Published in final edited form as:

J Health Soc Behav. 2014 December ; 55(4): 504–518. doi:10.1177/0022146514555224.

Race and Ancestry in the Age of Inclusion: Technique and Meaning in Post-Genomic Science

Janet K. Shim,

University of California, San Francisco

Sara L. Ackerman,

University of California, San Francisco

Katherine Weatherford Darling,

University of California, San Francisco

Robert A. Hiatt, and

University of California, San Francisco

Sandra Soo-Jin Lee

Stanford University

Abstract

This paper examines how race and ancestry are taken up in gene-environment interaction (GEI) research on complex diseases such as heart disease, diabetes, and cancer. Using 54 in-depth interviews of 33 scientists and over 200 hours of observation at scientific conferences, we explore how GEI researchers use and interpret race, ethnicity, and ancestry in their work. We find that the use of self-identified race and ethnicity (SIRE) exists alongside ancestry informative markers (AIMs) to ascertain genetic ancestry. Our participants assess the utility of these two techniques in relative terms, downplaying the accuracy and value of SIRE compared to the precision and necessity of AIMs. In doing so, we argue that post-genomic scientists seeking to understand the interactions of genetic and environmental disease determinants actually undermine their ability to do so, by valorizing precise characterizations of individuals' genetic ancestry over measurement of the social processes and relations that differentiate social groups.

The story of the genomic revolution often begins with the June 26, 2000 announcement of the completion of a working draft of the human genome. Many interpreted this much-hailed millennial event as substantiating claims about a universal humanity: the commonalities that bind us to each other could be read in the more than 99% of our genetic make-up that is shared among all humans. These claims were mobilized explicitly to argue against social divisions and hierarchies, and that there simply was no biological basis for making fundamental distinctions among groups and populations. But as is well known by now, post-genomic science continues to be animated by the logic of difference, of examining that less

* Address correspondence to Janet K. Shim, Department of Social and Behavioral Sciences, Box 0612, University of California, San Francisco, 3333 California Street, Suite 455, San Francisco, CA, 94143-0612 (janet.shim@ucsf.edu).

than 1% of the genome that varies from individual to individual, and its significance for health.

In addition to a focus on difference (Shim et al. 2014), we argue that post-genomic research may also be characterized with reference to two other motifs. The first is encapsulated in Steven Epstein's (2007) *Inclusion*. As he traces from the 1990s onward in the United States, we are now fully in an age of inclusion, in which the regulatory mandate to include under-represented populations has become both scientific and social common sense. A second point of departure is embodied in a number of publications, many of them authored by prominent leaders from the U.S. National Institutes of Health (e.g., Collins et al. 2003; Manolio, Bailey-Wilson, and Collins 2006), that call for research exploring the mutual and synergistic interactions of genetic and environmental determinants on human health. A new "sociogenomic" paradigm, in Bliss' (2012) terms, now combines both social and biological inquiry with a race-conscious genomics that explicitly accounts for the social and embodied realities of race.

Here, we explore how genome scientists are taking up these two points of reference, of inclusion on the one hand, and interaction on the other. In particular, we seek to understand how research that is explicitly intended to illuminate the interactions of genes with environment, and their joint impacts on disease incidence and distribution, incorporate and conceive of race, ethnicity, and ancestry. We find that those conducting gene-environment interaction (GEI) research invoke an inclusionary common sense: they see the inclusion of multiple racial and ethnic groups in research participation as a representational and social justice issue. The scientists we observed and interviewed seem to also be guided not just by a moral commitment to inclusion but also by a *scientific* sensibility that diversity and difference are necessary to properly study complex disease causation within broadly representative populations, address health disparities, and pose questions of significant population health impact. We describe how such scientific questions are pursued via practices of using self-identified race and ethnicity (SIRE) and ancestry informative markers (AIMs) and how these techniques are invested with meaning. But in doing so, post-genomic scientists also tend to displace political issues of inclusion and scientific accountability with technical questions of classification, adequate measurement, and precision. We argue that such displacements in turn have consequences for our understandings of human difference and the kinds of research and interventions that can materially improve the public's health.

BACKGROUND

With this article, we aim to contribute to an expanding literature on the invocation, uses, and constructions of race and population differences in genomic research (e.g., Abu El-Haj 2007; Bliss 2012; Braun and Hammonds 2008; Fujimura and Rajagopalan 2011; Fullwiley 2007, 2008, 2011; Lee 2006; Lee et al. 2008; Montoya 2011; Reardon 2005, 2009; Whitmarsh 2008). Many of these authors have examined the question of whether genomic and post-genomic science simply recapitulates categorical and essentialist notions of race or whether they rework those concepts in any significant way. One widespread claim in the genomics literature is that the approaches and techniques of human population genetics enable a shift from typological notions of *race* to statistical notions of difference among

populations (e.g., Cavalli-Sforza and Bodmer 1974/1999; Dobzhansky 1963; Lewontin 1972). However, multiple scholars who have analyzed genomic and post-genomic research practices take serious issue with the claim that ‘old’ ideas about racial difference and hierarchy have been eclipsed.

Gannett (2001), for example, found that the notion of population has not attenuated scientists’ inclination to identify and measure group differences that incorporate cultural meanings of race or precluded stereotypical and racist interpretations of statistical patterns. In fact, Lee (2005) identified an “infrastructure of racialization” in post-genomic research wherein biobanks incorporated racialized labels for groups whose DNA samples they store. In so doing, they privileged the human body as the grounds in and from which racial difference can be read and genomics as the official language for reading race. Fullwiley (2007, 2008) found much the same in her ethnography of two medical genetics laboratories in the U.S. There, the comparison of purportedly pure ‘Old World’ reference populations to ‘New World’ admixed populations led scientists to read race in DNA and revive race as biogenetically valid.

In examining race-based genetic research on asthma conducted in Barbados, Whitmarsh (2008) found that while initially the biological, medical, geographical, and socially constructed were all simultaneously seen as potentially defining or diagnostic of race, this capaciousness disappeared once data were analyzed. When scientists had to adjudicate, for example, the representativeness of a population to stand in for other populations, or the generalizability of a set of findings, they then resorted to pragmatic and expedient efforts to measure ‘race’ precisely by admixture, for instance, or by presence of genetic variations. And similarly, Montoya (2007, 2011) found in a Type 2 diabetes study that enrolled Mexican Americans in a county on the Texas-Mexico border that researchers on the one hand acknowledged the joint genetic and environmental determinants of diabetes. However, on the other hand, “by deploying the ideology of admixture and hereditary disease etiology as the rationale for Mexicana/o sampling, researchers and DNA donation field staff construct research participants as genetic carriers.... By accepting the biogenetic risk narrative, affected people cede the sociocultural risk factors so central to disease” (2011:103–4).

Fujimura and Rajagopalan (2011) report that in genome-wide association studies, while many researchers used racial classifications in their analyses, others explicitly avoided them. This ostensible avoidance was facilitated through the use of statistical software that in *theory* allowed geneticists to identify and define population clusters without an a priori use or conceptualization of race. However, some of these same researchers then began to link genetic similarities to shared “genetic history” and shared “ancestry” in ways that invoked geographic origins. This intertwining of genetic analysis technologies, the populations they generate, attributions of genetic similarities, and circulating geographical assessments and labels constitute, in Fujimura and Rajagopalan’s terms, “genome geography.” Thus, although new technologies and scientific commitments offered opportunities to uncouple genetic research on disease from racial categories, Fujimura and Rajagopalan argue that genome geography often—though not inevitably—served to overlay older concepts of race into and onto genomic research.

Finally, Bliss (2011, 2012) found in her study of elite genome scientists that even while these scientists acknowledged the great heterogeneities within racial groups, they attended to race in research participation as a means to ensure racial equality in research and drug development. In the shift from “colorblind science to race positive inquiry” (2012:67), Bliss observed that elite scientists were no longer content to use ‘race’ as a catch-all proxy or black box variable for other potential disease determinants. Instead, they were committed to measuring those previously uninterrogated factors in and of themselves. Bliss argues that within an “anti-racist racialism,” researchers strategically incorporated racial populations into genome science with the goals of deconstructing biological race and producing an inclusive genomic science.

Given these emergent and enduring connections between genome science, race, and racism, we track in this article the shifting practices of using self-identified race and ethnicity (SIRE) and ancestry informative markers (AIMs) and scientists’ specific concerns about and commitments to the investigation of human and social differences. We argue that their uses of techniques of self-identification and genetic admixture measurement both construct and reflect particular meanings about racial subjects and the effects of race on health. We find that post-genomic scientists involved in gene-environment interaction research often readily express the social and political value of inclusion and justify their research on and practices vis-à-vis specific racial and ethnic populations in those very terms. But how they do so, the relative uses they put SIRE and AIMs to, and the relative values they invest in those practices can all too easily sidestep and subsume accountability for investigating and intervening on health disparities, with a preoccupation with technical questions of classification, measurement, and precision. Thus we—like other scholars—find that while post-genomic scientists may (unevenly) break from ‘old,’ essentializing, categorical understandings of race, there is no clear concomitant turn to ‘new,’ statistical notions of difference. Nor have such breaks resulted as yet in widely changed thinking about the connections of race to health. By examining the increasing use and valorization of AIMs to solve technical problems in GEI research, our study shows how scientists’ specific techniques for measuring and including race may shift, yet the meanings invested in racial difference may still prevail.

DATA AND METHODS

From 2010–2014, we recruited U.S.-based genetic epidemiologists and other genome scientists currently conducting research that fit a number of eligibility criteria: (1) identifiable as GEI research, that is, studies that seek to account for the interactions of genetic with environmental factors; (2) inclusion of at least one non-white population or studies designed to enable comparisons between racial/ethnic groups; and (3) research on the etiology of heart disease, type 2 diabetes, and cancer—all conditions widely acknowledged to be caused by multiple, interacting genetic, behavioral, and environmental determinants. These criteria were chosen to ensure an interviewee sample comprised of scientists who were explicitly engaged in research across the causal spectrum (from the genetic to the behavioral to the environmental), had some experience conducting research on diverse populations and on disease conditions commonly understood to be multicausal and that have disproportionate impact by race and ethnicity. We also conducted observations and

informal interviews at scientific conferences where findings from GEI research were presented. The institutional review boards at both the University of California, San Francisco and Stanford University approved this study.

This paper is based on 54 in-depth interviews with 33 GEI scientists (21 of these interviews were follow-up interviews conducted approximately one year after the first interview) and over 200 hours of observation and informal interviews at nine scientific conferences. All interviews were transcribed verbatim, while observation data consisted of fieldnotes along with selected verbatim transcriptions of podium presentations and question-and-answer sessions.

All of the 33 participants whose interview data were analyzed for this paper are affiliated with either U.S. academic institutions or the U.S. National Institutes of Health. GEI research is an inherently multidisciplinary field, and we therefore asked our participants to characterize the work that they do rather than to locate themselves within a particular discipline. In the course of answering this question, our participants reported a range of disciplinary affiliations (numbers total greater than 33 because many participants listed more than one affiliation): “epidemiology” (n=9) or some specific field of epidemiology (n=6); “genetic epidemiology”¹ (n=10); “molecular epidemiology” (n=3); and “genetics” (n=8). Three other disciplinary affiliations were mentioned by just one participant each (e.g., “physician” or “cancer researcher”), and an additional two scientists did not name any disciplines or sub-fields. Given this disciplinary heterogeneity, in this paper we refer our study participants more generically as “GEI researchers” or “GEI scientists.” Moreover, given the relatively small numbers of participants within specified sub-fields, we opted not to make claims about the potential connections between disciplinary affiliation and the kinds of research practices they engage in.

All data were uploaded to the qualitative data analysis software ATLAS.ti. The research team undertook a collaborative reading and analysis of a sub-set of interviews to generate inductively the initial codes. Through successive waves of independent and joint coding and comparison of coding across research team members, we then finalized a list of approximately 80 categories and codes. The ATLAS.ti query tool was used to extract data tagged with particular codes and intersections of codes (e.g., the code “study design” and the intersection of codes “race/ethnicity” × “recruitment”). We wrote, circulated, and revised memos on these queries in order to generate the findings described in this paper.

RESULTS

In what follows, we explore how the diversity of populations is being accounted for by the GEI scientists in our study, through a close examination of two primary measures or constructs of race, ethnicity, and ancestry employed in contemporary research on complex disease. The sections below address in turn these two constructs, self-identified race and ethnicity (SIRE) and ancestry informative markers (AIMs). For each, we describe how

¹We acknowledge here Montoya’s (2011:58) description of genetic epidemiology as Janus-faced, where *genetic* epidemiology refers to “the use of populations to understand the genetics of disease,” and genetic *epidemiology* refers to “the use of genetics to understand disease in populations.” We also saw this difference in emphasis among our participants as well.

scientists articulate both the “how” and the “why” of SIRE and AIMS, that is, SIRE and AIMS as both technique and meaning. We then offer a discussion and reflection, as viewed through the lens of vignettes from our fieldwork, on how our scientist-participants’ use and understandings of SIRE and AIMS work to displace political, moral, and scientific commitments to representation and understanding disease disparities with more technical issues of precision.

SIRE as technique

The vast majority of our participants use SIRE to refer to the racial and ethnic categories defined by the U.S. Office of Management and Budget (OMB) to be used in biomedical research that receives funding from the U.S. National Institutes of Health (NIH). This use of SIRE is seen by our participants as now embedded and entrenched in the social practices of conducting research, as mandated by NIH. One told us, “All NIH research requires that we use a category, so I have to do that.... Right now, that’s what we have. So I have to use that.” Another related that “[We’re] pretty much conforming with what we have to report to the NIH.... So we kind of make that question fit those requirements.” This use of SIRE underscores its characterization as a technique of convenience. As one scientist rationalized, “I think the path of least resistance is to stick with self-identified race/ethnicity,” and another explained that, “When we’re measuring it de novo ... I try to ask it in categories that conform to the NCI [National Cancer Institute, one of the institutes that make up the NIH] reporting requirements. Make it easy on everybody.”

A major part of this compulsion to use pre-existing NIH racial and ethnic categories to “make things easy” is to facilitate combining their study’s data with those of others. While the imperative to have samples of sufficient size in order to achieve meaningful statistical power is relevant for all quantitative studies, it is of particular concern to the scientists we interviewed and observed who aim to study gene-environment interactions. In order to accommodate the larger number of variables that such studies typically involve, as well as the possibility for identifying significant interactions between genetic and environmental factors, extremely large sample sizes are required. This in turn makes a study sample and data set that are harmonizable with other samples—where data are combinable and variable definitions and measurements can be easily reconciled—highly desirable. Using routinized racial and ethnic categories was one clear way our participants structured their samples in order to maximize their potential to be recombined with other data sets. Thus when scientists invoke the quality of “ease,” they refer almost exclusively to the ease for researchers—rather than, for instance, to the ease of their human subjects in finding categories that align with their own identities. Moreover, even when scientists speak of collecting racial-ethnic data “de novo,” they are not doing this truly anew, given that categories developed and promulgated by federal agencies pre-exist their own studies. Such upstream standardization works to prevent downstream harmonization problems and thereby facilitates efforts to amass the large datasets that are necessary to make claims about the joint and interacting effects of genetic and environmental risks on disease.

On the other hand, what may appear on the surface to be a widely standardized, unproblematic practice can still come with complications. That is, what may be “easy” and

“convenient” in fact may not be very clear or clean-cut. For example, several of the scientists mentioned the confusion regarding the status of “Hispanic/Latino” as an ethnicity but not a race. One participant said explicitly that while they report Hispanic *ethnicity* to the NIH as mandated, when they do data analyses in their study, they consider Hispanic to be another *race*. Moreover, because this study allowed for claiming multiple races in self-identification, it further complicated the use of SIRE. As this scientist explained:

When we’re actually analyzing data, we create Hispanic as another one of the basic categories.... If people report multiple things, then we say African-American trumps everything else. Then Hispanic then trumps everything else that’s remaining. And then the Asian trumps everything else that’s remaining. That’s kind of what I think the standard thing that people do. And of course we may end up misclassifying some of the people in terms of how they ... actually perceive themselves.

Notably, this participant on the one hand affirms the right for subjects to self-identify as Hispanic/Latino or as multi-racial, but then on the other hand needs to assign them to only one category in the analysis phase. As this scientist alluded, this was a commonly mentioned and utilized practice by our other participants in order to maximize the size of sub-populations within their samples. However, rather than simply being a technical solution to facilitate sub-group analyses that might “misclassify” some people, this practice of reassignment also reflects and reifies the notion that counting and knowing about African Americans are more valuable than about Hispanics, or about Asian Americans, in turn.

Other scientists spoke of how they resolved the issue of identification when human subjects chose not to self-identify. For example, one of our participants explained the procedures for assignment into a racial and ethnic category based on surname:

We use a Bayesian algorithm developed actually by RAND [Corporation], that’s based on a person’s surname and where they live for Census information to give them a probability of being in one of six buckets: white, black, Asian, Hispanic/Latino, Native American or mixed.... When we validate that across the people we have ... we find that it’s not perfect; using that you would probably underestimate blacks, [and] overestimate Asians.

In direct contrast, other scientists in fact expressly counter the OMB/NIH need for pre-designated categories, instead allowing for more capacious notions of “self-identification”:

I always use [OMB categories] in conjunction with my own thing, whatever I believe is most useful for that particular study. So I will ask people to identify themselves based on how they do it in the Census and give those categories and have people check those boxes. And then I often, depending on the study, if it’s a small study, I’ll have an open-ended box for people to write how they describe themselves or whatever.

In making such comments, GEI scientists express both a sense of curiosity about the myriad ways in which people see themselves and define their own identities as well as the desire to accommodate and respect such diverse identity practices with ever more precision.

But yet, even when some considered alternatives to the usual OMB categories, the weight of their standardization as a taken-for-granted practice ultimately made them the fallback or default mode of classification used in analyses. As one participant recalled:

We've often gone to try to get more granular self-report information ... having lots of different possibilities, having lots of different multi-ethnic designations.... What we do is we give them the opportunities, then we look at the numbers [of participants choosing those other designations], and it's so small ... that we ended up having to fold them back into the same categorization that we've talked about here that has been used for a long time.

Thus even for those scientists who offer multiple possibilities for self-identification that try to go beyond just picking one box, the numbers of subjects who took advantage of such opportunities were so negligible as to render those more capacious identifications irrelevant when it came to data analysis.

As these quotes from our participants illustrate, the technique of using SIRE, while seemingly over-determined by OMB regulations, actually involves a number of predicaments and uncertainties, in the face of which scientists develop a range of workarounds. But even so, conducting research in a “world of standards” (e.g., Timmermans and Berg 2003; Timmermans and Epstein 2010) tips the scales in favor of reusing and replicating those same standardized categories, though not always and not strictly in the ways they were intended. The simultaneity of the workarounds scientists have arbitrated alongside the continued implementation of SIRE therefore seem to exemplify what Michelle Murphy (2012:183, n3) calls “counter-conduct,” or “modes of undoing, remaking, and antagonism that are immanent with and animated by hegemonic formations.” That is, while these alternate practices attempt to remake how research accommodates self-identification with more meaningful practices, they often do so in ways that work to reanimate the relevance and circulation of SIRE. This, we argue, reveals scientists' ambivalence about the *scientific* utility of study subjects' actual, contingent, and complex self-identification, even as they express respect for and commitments to individuals' self-representation and self-determination in racial and ethnic terms.

As a technique, then, SIRE reflects some of the defining attributes of standards and standardization (Bowker and Star 1999; Lampland and Star 2009; Timmermans and Epstein 2010) in that the racial and ethnic categories that individuals are asked to identify with are more or less made uniform through NIH regulations governing the conduct of studies supported by U.S. federal funding. And in a nominal way, the practice of self-identification itself has become a standard operating procedure—although as we describe previously, the particulars of implementation are by no means widely agreed upon. Thus SIRE constitutes two subtypes of standards (Timmermans and Berg 2003): both a terminological standard that attempts to stabilize the meaning of racial and ethnic categories over time and space, as well as a procedural standard that specifies how the practice of self-identification should be performed. And as Lampland and Star (2009) argue, standards like SIRE invoke ethics and values that are enormously consequential for the individuals implicated by them. In the following, we examine these questions by asking what kinds of meanings are mobilized in the use of SIRE.

SIRE as meaning

Previously, we described the meanings attached to SIRE as a technique that is “easy” and “convenient” and one that facilitates future, anticipated data recombinations. In addition, many of our scientists had a great deal to say about their understanding of SIRE as a crude, imprecise, and overly confounded entity:

I think if you equate the example [of] body mass index [to using SIRE], it's not necessarily the best measurement. It's just easy and it's important, and that's why people use it left and right.... If you ask me how to measure adiposity, BMI [body mass index] is not necessarily the best one.... By itself, it's not really a good measure, if you ask me. But it's important, to lots of purposes. And if I were to study obesity, I would measure BMI, but I'm also gonna try ... to measure, say, visceral fats and subcutaneous fat, and whatnot.

In this scientist's argument, SIRE, like the metric of body mass index, is easy to use to measure population differences but a blunt instrument that encompasses too many things. But with a little bit of investment and work, this investigator asserts, one could get to a better measure of obesity than body mass index. Similarly, she implies, GEI scientists ought to be able to get to a better measure of the genetic and environmental determinants that may be making up racial differences.

Many of our participants also took up this same call that SIRE was a stand-in for too many other factors:

Race is a convenient label to use for studies because you can get it easily. You can get people to answer these questions. But you know, we don't tend ... to think a lot about that variable, what it means, how it's defined, how it's being used. We just sort of use it blindly.... If we were laboratory scientists and we were developing assays, we wouldn't just use the assay somebody else came up with some other purpose. We would develop our own assays and test them and figure out what the coefficients of variability are and the properties of that assay before we applied them in experiments. And so I think that's what should be done ... with a variable like race is that we should develop metrics for what it is we actually want to learn, think about them very carefully, refine them, validate them, et cetera, and then use them in studies.

By the “blind” use of SIRE—a measure that someone else developed for other purposes and used only because it is “easy”—this scientist argues that we actually undermine our ability to “see” the determinants important for disease etiology. That is, even though many of our participants understood racial self-identification and social and cultural membership (and indeed, racial assignment by others as well²) to stand in for social experiences and environments that may affect health, the use of SIRE as a convenient proxy in fact served to truncate or preclude examination of those other factors. The participant quoted above also

²A few of our participants argued that in addition to *self*-identified race and ethnicity, racial ascription or assignment by *others*—that is, the race or ethnicity one is seen or perceived to belong to—can also be important for GEI research to attend to because it shapes one's interactions with other social actors and institutions, which can have implications for one's health.

draws upon a comparison to “laboratory science” as an implicitly superior, more accurate form of science as a means to critique the measure of SIRE.³

Thus the meanings ascribed to SIRE as a technique are that it is rote, a default, insufficient for the purposes for which it is being used, and unscientific in that investigators are using a technique of measurement that they did not themselves develop or validate. While understandable as an attempt to honor the very personal nature of racial and ethnic identity and the act of self-identification itself, SIRE is nonetheless seen as encompassing too much, as being always already, overly confounded and confounding.

What then about the meanings of the human subjects who are constructed and conjured up by the technique of SIRE? For the most part, scientists in our study expressed a desire to honor their subjects’ right to self-identify but asserted that the nature of the OMB/NIH categories—as a technique that over-lumped and under-specified human individuals—undermined this very commitment. One participant told us, from a more general perspective:

It’s been on my mind about how we do clump everybody together.... In general the studies that I have worked on don’t have a big enough sample size where we would even be able to get down to that level; but it does make us aware when we are interpreting our data that there is some measurement error.... It’s a mish-mash.

With respect to specific racial-ethnic groups, another scientist related:

The group called “white” comprises people from Europe, people from the Middle East, people from North Africa, I mean, a vast variety of people from different cultures. Of course, any one of these groups do comprise people from different countries, different regions, but the group “white” for me is most problematic in terms of the conglomerate of people in that group.

And yet another participant levied a similar critique of the category of Hispanic/Latino:

[For] a lot of the Dominicans, Puerto Ricans, the physicians reported them as African-American, and they reported themselves as Hispanic.... The real world is more heterogeneous than African Americans whose ancestors came over here a century or two ago. The real world is very heterogeneous, and it would be great if we could look at that and try to sort that out.... [We encountered] this very adamant “I’m Hispanic. I’m not Caucasian, I’m not African-American; I’m Hispanic.” But even Hispanic, there’re Native Americans, there’s people from Africa. So even when you think about Hispanic, to me that’s not a very representative category at all of anybody.... Who are these Hispanics? Are these Mexicans? Are they Dominicans? Puerto Ricans?

Our participants frequently invoked Barack Obama not only as an icon of American multi-raciality but also as an example of the inability of under-specified or over-lumped categories

³By characterizing the practice of categorical racial classification as bureaucratic and rote, this participant appears to be unaware of the extensive body of research and literature on such classification practices (e.g., Brodtkin 1999; Campbell and Rogalin 2006; Frank, Akresh, and Lu 2010; Harris and Sims 2002; Lee and Bean 2004). We did not hear from our scientist-participants that they had consulted this literature or scholars engaging in such research, despite their very frequent struggles to understand how best to handle racial identification and classification.

to capture an individual's unique ancestry, experiences, and identities. As one of our participants said, SIRE "is what I call a dirty trait.... Take Obama for example. His father is African and his mother is Caucasian.... People identify him as African American.... [But] actually, he lived with his mother [who was white] most of his life."

Thus, as a representational exercise in identification, SIRE falls far short of signifying and denoting the heterogeneity of the subjects it attempts to name. Scientists use SIRE as a means to respond to the twin calls for inclusion and interaction and to participate in the new sociogenomic paradigm in genetics (Bliss 2012). The capaciousness of self-identification in theory allows participants to self-identify as they would wish, and such self-identification can refer to a range of individuated meanings of racial difference, from the social, to the cultural, environmental, and economic, as well as the ancestral or genetic. However, in practice, SIRE relies on an infrastructurally embedded history of categories that are now understood to be unstable and under-determined, and this in turn undermines its own ability to represent human individuals faithfully. The imprecision of SIRE as a proxy for genetic differences is seen as a more or less technical problem, as being a "dirty," confounded, unscientific, and imprecise metric. But there is also moral critique here: SIRE, as a means to specify human subjects, immorally amalgamates those whose distinctiveness should be recognized. Such comments about the limits of SIRE demonstrate a tacit recognition of the fraught politics of identity in the United States and the failure of institutionalized, regulatory categories to reflect the more hybrid, shifting social groups with which people identify. In our study, scientists negotiate a moral imperative to represent human diversity with much greater accuracy and refinement through an appeal to the value of 'precision.' And this brings them to a consideration of genetic ancestry and its measurement via ancestry informative markers.

AIMs as technique

For many—but not necessarily most—of our participants, one way to represent human diversity with greater refinement is the use of ancestry informative markers (AIMs). AIMs are a set of genetic variations for a particular DNA sequence that appear in different frequencies in populations from different regions of the world. The use of AIMs compares an individual's genetic polymorphisms at these markers with previously analyzed genomic reference sets from people whose ancestral history is purportedly fairly well known. Many scholars (e.g., Fullwiley 2008) have pointed out the limitations of these reference sets, which are comprised of the genomes of relatively few sampled individuals who are themselves from a relative few, geographically conscribed populations that are considered to stand in for larger, presumably more heterogeneous populations spread over wider regions. Notwithstanding and indeed, eliding, this limitation, however, participants we interviewed and observed routinely used AIMs to estimate the geographical origins of an individual's ancestors, typically expressed as admixture, or the proportion of one's ancestry that comes from different continental regions.

For some of the scientists in our study, the use of AIMs is seen as a substantial shift in technique from SIRE, while for others, it is simply the next logical step to better ascertain some of the work that SIRE used to do for them. A fair number of our participants narrated

how they tried to move beyond just the crude measure of SIRE by asking a series of questions about birthplace and country of origin of parents and grandparents. For example:

People have certainly spent a lot of time trying to develop more complex questionnaires that might get down to ethnicity or country of origin or something like that as opposed to just race.... But it's hard to know what to make of it. If you say, well, my maternal grandmother was Irish, my maternal grandfather was German and Swedish, you know, you go on and on—what do you do with that information because it's so complex?

Given the complexity of this kind of self-reported information, AIMs represent a technique that greatly simplifies ancestry-related data. By producing a set of numbers—the proportions of genetic ancestry that originate from different continental regions, and sometimes sub-regions—AIMs not only distill complex family histories into numbers, but standardize them so that they are comparable and can be incorporated into statistical analyses.

All of our participants agreed that AIMs is a technical routine that is necessary in any research examining genetic determinants of disease, as a means to control for what is termed population stratification. Population stratification is the presence of systematic differences between groups in the frequencies of genetic variations, possibly due to differences in the ethnic admixture of those groups. Population stratification has long been a problem for genomic researchers: as they seek to establish that a statistical association between a genetic risk factor and a disease is, in fact, due to the factor itself, they must ensure that any systematic differences in the frequency of genetic traits between the populations being compared are controlled for. AIMs provide a way to do this, and thus have become part of conventional post-genomic science. As one researcher asked rhetorically,

Do we need to keep track of ethnic differences as a surrogate for population ancestry? Absolutely. And the reason for that is it is well-known that allele frequencies at many loci vary as you move across different ancient geographic regions.... [It's] garbage in, garbage out.... You put bad data into a computer program, it doesn't matter how good it is, it will give you bad information out.

That is, data taken from groups who may differ in ethnic admixture are, by definition, “garbage” and “bad data” unless those underlying differences have been controlled for; AIMs provide scientists a technical means to do this.

But at the same time, a small number of our participants disagreed that population stratification was always an issue in GEI research, and they argued therefore that AIMs are not universally useful. These quasi-dissenters saw AIMs as sometimes being an empty technical routine, devoid of meaning in some research situations, and instead performed because others (particularly funding agencies and grant and manuscript reviewers) mandated that it be done. One investigator told us:

Now I think that if you're proposing to do any study that's related to race, if you don't talk about ancestry informative markers, you're not going to get funded. Whether they really make a difference or not, I think that reviewers are thinking it's essential.... [For one grant] we had to submit it three times to get it funded.... In

the first review, it was, “Why aren’t you using ancestry informative markers?” And I said, “Well, we can categorize or classify African American enough [with SIRE] to say African-American women are more likely to be diagnosed when they’re young and to be diagnosed with aggressive breast cancer. We should be able to categorize people to look at tumor differences in the same way. But to make you happy, we’ll put these ancestry informative markers in.” ... A lot of stuff I think becomes kind of like fad and, you know, this is what you’ve got to do now.

For the most part, however, the scientists in our study perceived AIMs to be requisite and necessary to account for genetic variability. And in the process, the increasingly ritualized use of AIMs works to produce certain understandings of human difference and populations, as the next section explores.

AIMs as meaning

First, it is important to note that while our participants believe that AIMs are necessary to account for population stratification, this stratification was primarily a statistical control, intended mostly (or only) to adjust for possible associations between ethnic admixture and genetic variations. That is, our researcher-participants were equivocal in terms of whether AIMs in and of themselves were meaningful as markers of disease risk per se. Rather for a few, they left the door open for the possibility that variants related to ancestry, or markers found in one population and not in another, or that appear with differing frequency in different ancestral groups, could be associated with disease risk. But even so, they would vacillate, saying:

Association alone is not enough. Ultimately, you would need to understand ... the underlying biological mechanism. So for example, if you could say that a particular pattern of genomic markers was related to biological functions that were different among groups ... and that would be mechanistically associated or shown to be associated with disease, then obviously that would be the link. You would have to do some other basic science kind of studies.

Finding these kinds of links that connect race and ancestry to disease risk is our popular image of how a post-genomic focus on difference manifests in health research. However, the GEI researchers in our study were markedly tepid about the direct connections to be made between race, ancestry, and disease risk.

In this regard, the human bodies summoned by AIMs as a technique are not bodies that are differentially susceptible to diseases based upon ancestry. Rather they are bodies that may differ genetically in *other* ways that require controlling for population stratification, but that may not be related to disease disparities in and of themselves. Instead, the kinds of human subjects called forth by AIMs are admixed bodies, in a way that seems particularly (though not exclusively) American. This American particularism can be seen in the comparisons made between doing research in the U.S. versus in locales with much more homogeneous populations. As one scientist told us:

In Finland, which is a very homogeneous population ... I’m not going to have this problem, particularly if I say, “Were all four of your grandparents from Finland?”

At that point I will have zero ethnic admixture ... Finnish ancestry from back to the time of the Plague! I'll have a very homogeneous population, and I can just compare those cases and controls and it will be beautiful. But in the U.S., come on. We're the mongrels of the world. We are the population that is the admixture of all admixture. Seriously. I mean just my own family. My God, I couldn't tell you what I am. I am the biggest little mutt in the world. I know that I have ancestors from all over Europe, but also my family's been here so long I could not possibly tell you all of our ancestors. It only gets more complex as we move on down the next 20, 30, 40 years.

Thus AIMS are seen to offer GEI scientists the opportunity to not only do better science, but also to more completely capture all the heterogeneity and variability that exists within a singular person. They allow scientists, as one participant put it, to “de-mystif[y] the self-identification of race for us ... and mak[e] sure we understand what the makeup is of the participant we have.” The technique of AIMS represents the ability to more comprehensively characterize the ancestral complexity of contemporary individuals *as* individuals. In so doing, they infuse human subjects' bodies with the provocative valences and weighted language of “clean” and “beautiful”, and of “mutts” and “mongrels.”

In addition, the rhetoric of a future of ever-increasing admixture was a very common theme. Scientists we interviewed and observed spend a good deal of time anticipating the “American mix,” often refracted through personal stories of admixture in their own families. For example:

I showed this picture of my niece and her husband and their ... three kids. The three kids are between six genomes, right? And of the six genomes ... two are of European origin, two are of South Asian origin, and two are of African origin.... There's going to be more and more of this. More families are going to be mixed.... More and more people whose ancestors within them are going to ... have ancestry from different portions of the world.

This anticipated future is then one that invests AIMS with an ever-increasing mandatory nature. As one scientist told us, “I think anyone in the future generations and coming up are more mixed race, that you can't capture [with SIRE].... I've picked up that message that as we go forward, the demographics of the nation are changing. So we need to anticipate doing [AIMs].” And in the words of another:

There's much less barriers between ... the ethnic groups in our country as there used to be, which as far as I'm concerned is a great thing socially. It does make a geneticist's job harder.... What we then do is look at our mixed families and say, all right, what [AIMs] do we really need to use for them?

Thus AIMS as a technique becomes increasingly more meaningful as multiraciality becomes a larger issue; indeed, this *future* mixture creates an imperative to study ancestry and admixture *now*. In this future, human subjects are seen as individual unique recombinations of the genomes of their heterogeneous ancestors, potentially from far-flung places. And while some of these ancestral genomes can be labeled as coming from a particular place and of a particular biogeographical nature—“of European origin ... of South Asian origin” and

so on—increasingly their present-day descendants cannot be so labeled. Instead their genomes can now be thought of as mosaics of various origins: a stretch of DNA coming from a distant European ancestor, another coming from a Native American one, and yet another from an African ancestor. If nothing else, our scientists reiterate, the examples of African Americans and Latinos as admixed populations show us the ultimate fallacy of group identity, at least in the genetic context.

DISCUSSION AND REFLECTIONS: Reframing the Question?

As Abu El-Haj (2007) points out, by definition AIMs and admixture presume race-mixing and she therefore asserts, following Michaels (2000), that to choose between “pure” or “hybrid” race is still to have already chosen race itself. In turn, she argues that AIMs are purported to provide biologically or historically useful information in calculating disease risk for a particular population. Among our participants, however, they are pointedly circumspect about any straightforward connections to be made between AIMs, genetic ancestry, and disease risk (see also Shim et al. 2014). Nonetheless, despite their questionable utility as clues to differential health, AIMs comprise a highly routinized and increasingly institutionalized practice among the scientists in our study. In large part their routinization is a product of their signification as a means to deal with the current and future American mix; we therefore argue that AIMs constitute a practice of anticipation, preparedness, and possibility (Adams, Murphy, and Clarke 2009), as well as an investment in collecting data that will be useable and useful in the future. Simultaneously, AIMs are also seen as a practice of more specific, and therefore more scientifically and socially responsible, representation.

Thus we find that AIMs are understood as more accurate and precise measures of unique ancestral recombinations and fractional identities (Lee 2013) than SIRE, which are seen as blunt and crude categories that over-emphasize the differences between and under-represent the heterogeneity within them. By evoking attributions of precision and accuracy, the valorization of AIMs relative to SIRE invokes motifs of what Lee (2008) characterizes as “racial realism and responsibility.” In our case, we find that these themes refer to the manifold ideas that ancestral differences are “real” (as can be seen in the differing frequencies of AIMs by continental ancestry), that we need to get “real” about such human differences (rather than a resolute and misguided colorblindness), and that such realism is the responsible path for post-genomic scientists to take (see also Bliss 2012).

This juxtaposition of the relative utility of AIMs versus SIRE, we argue, serves to elide important debates about the relationships of race, ancestry, genes, environment, and health. That is, the ways in which the perceived precision of AIMs resonates with ideas about “racial realism” and “responsibility” are critical moves in reframing the question, from thinking about how genes and environment interact, to how best to classify individuals. Such reframing is highly consequential, as can be seen in two exchanges we observed during the course of our fieldwork at a scientific conference.

The first took place during the question-and-answer portion of a panel, in which one audience member spoke about her conversations with medical professionals about health

disparities between black and white women. She reported that hospital executives “brush[ed] aside” more fundamental problems with the “financial underpinnings of the health care system” as potentially playing a role in producing such disparities, and instead pointed her to genetic studies that attempt to document biological differences. One panelist responded to her comment as follows:

This is not a personal criticism, but if we are going to talk among disciplines, the term “Caucasian” is genetically undefined. It is an old, racist term, and where would you classify Afghani? They are in the middle of Asia. They have light skin. They’re not East Asians, they’re not Europeans, are they Caucasian? Well, I don’t use that term. They are Afghani. And I think we have to be very careful about use of terms, and the medical literature uses “Caucasian” pervasively, and when I get papers to review from *New England Journal of Medicine* or *JAMA*, they’re usually about population genetics, and I have a standard paragraph that goes in virtually every review, that you are using undefined, indefinable terms.

The second exchange occurred later at the same conference, when we engaged a poster presenter in conversation about his findings on a particular type of cancer in an Asian sub-population. In explaining how he generated the reference group, he related his problems discerning the sameness or difference of research subjects from different studies who were variously described as “white,” “European,” and other labels. He went on to remark that rather than the Census, “they should call it the non-sense-sus,” as he found the U.S. Census categories to be nonsensical conceptually, methodologically, and scientifically.

In these two vignettes from the field, we saw how a larger conversation about racial disparities in health was contracted into a discussion of race and ancestry in genomic science, and in turn to questions of how to properly classify people (as with Afghani outliers), and how to properly and consistently label them (without the variability of “white” sometimes, or “Caucasian” other times). By characterizing the category of “Caucasian” as illegitimate and “racist” because it is “genetically undefined,” the panelist in the first exchange described above questioned the integrity and legitimacy of standardized racial categories and labels. The emphasis on the “old racist” and “non-sense-sus” categories in both vignettes narrowed the focus of the discussion, and of the articulation of the problem, to the meaninglessness of racial classifications and to the lack of standardization in how they are being used. Given that in the first vignette, the audience member’s question was raised within the context of the racial politics of health care resource allocation, and the study at the center of the second vignette dealt with cancer disparities, this narrowed focus is a particularly telling move.

We suggest that these critiques of SIRE and the valorization of AIMs in GEI research may work to serve the objectives of inclusion, but not those of interaction, to return to the two points of reference with which we opened this paper. Rather than putting the social and biological on more equal footing, the understandings of SIRE and AIMs and their relative utility instead reinforce a ‘more-and-less’ focus, with a primary emphasis on the biological combined with supplemental or subsidiary attention to the social. By characterizing SIRE as inexact, crude, and ill-defined, GEI scientists both acknowledge the social, cultural, economic, and environmental heterogeneity of populations at the same time that they also

undermine the practice of using SIRE as a means to explore that very heterogeneity. That is, the use of SIRE had previously opened the door to finding racial and ethnic differences in disease risks, that in turn allowed scientists, at least in theory, to then examine social and environmental as well as biological influences on disease. Yet the persistent and continued use of SIRE as only a proxy, and the simultaneous use of AIMs as a necessary technique that says something about ‘the American mix’ but not about American inequality, enable researchers to avoid the conceptually and methodologically more challenging task of measuring environmental, cultural, and social determinants in and of themselves.

We suggest that conversations like those we encountered in our fieldwork alert us to the possibility that an acknowledgement of racial specificity, in the name of better representation and precision (both of which are hybrid social-political-scientific goals), becomes a way that post-genomic science feels it can account for population differences and health disparities. The turn to classification as ‘the problem,’ as we describe, obscures other ways of understanding how difference matters. Scientific talk critiquing classification and categorization practices signals an empathy and sensitivity to the complexity of self-identification and inter- and intra-group differences, but in ways that seem to elide group-based social processes and practices that affect and produce health and disease. Abolishing “racist categories” through careful language and more precise genetic distinctions both sidesteps and distracts from the recognition that racial categorizations are embedded within institutions and processes that sustain configurations of power and reproduce racial inequalities in health.

CONCLUSION

To conclude, our scientist-participants see SIRE as a technique that makes convenient yet ultimately erroneous claims about the identities and nature of the human subjects it characterizes, representing them as categorizable and standardizable. In contrast, AIMs are seen by many as “just” technique—a necessary methodological step to deal with the problem of population stratification, but not necessarily in and of itself making meaning of human subjects. However, as we argue, the technique of AIMs in fact does invest subjects with meanings, as highly unique genetic recombinations of an ever-widening variety of ancestors. Therefore, both technically and symbolically, AIMs centrally concern the meanings of ‘American mixtures.’ While our participants frame this as an acutely American issue (in alignment with characterizations of race as a peculiarly American obsession), GEI scientists and their research indicate that it is not an exclusively American one. Indeed population stratification must be addressed in all genomic analyses, and moreover, U.S.-based conceptions of race both are increasingly internationalized but also bump up against more local understandings in the process.⁴

In our interviews and observations of scientists conducting gene-environment interaction research, one refrain seems to be a constant: that the first iteration of the inclusion-and-difference paradigm (Epstein 2007)—in which under-represented groups were brought into the research fold and dimensions of difference accounted for—does not go far enough. Too

⁴See, for example, the July 2014 special issue of *Science, Technology, & Human Values*.

often investigators found themselves struggling with the challenges of under-specification and over-lumping, of not knowing how to properly characterize the unique individuals that comprise their samples. What mitigates against scientific progress on this point is in part the constant pressure to maximize the size of study samples—to get big data—to facilitate subgroup analyses and claims-making about subpopulations. Thus, scientists constantly weighed the potential upsides to precision by splitting against the known downsides to statistical power; the upsides to sample size by lumping versus the downsides to accuracy and nuance. The calculus and stakes of these tradeoffs are particularly pressing for GEI researchers for at least two reasons. First, many believe that widespread debates about genetic studies' failure to replicate their findings or to explain much of disease etiology (e.g., Editors 2010; Goldstein 2009; Hardy and Singleton 2009; Hirschhorn 2009; Visscher et al. 2012) can be resolved by research that accounts for gene-environment interactions. And second, their commitment to looking at the causal continuum, from the genetic to the social and in between, compels them towards precision in measurement of as many determinants as possible in as many subjects as possible, goals that are all too often at odds with one another.

What is striking to us is that the shift in value from thinking about SIRE as encompassing a wide range of social, behavioral, environmental, and possibly biological determinants of disease, to the accuracy of describing human diversity through AIMs, seems to be emblematic of a larger shift from population health to precision medicine. That is, population health—as expressive of the value of studying disease at the population level in and of itself—may be giving way to the idea that prevention, diagnostics, and treatment should be more precise and customized to the individual. Although many of our participants profess a commitment to intervene to promote health at the population level, the ways in which they now seek to do so seem to be underwritten by logics and techniques that are much more concerned with refinement, accuracy, and exactness that go beyond the category, the group, the population. This is what GEI scientists perceive to be necessary in order to be able to better account for the vastly complex processes of disease etiology. Yet as several of our participants have pointed out, as a society, we have invested in learning how to precisely parse some kinds of complexity, and not others. As importantly, even while there are multiple lines of active scientific inquiry that analyze human difference, some approaches (e.g., genomic) are attended to while others (e.g., sociological, epidemiologic) less so (see note 3). This differential attention to existing research and differential investment in new research both reflect and reinforce our thinking and priorities about what dimensions of difference require greater precision and which should and *can* be attended to first. More specifically, while *some* of our participants strive to collect ever-more precise data on cultural factors, habits, behaviors, and exposures, *all* of our participants use AIMs as a means to capture individual ancestral variability.

We heartily agree with our participants' concerns about what kinds of precisions are favored first and thought to be amenable to science—versus those that are seen as recalcitrant to measurement and/or too complex. But we would also add that the imperative of and faith in precision is giving rise to practices that explicitly and implicitly redefine race and ancestry in ways that preclude important claims that categorical race can mobilize. Specifically, we

suggest that the concentrated attention to and investments in the measurement and definition of differences in individual ancestry, even from a GEI perspective, can obscure alternate politics and scientific knowledge production about the connections to be made between race and health inequalities.

The kind of difference-sustaining that is going on in GEI research constitutes one thread in a larger landscape that is moving away from group-based niche standardization (Epstein 2007) towards personalized and precision medicine, big data, and data on, specific to, and for everyone. This landscape gives the appearance of—and is probably also motivated by the values of—democratization, inclusion, and better care and health for all. Yet, as we, along with many other others, have argued, the kinds of data and information sought in this age of inclusion demand greater precision along some dimensions (e.g., genetic ancestry, genetic disease determinants) and seem to ignore it along others (e.g., racial self-identification, environmental disease determinants). As importantly, this focus on accuracy and precision functions to efface social situations and dynamics that work not solely at the individual level, but at the group and population level. The reasons for these omissions are both deep and often hidden; how they work and what their consequences may be demand our continued attention and interrogation.

Acknowledgments

Our thanks go first to our study participants, for the time and attention they gave to answer our many questions. We also greatly appreciate the Editor, Gilbert Gee, and the anonymous reviewers of *JHSB* for their very thoughtful comments. Research reported in this publication was supported by the National Human Genome Research Institute's Ethical, Legal, and Social Implications Research Program of the National Institutes of Health under award number R01HG005848. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- Abu El-Haj, Nadia. The Genetic Reinscription of Race. *Annual Review of Anthropology*. 2007; 86:283–300.
- Adams, Vincanne; Murphy, Michelle; Clarke, Adele E. Anticipation: Technoscience, Life, Affect, Temporality. *Subjectivity*. 2009; 28:246–265.
- Bliss, Catherine. Racial Taxonomy in Genomics. *Social Science & Medicine*. 2011; 73(7):1019–1127. [PubMed: 21840637]
- Bliss, Catherine. *Race Decoded: The Genomic Fight for Social Justice*. Stanford, CA: Stanford University Press; 2012.
- Bowker, Geoffrey C.; Star, Susan Leigh. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press; 1999.
- Braun, Lundy; Hammonds, Evelyn. Race, Population, and Genomics: Africa as Laboratory. *Social Science & Medicine*. 2008; 67(10):1580–1588. [PubMed: 18755531]
- Brodkin, Karen. *How Jews Became White Folks and What That Says About Race in America*. New Brunswick, NJ: Rutgers University Press; 1999.
- Campbell, Mary E.; Rogalin, Christabel L. Categorical Imperatives: The Interaction of Latino and Racial Identification. *Social Science Quarterly*. 2006; 87(5):1030–1052.
- Cavalli-Sforza, L Luca; Bodmer, Walter. *The Genetics of Human Populations*. New York: Dober Publications; 1974/1999.
- Collins, Francis S.; Green, Eric D.; Guttmacher, Alan E.; Guyer, Mark S. on behalf of the US NHGRI. A Vision for the Future of Genomics Research: A Blueprint for the Genomic Era. *Nature*. 2003; 422(6934):835–847. [PubMed: 12695777]

- Dobzhansky, Theodosius. A Debateable Account of the Origin of Races. *Scientific American*. 1963; 208:169–172.
- Editors. On Beyond GWAS. *Nature Genetics*. 2010; 42:551. [PubMed: 20581872]
- Epstein, Steven. *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press; 2007.
- Frank, Reanne; Akresh, Ilana Redstone; Lu, Bo. Latino Immigrants and the US Racial Order: How and Where Do They Fit In? *American Sociological Review*. 2010; 75(3):378–401.
- Fujimura, Joan H.; Rajagopalan, Ramya. Different Differences: The Use of ‘Genetic Ancestry’ Versus Race in Biomedical Human Genetic Research. *Social Studies of Science*. 2011; 41(1):5–30. [PubMed: 21553638]
- Fullwiley, Duana. The Molecularization of Race: Institutionalizing Human Difference in Pharmacogenetics Practice. *Science as Culture*. 2007; 16(1):1–30.
- Fullwiley, Duana. The Biological Construction of Race: ‘Admixture’ Technology and the New Genetic Medicine. *Social Studies of Science*. 2008; 38(5):695–735. [PubMed: 19227818]
- Fullwiley, Duana. *The Enculturated Gene: Sickle Cell Health Politics and Biological Difference in West Africa*. Princeton, NJ: Princeton University Press; 2011.
- Gannett, Lisa. Racism and Human Genome Diversity Research: The Ethical Limits of ‘Population Thinking’. *Philosophy of Science*. 2001; 68(suppl 3):S479–S492. [PubMed: 15948338]
- Goldstein, David B. Common Genetic Variation and Human Traits. *New England Journal of Medicine*. 2009; 360(17):1696–1698. [PubMed: 19369660]
- Hardy, John; Singleton, Andrew. Genomewide Association Studies and Human Disease. *New England Journal of Medicine*. 2009; 360(17):1759–1768. [PubMed: 19369657]
- Harris, David R.; Sims, Jeremiah Joseph. Who Is Multiracial? Assessing the Complexity of Lived Race. *American Sociological Review*. 2002; 67(4):614–627.
- Hirschhorn, Joel N. Genomewide Association Studies — Illuminating Biologic Pathways. *New England Journal of Medicine*. 2009; 360(17):1699–1701. [PubMed: 19369661]
- Lampland, Martha; Star, Susan Leigh, editors. *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca, NY: Cornell University Press; 2009.
- Lee, Jennifer; Bean, Frank D. America’s Changing Color Lines: Immigration, Race/Ethnicity, and Multiracial Identification. *Annual Review of Sociology*. 2004; 30:221–242.
- Lee, Sandra Soo-Jin. Racializing Drug Design: Implications of Pharmacogenomics for Health Disparities. *American Journal of Public Health*. 2005; 95(12):2133–2138. [PubMed: 16257939]
- Lee, Sandra Soo-Jin. Biobanks of a ‘Racial Kind’: Mining for Difference in the New Genetics. *Patterns of Prejudice*. 2006; 40(4–5):443–460.
- Lee, Sandra Soo-Jin. Racial Realism and the Discourse of Responsibility for Health Disparities in a Genomic Age. In: Koenig, BA.; Lee, SS-J.; Richardson, SS., editors. *Revisiting Race in a Genomic Age*. New Brunswick, NJ: Rutgers University Press; 2008. p. 342–358.
- Lee, Sandra Soo-Jin. Race, Risk, and Recreation in Personal Genomics: The Limits of Play. *Medical Anthropology Quarterly*. 2013; 27(4):550–569. [PubMed: 24214161]
- Lee, Sandra Soo-Jin; Mountain, Joanna; Koenig, Barbara; Altman, Russ; Brown, Melissa; Camarillo, Albert; Cavalli-Sforza, Luca; Cho, Mildred; Eberhardt, Jennifer; Feldman, Marcus; Ford, Richard; Greely, Henry; King, Roy; Markus, Hazel; Satz, Debra; Snipp, Matthew; Steele, Claude; Underhill, Peter. The Ethics of Characterizing Difference: Guiding Principles on Using Racial Categories in Human Genetics. *Genome Biology*. 2008; 9:404. [PubMed: 18638359]
- Lewontin, Richard. The Apportionment of Human Diversity. *Evolutionary Biology*. 1972; 6:381–398.
- Manolio, Teri; Bailey-Wilson, Joan E.; Collins, Francis S. Genes, Environment and the Value of Prospective Cohort Studies. *Nature Reviews Genetics*. 2006; 7(10):812–820.
- Michaels, Walter Benn. *Political Science Fictions*. *New Literary History*. 2000; 31:649–664.
- Montoya, Michael. Bioethnic Conscription: Genes, Race, and Mexicana/o Ethnicity in Diabetes Research. *Cultural Anthropology*. 2007; 22(1):94–128.
- Montoya, Michael. *Making the Mexican Diabetic: Race, Science, and the Genetics of Inequality*. Berkeley, CA: University of California Press; 2011.

- Murphy, Michelle. *Seizing the Means of Reproduction: Entanglements of Feminism, Health, and Technoscience*. Durham, NC: Duke University Press; 2012.
- Reardon, Jenny. *Race to the Finish: Identity and Governance in an Age of Genomics*. Princeton, NJ: Princeton University Press; 2005.
- Reardon, Jenny. *Finding Oprah's Roots, Losing the World: Beyond the Liberal Anti-Racist Genome*. Presented at the Annual Meeting of the Society for Social Studies of Science; Washington, DC. 2009.
- Shim, Janet K.; Darling, Katherine Weatherford; Lappe, Martine D.; Thomson, L Katherine; Lee, Sandra Soo-Jin; Hiatt, Robert A.; Ackerman, Sara L. Homogeneity and Heterogeneity as Situational Properties: Producing—and Moving Beyond?—Race in Post-Genomic Science. *Social Studies of Science*. 2014; 44(4):579–599. [PubMed: 25272613]
- Timmermans, Stefan; Berg, Marc. *The Gold Standard: The Challenge of Evidence-Based Medicine*. Philadelphia, PA: Temple University Press; 2003.
- Timmermans, Stefan; Epstein, Steven. A World of Standards But Not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology*. 2010; 36:69–89.
- Visscher, Peter M.; Brown, Matthew A.; McCarthy, Mark I.; Yang, Jian. Five Years of GWAS Discovery. *American Journal of Human Genetics*. 2012; 90(1):7–24. [PubMed: 22243964]
- Whitmarsh, Ian. *Biomedical Ambiguity: Race, Asthma, and the Contested Meaning of Genetic Research in the Caribbean*. Ithaca, NY: Cornell University Press; 2008.

Biographies

Janet K. Shim is associate professor of sociology in the Department of Social and Behavioral Sciences, University of California, San Francisco. Her current research focuses on analyzing the science of health disparities and the production of health care inequalities. She is the author of *Heart-Sick: The Politics of Risk, Inequality, and Heart Disease* (2014, NYU Press).

Katherine Weatherford Darling is a doctoral candidate in sociology and Rosenberg-Hill Graduate Research Fellow at University of California, San Francisco. Her dissertation is an ethnography tracing the uses of health information technologies within clinical and policy strategies for managing human immunodeficiency virus (HIV) as a chronic illness in US health care.

Sara L. Ackerman is a medical anthropologist and assistant professor in the Department of Social and Behavioral Sciences, University of California, San Francisco. Her research interests include the uptake and sociocultural implications of new medical technologies, including genomic testing in research and clinical settings.

Robert A. Hiatt is professor and chair of the Department of Epidemiology and Biostatistics, University of California, San Francisco, and associate director for Population Science of the UCSF Comprehensive Cancer Center. His research interests include cancer epidemiology, health services and outcomes research, and the social determinants of cancer.

Sandra Soo-Jin Lee is a medical anthropologist and senior research scholar at the Stanford Center for Biomedical Ethics, and faculty in the Program in Science, Technology and Society at Stanford University. Dr Lee's research interests are in the sociocultural and ethical dimensions of emerging genomic technologies and their translation into biomedical practice.