddes Gölge, Emmeline W. Hill, Anna Jeziorowska, Luba Kalaydjieva, Manfred Kayser, Toomas Kivisild, Sergey A. Kravchenko, Astrida Krumina, Vaidutis Kučinskas, João Lavinha, Ludmila A. Livshits, Patrizia Malaspina, Syrrou Maria, Ken McElreavey, Thomas A. Meitinger, Aavo-Valdur Mikelsaar, R. John Mitchell, Khedoudja Nafa, Jayne Nicholson, Søren Nørby, Arpita Pandya, Jüri Parik, Philippos C. Patsalis, Luísa Pereira, Borut Peterlin, Gerli Pielberg, Maria João Prata, Carlo Previderé, Lutz Roewer, Siiri Rootsi, D. C. Rubinsztein, Juliette Saillard, Fabrício R. Santos, § Gheorghe Stefanescu, Bryan C. Sykes, Aslihan Tolun, Richard Villems, Chris Tyler-Smith, and Mark A. Jobling. 2000. "Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language." *American Journal of Human Genetics* 67(6):1526-43.

Sankar, Pamela. 2008. "Moving beyond the Two Race Mantra." Pp 271-84 in *Revisiting Race in a Genomic Age*, edited by B. Koenig, S. Lee, and S. Richardson. New Brunswick, NJ: Rutgers University Press.

Schwartz, Michael K. and Kevin S. McKelvey. 2009. "Why Sampling Scheme Matters: The Effect of Sampling Scheme on Landscape Genetic Results." *Conservation Genetics* 10:441-52.

Serre, David and Svante Pääbo, S. 2004. "Evidence for Gradients of Human Genetic Diversity within and among Continents." *Genome Research* 14:1679-85.

Sesardic, Neven. 2010. "Race: A Social Destruction of a Biological Concept." *Biology and Philosophy* 25:143-62.

Shiao, Jiannbin Lee, Thomas Bode, Amber Beyer, and Daniel Selvig. 2012. "The Genomic Challenge to the Social Construction of Race." *Sociological Theory* 30(2):67-88.

Stepan, Nancy Leys (1986) 'Race and Gender: The Role of Analogy in Science', *ISIS* 77(2):261–277.

Sturm, Richard A. and David L. Duffy. 2012. "Human Pigmentation Genes under Environmental Selection." *Genome Biology* 13:248-63.

Telles, Edward Eric. 2004. *Race in another America: The Significance of Skin Color in Brazil*. Princeton, NJ: Princeton University Press.

Templeton, Alan R. 1999. "Human Races: A Genetic and Evolutionary Perspective." *American Anthropologist* 100(3):632-50.

Templeton, Alan R. 2013. "Biological Races in Humans." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44(3)262-71.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, Agnes A. Awomoyi, Jean-Marie Bodo, Ogobara Doumbo, Muntaser Ibrahim, Abdalla T. Juma, Maritha J. Kotze, Godfrey Lema, Jason H. Moore, Holly Mortensen, Thomas B. Nyambo, Sabah A. Omar, Kweli Powell, Gideon S. Pretorius, Michael W. Smith, Mahamadou A. Thera, Charles Wambebe, James L. Weber, and Scott M. Williams. 2009. "The Genetic Structure and History of Africans and African Americans." *Science* 324(5930):1035-44.

Turkheimer, Eric. 2011. "Commentary: Variation and Causation in the Environment and Genome." *International Journal of Epidemiology* 40(3):598-601.

Underhill, Peter A. and Toomas Kivisild. 2007. "Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations." *Annual Reviews of Genetics* 41:539-64.

Walsemann, Katrina M., Arline T. Geronimus, and Gilbert C. Gee. 2008. "Accumulating Disadvantage over the Life Course—Evidence from a Longitudinal Study Investigating the Relationship between Educational Advantage in Youth and Health in Middle Age." *Research on Aging* 30(2):169-99.

Washington, Scott L. 2011. "Hypodescent: A History of the Crystallization of the One-Drop Rule in the United States, 1880-1940." PhD dissertation, Princeton University. Publication Number: AAT 3480237.

Weiss, Kenneth M. and Jeffrey C. Long. 2009. "Non-Darwinian Estimation: My Ancestors, My Genes' Ancestors." *Genome Research* 19:703-10.

Willson, Andrea E., Shuey, Kim M., and Glen H. Elder, Jr. 2007. "Mechanisms of Inequality in Life Course Health." *American Journal of Sociology* 112(6):1886-924.

Xiao, Feng-Xia, Vania Yotova, Ewa Zietkiewicz, Alan Lovell, Dominik Gehl, Stéphane Bourgeois, Claudia Moreau, Cleanthe Spanaki, Andreas Plaitakis, Jean-Paul Moisan, and Damian Labuda. 2003. "Human X-Chromosomal Lineages in Europe Reveal Middle Eastern and Asiatic Contacts." *European Journal of Human Genetics* 12:301-11.

Yamaguchi-Kabata, Yumi, Kazuyuki Nakazono, Atsushi Takahashi, Susumu Saito, Naoya Hosono, Michiaki Kubo, Yusuke Nakamura, and Naoyuki Kamatani. 2008. "Japanese Population Structure, Based on SNP Genotypes from 7003 Individuals Compared to Other Ethnic Groups: Effects on Population-Based Association Studies." *American Journal of Human Genetics* 83(4):445-56.

AUTHOR BIOGRAPHIES

Joan H. Fujimura is professor in the Department of Sociology and in the Holtz Center for Research on Science, Technology, and Medicine at the University of Wisconsin-Madison. Her research has focused on the sociology of science, medicine, and technology–especially on developments in genetics. She has studied contemporary genomics and its engagements with questions of race and ethnicity since the early days of human genetic variation studies following the completion of the Human Genome Project. Her current research focuses on developments in epigenetics and systems biology and on the topics of "big data" analysis and interdisciplinarity in the life sciences.

Deborah A. Bolnick is associate professor in the Department of Anthropology and a faculty research associate of the Population Research Center at the University of Texas at Austin. She is a genetic anthropologist and population geneticist, and her research focuses on human genomic variation, race, genetic ancestry, and population history, with a particular focus on Native American populations.

Ramya Rajagopalan has been studying human genomics and its implications for notions of race, identity and difference since 2007. As a postdoctoral fellow at the University of Wisconsin-Madison, she has co-authored several articles, including "Different Differences" (2011) which was published in *Social Studies of Science*. Her current research examines how researchers and clinicians mobilize different sociopolitical, technical and regulatory frames to bring tools for personalized medicine to the clinic, including ideas about race, genetics and difference.

Jay S. Kaufman is professor and Canada Research Chair in Health Disparities in the Department of Epidemiology, Biostatistics and Occupational Health at McGill University and holds appointments at the Centre on Population Dynamics and the Institute for Health and Social Policy. He conducts research in population health inequalities in Canada, the United States, Chile, Peru and other settings.

Richard C. Lewontin is emeritus professor of biology and emeritus Alexander Agassiz professor in zoology in the Museum of Comparative Zoology, at Harvard University. He is an evolutionary geneticist, and his research has concentrated on both theoretical and experimental studies of genetic variation within and between populations, including humans.

Troy Duster is chancellor's professor and senior fellow at the Warren Institute for Law and Social Policy at the University of California, Berkeley. He is the author of many articles and a book on the relationship between recent developments in molecular genetics and shifting understandings of human taxonomies.

Pilar Ossorio is professor of law and bioethics at the University of Wisconsin-Madison, where she is on the faculties of the Law School and the Department of Medical History and Bioethics at the Medical School and is affiliated with the Morgridge Institute for Research. Her work focuses on ethical issues in data sharing and data use, research with human subjects, research regulation, and the use of race and ethnicity in biomedical research.

Jonathan Marks is professor of anthropology at the University of North Carolina at Charlotte, with primary interests in human evolution and diversity. He is the author of many scholarly articles and books, including *What It Means to Be 98% Chimpanzee* (2003) and *Why I Am Not a Scientist* (2009).