2

A Colorful Explanation

Promoting Genomic Research Diversity Is Compatible with Racial Social Constructionism

Tina Rulli

This chapter explores the possible tension between the call for more diversity in genomic research and the view that races are socially constructed and not biologically real. Does the claim that we need more diversity in genomic research, often understood in racial terms, rely upon an explicit commitment to biological race realism?

Proponents of genomic medicine hope to employ associations between gene variants and disease states and drug metabolism to predict, diagnose, or treat disease in individuals through genetic testing, including in preimplantation genetic diagnosis and prenatal screenings, and to develop targeted gene therapies or interventions. Genomic medicine relies upon genome-wide association studies (GWAS), where individual genomic samples are assessed and compared for patterned associations between known gene variants and disease states or drug responses. The targets of GWAS are usually complex diseases, those associated with multiple genes. Since the effects of each gene may be tiny, GWAS requires databases of genomic samples from a very large number of individuals to sufficiently power the associations. Currently, however, individuals of primarily European descent are vastly overrepresented in GWAS. The GWAS Diversity Monitor, which tracks real-time diversity statistics for participants, reports that 95.05 percent of participants are of European descent, with slightly more than 3 percent of Asian descent.[1]

There is a widespread call to racially and ethnically diversify genomic research.[2] Proponents of diversification claim that population diversity—often described at the continental level, echoing familiar continental conceptions of race—is needed to ensure the accuracy of genomic medicine and to extend the benefits of genomic medicine to all people.

The call for racial diversity in genomic research might imply that race must be biologically real, that race is encoded at the genetic level. Why else would racial inclusion be important in genomic research? The call for diversity may also seem to imply that differently racialized people have different genes. But neither claim is true. Here, I argue that racial diversity efforts in genomic research are compatible with the denial of a biological reality for race and compatible with social constructionism about race. Thus, genomic researchers advocating for racial diversity in genomic research need not be committed to or seen as advocating for the view that racial categories are biologically real.

A few disclaimers at the outset. I am not advocating for genomic medicine. The majority of race-based differences in disease have socioenvironmental explanations.[3] Nor do I think increasing racial diversity is the *best* way to go about increasing genomic diversity. The use of genetic similarity, a continuous measure based on genes themselves, would better ensure representation of human population diversity. But if we take geneticists at their word—that genomic medicine will bear fruit—it is incumbent upon us that these putative benefits be equitably distributed. The calls for inclusivity in genomic research often take the form of racial diversity. I argue that it is not incoherent to advocate for racial diversity in genomic research and to embrace the dominant, most defensible view of what racial categories are, the social constructionist view. That is, one is mistaken if one sees these calls for racial diversity as requiring the truth of biological race realism. Instead, a call for racially diversifying genomic research can be a practical strategy in the just allocation of benefits across diverse people, even among social constructionists about race.

In what follows, I will center the US conception of race, which identifies five races pertaining to five continents. In this conception, the categories are white (European descent), Black (African descent), Asian, Pacific Islander, and Indigenous American.[4] This continental race-based classification is widely adopted by geneticists and invoked even when not talking directly about race—for example, when making population or ancestral group assignments for people.

I face a difficulty in citing studies that use population descriptors, referring to *race, descent, genetic ancestry*, or continental level *populations*. There is a lack of consistency among scientists in the use of these terms, and, further, these groupings are typically given at the continental level, reifying the idea that there are meaningful biological groupings that map onto our conventional notion of race. But it is this very idea that I am arguing against here. Recent, prominent efforts have been made to scrutinize descent-based descriptors and to render their usage more consistent, intentional, and transparent. In 2023 the National Academies of Sciences, Engineering, and Medicine (NASEM) issued a report whose mission is to clarify the use of group labels for individual research participants in scientific studies out of concern with the unstandardized, unscientific use of racial or ethnic categories in population descriptors.[5] The committee does not recommend terms

of use; rather, it outlines a shared approach to the use of population descriptors in accordance with the principles of respect, beneficence, equity, justice, and transparency, among other values. This chapter is part of a critical literature on these race-based concepts. In citing or referring to studies, I am not advocating for the use of these race-based terms. But were I to change the nomenclature these studies employ, I might change their intended meaning, whatever it is. Thus, I have opted to report in the terms they use.

Indeed, the NASEM report's first recommendation is that "race should not be used as a proxy for human genetic variation. In particular, researchers should not assign genetic ancestry group labels to individuals or sets of individuals based on their race, whether self-identified or not."[6] However, I argue that *racial diversity* can be a proxy for genomic diversity. This may seem at odds with the NASEM recommendation. To the contrary, I see this chapter as addressing a pressing question and need that lingers over their recommendation. Race itself is not a proxy for the genotype of individuals because races are not biologically real. But we do need racial diversity in genomic research. I doubt the authors of the NASEM report would deny that. Thus, the inevitable question I raise here about how to square the call for diversity in genomic research with social constructionism about race needs addressing. I believe that clarity on this very limited way in which racial diversity can be helpful to genomic science and medicine, with extensive clarification on the limits of race's usefulness as a proxy, advances the same goals as the report. Race and racialized genetic ancestry themselves are not proxies for an individual's underlying genotype. But racial diversity in genomic research is needed to justly extend the putative benefits of genomic medicine to all.

In the first section of this chapter, I further discuss the importance of genetic diversity in genomic research and the call for racial diversity. In the second section, I explain the different conceptions of race: biological race realism, statistical race realism, and social constructionism. I explain why social constructionism is the most defensible conception of race and proceed through the rest of the chapter on the assumption that it is the correct view. Yet I will show how racial diversity in GWAS can be a proxy for genomic diversity, broadly speaking, even if race is a socially constructed category. In the third section, I use a novel analogy to do so. In the fourth section, I caution against the use of race as a proxy for individuals' genotypes in the clinical setting. Thus, even if race can be useful in promoting genomic diversity, its use as a proxy is quite limited and specific.

NEED FOR DIVERSITY IN GWAS

There is a need for genomic data that come from a diverse range of people. We cannot accurately extrapolate findings about gene variants and disease traits or drug responses from one population to another.

Populations are, roughly, interacting, interbreeding groups of individuals cooperating for survival.[7] Populations themselves are scientific constructs, not biological entities, that scientists posit for research purposes. Race and population are not interchangeable concepts. Change who you interact with, and you change your population. But this is not true of race.[8] Nonetheless, scientists frequently racialize populations, describing groups of people at the continental level because this level of grouping is familiar to and precedes population genetics. The definition of population does not preclude the possibility of interracial populations, obviously, but many genetic scientists construe populations along racialized lines in order to ensure roughly (what they think is) homogeneous ancestry among individuals within populations, an issue that will be discussed at greater length in the chapter by Carlos Andrés Barragán, Sivan Yair, and James Griesemer and in the chapter by Lisa Ikemoto. Predictions based on associations in one population may give rise to false positives in another population.[9]

This is for several reasons.[10] Allele frequencies vary among people by geography. When looking for medically relevant variants, geneticists compare those who exhibit the disease in question to controls who do not. If studies use people from different, geographically circumscribed populations, there is a risk of confounding alleles that vary among individuals due to a difference in ancestry with those that are associated with the disease in question. Controlling for population is meant to eliminate this confound. Additionally, a GWAS identifies *associations* between genes and traits, not the genetic *causes* of the trait. Given that alleles vary among populations, an identified marker of a trait may be linked to both common and rare alleles that cause the trait. The rare alleles may be frequent in some populations but not in others. Thus, a marker that is accurate in one population (where the rare allele is present) may give rise to a false positive in another population (where the allele is not present).

Another reason population diversity in genomic research is important is that scientists predict that rare variants (those that occur in less than 5 percent of the world population) will be more informative in predicting disease occurrence and drug response. Rare variants are often specific to populations.[11] These variants may be uncommon among people of European descent but present among other groups. Without genomic diversity, we are presumably failing to find many such rare variants.

This failure is especially acute because modern humans evolved in Africa. Some humans migrated out of Africa and populated the rest of the world. But these small, migrating groups carried with them only a subset of the genetic diversity that remained within Africa. Due to this genetic "bottleneck" and the fact that humans have been in Africa the longest, there is more genetic diversity among people with African ancestry than in other ancestral groups.[12] The exclusion of people from the African continent in genomic research poses an opportunity cost in identifying meaningful variants. For example, the discovery of *PCSK9*

variants in people of recent African descent, which lower cholesterol in other ancestral groups, resulted in the successful development of the drug evolocumab, considered "the most important trial result of a cholesterol-lowering drug in over 20 years."[13] What other such discoveries are we failing to make for lack of diversity in genomic research?

This also signals a problem with geneticists' habit of using racialized population designations corresponding to continental-level populations. Given that there is much more genetic diversity in the "African population" than in other continental populations, lack of African diversity in genomic research may result in weaker associations between genetic markers and gene variants in African populations compared to European populations.[14]

The underrepresentation of certain non-white people in the data already entails a health-care disadvantage. Popejoy and Fullerton report that individuals of African and Asian ancestry—those often racialized as Black and Asian, respectively—more frequently receive nondefinitive test results or have variants of unknown significance.[15] Without racial and ethnic diversity in genomic research, those who are already underserved in the medical community—historically oppressed racial and ethnic minorities—will be further disadvantaged by a genomic medicine that does not include them.[16] For these reasons, many geneticists have called for racial diversification of GWAS to ensure the future potential benefits of genomic medicine apply to all.

WHY RACE IS NOT BIOLOGICALLY REAL

The need for genomic diversity, often construed as racial diversity, in genomic research may suggest to some that racial differences are reflected at the genomic level. The view that race has a genetic basis is a kind of biological race realism. Biological race realism is the view that race is a meaningful biological category that distinguishes differently racialized individuals on a biological or genetic level. Biological race realism is the conventional and perhaps common lay view of race. It is commonly held by scientists and physicians as well. In its original and crudest form, it is essentialist; it assumes that race is grounded in some biological essence—perhaps phenotype (e.g., physical features) or genotype (e.g., race-related genes)—that is inherited. In this view, race is discrete, meaning all of the people within one race share the essential features, while all of those outside the race lack these essential features. In this view, there are mixed-race people. But even this idea implies that there are "pure" racial groups that can then be blended, a widespread but mistaken belief that will be taken up in the chapter by Lisa Ikemoto.

This crude race realism has been widely dismissed by social scientists and philosophers. There are no biological features that comprise a discrete racial essence. Populations that correspond to the large continental groupings do not vary from

one another in stark, discrete ways. Rather, phenotype and genotype among and within these large populations vary gradually—that is, *clinally*. We perceive there to be drastic and discrete morphological differences between groups of people (and thus infer discrete genetic differences) only when we compare individuals in (or with ancestors from) locations that are geographically distant from one another. If we look at people in (or with ancestors from) the places in between, we see gradual transitions in phenotype and genotype.

Crude race realism is obviously false. It is now being replaced by a *statistical race realism* in genetics. Some scientists and philosophers emphasize that, while there are no discrete populations corresponding to our common racial categories, there is structure to clinal genomic data.[17] The claim is that groups of individuals can be identified by genetic clustering among them—some individuals share distinctive groupings of genomic variants called haplotypes—that statistically correlates to having recent ancestry from particular geographic regions that are roughly continental. Perhaps these genetic clusters signify races.

But this new statistical view of race faces many criticisms. Some of the concerns are methodological. The clusters may be the artifacts of sampling strategies—for instance, using predefined populations that bias the data to produce racialized outputs; using small sample sizes for large, diverse geographic regions; preserving geographic distance between samples, which makes clinal differences look larger.[18] Further, generating statistically meaningful genetic clustering of populations that correspond to the familiar racialized continental-level groupings requires scientists to choose a number of clusters that reflects our race realist conception of the races.[19] The choice is arbitrary. Instruct the computer program to generate a large number, and you end up with 50 races, for instance, rather than the conventional 5. But this example shows that achieving an output that corresponds to our continental conception of race is possible only through human intervention in the data. In other words, these genomic clusters do not emerge from the data but are imposed on them.

There are many other concerns with statistical race realism.[20] But it will suffice to say here, even setting those important worries aside, that this statistical conception of *race* is far too revisionary to warrant the name. Races were originally theorized to be discrete, essentialist, and hierarchical. This new conception of race as continental genetic clusters is clinal, nonessentialist, and nonhierarchical. Shiao et al., who argue that these genetic clusters represent "clinal classes" homologous to race, see this departure of the race concept from its racialist roots as a defense of their argument. They say:

> Arguably, the origin of the essentialist criterion for biological differences lies less in actual science than in its use in the historical justifications for the categorical exclusion of nonwhites from political, economic, social, and cultural citizenship in the United States. By contrast, biological science does not require the white supremacist belief in species-level, much less greater, differences between human subspecies.[21]

But this is hardly a defense; it is a refutation. This view of race attempts to recuperate the old race realist categories for no scientifically motivated reason; the most benign reason is merely that these categories are familiar to us. Why adopt a term that is entirely inapt and loaded with a racist history to describe a novel putative biological phenomenon? A major worry is that calling this conception of human population structure "race" reifies race realism in the conventional understanding. This borrowed nomenclature facilitates the slide back into racialist thinking. Indeed, the move to a statistical conception of race is the continued social construction of race occurring in real time.[22] As I'll note in what follows, this same move happens in race-based medicine. In summary, statistical conceptions of populations are not races, and we shouldn't use race terminology to describe this position.

Race realism is in deep tension with the dominant academic view of race—one shared by many scientists and the majority of social scientists and humanities scholars—that our race concept is a social construction with no deep, meaningful biological reality.[23] Social constructionists argue that race is a socially constructed category for sorting human beings that has social reality and real effects on people's life prospects. In other words, races are real, but they are grounded in social facts, not biological ones. But this means that any biological or medical differences between the races have their source in historical processes or socioenvironmental causes, not mythical race-based genes.[24]

The social constructionist position about race is supported by the historical record, which traces race formation through time. Consider the changing racial categories used in the United States throughout its history.[25] In 1790 the US Census categorized people by their legal standing, using the categories of "Slaves," "Free White Females and Males," and "All Other Free Persons." By 1820 "Slaves and Free Colored Persons" were grouped together in one category in contrast to "Whites" and "Other Free Persons," illustrating the conflation (and true reality) of legal, political categories of hierarchy and race. By 1850 the first category became fully racialized as "Black and Mulatto" (and by 1890 included further fine-grained categorizations of "Black," "Mulatto," "Quadroon," and "Octoroon"), reflecting an entrenched rule of hypodescent where offspring of Black, white, or mixed parents inherit the political status of the parent deemed socially inferior. Only a political system, rather than rules of biology, could explain why children with a smaller portion of African ancestry are categorized with other Black people rather than with white people. This taxonomy functioned to keep the white race "pure." It limited the number of white people with full property and other civil rights. These rules are obviously socially and politically constructed and biologically arbitrary.

In 1860 the US Census added other racial designations—"Indian" and "Chinese"—alongside "Black/Mulatto" and "White," reflecting the contested legal status of Native Americans and Asian immigrants and their descendants as neither white nor Black. These categories morphed through time to the present-day mix of census

race and ethnicity categories, always reflecting the social and political conditions and preoccupations of the day rather than biologically meaningful groupings.

The changes in the groups represented and rules about how people should classify themselves within them reflect social and political facts in the United States, including who could own land, changes in immigration demographics, and political solidarity. Regarding the last of these, the Asian American category came into existence in the 1960s as an explicit political rights movement echoing the civil rights achievements of the Black Power movement. Berkeley graduate students Yuji Ichioka and Emma Gee created the Asian American Political Alliance to unite discrete ethnic groups from Asia into one united political front, coining the term and hence race category *Asian American* at that time.[26] Another example is in the current racialization of Middle Eastern and North African people in the United States, many of whom, in the post-9/11 world, feel uneasy categorizing themselves as white and petitioned (unsuccessfully) to have the category *MENA* added to the US Census in 2020.[27] Social constructionism, not biological realism of any stripe, makes the most sense of these historically and socially grounded practices of race formation.

If races are socially constructed and not biologically real, then how can geneticists coherently advocate for racial diversity in order to achieve genomic diversity in research? Does the call for genomic diversity rely on the view that race categories are biologically real?

Before answering this question, it's worth noting that there are reasons to promote genomic research diversity that obviously do not assume race realism. One is that gene-environment interactions may differ by population—as a proxy for social circumstance.[28] Individuals, even with similar genes, in different environments may have different health outcomes. Including people with diverse backgrounds could eventually help scientists gain clarity on these interactions. This reason for genomic research inclusivity is not about genetic variation between racially defined groups but rather about socioenvironmental differences.

A COLORFUL ANALOGY

My aim is to show that increasing racial diversity among the genomic samples researchers use can increase genetic diversity in their research, even though race is not a biologically meaningful category. Specifically, what I'll argue is that diversity among socially constructed categories can be a proxy for diversity in some underlying physical reality, without the socially constructed categories being physically real. I endeavor to make the point through example, one that takes us away from the loaded debate about race. Take a natural phenomenon that is clinal—that is, gradual in variation—but upon which we've placed discrete, socially constructed categories. Variation in these socially constructed categories

may function as a proxy in some limited ways for the natural variation in the clinal, physical phenomenon upon which they are imposed. The color spectrum, our conventions about color names, and the more precise wavelengths that produce different colors offer an apt example. To understand, we need to get nerdy about color for a moment.

Color is the perception of electromagnetic radiation in the visible spectrum of light.[29] Objects absorb some wavelengths of light while reflecting others. The wavelengths that are reflected back to an organism's eyes are, depending on the color receptors that organism has, perceived as a color. Humans can see wavelengths measuring 390–750 nanometers (nm).[30] For example, a blue object is one that reflects wavelengths measuring 450–495 nm. The spectrum of visible light, composed of wavelengths that produce colors when perceived by us, is gradual in nature. Indeed, the word "spectrum" has come to mean the organization of things that vary gradually in some regard that can be arranged from one extreme to the other. There are no discrete boundaries between areas on the spectrum; the wavelengths on the spectrum gradually change into one another, and so do the corresponding perceived colors.

We humans have given ranges within the light spectrum particular color names. We've roughly carved up the visible spectrum into color bands of red, orange, yellow, green, blue, and purple. But our color designations, which demarcate these bands into bounded, discrete groups, are socially constructed, and they vary by culture through place and time. For instance, the standard ROYGBIV seven-color partition of the spectrum, familiar to English speakers, originated with Aristotle, who theorized that there were seven colors just as there are seven musical notes.[31] Isaac Newton added orange and indigo to the already recognized colors of his time and place in order to achieve the Aristotelian ideal of seven colors and to honor the tradition of alchemy, in which the number seven has significance.[32] This is a vivid example of how cultural preference and human choice dictate the number of categories we impose upon a spectral reality. We are not carving the spectrum at its natural joints. Indeed, the spectrum—being clinal—has no such joints.

Consider other cultural variations. Greek and Russian speakers have distinct words for two different shades of blue, while others—for instance, speakers of English—use a broad category of blue for all hues within this range.[33] But the Tahitian, Tzeltal, and Japanese languages group blues and greens into one color category. In English, we consider red and pink distinct colors. Yet we have no such discrete separation between a saturated blue and a pastel one, the light-blue analogy to pink.

Ultimately, which colors we identify as distinct and how fine-grained our choices are may be the result of whether or not we have a purpose for making distinctions within broad color groups. In brief, color categories are culturally

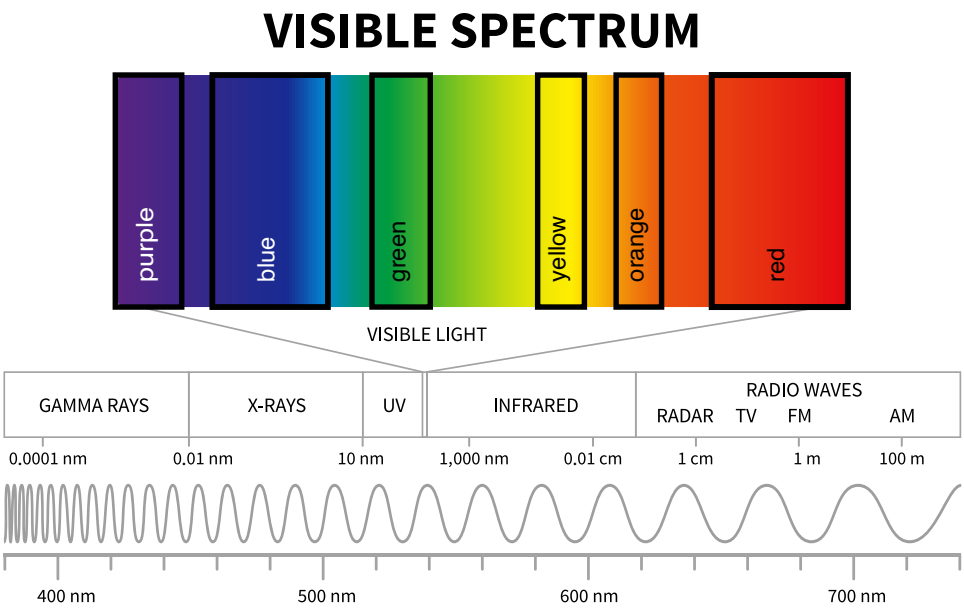

FIGURE 2.1. The conventional English-language colors are imposed on the visible light spectrum as discrete categories. Adobe Stock #229007362, modified by the author.

constructed. Colors are categories we impose on the spectrum. They are discrete, and while perhaps explainable by reference to culture, cultural development, or our physiology, they are arbitrary with regard to any distinguishing features of the light spectrum itself.

Yet our socially constructed color categories are still informative about the spectrum. Within each socially constructed color band is a group of wavelengths. For example, wavelengths of 620–750 nm produce the color red; those of 590–620 nm create orange. Let's say you are an eccentric collector: you collect electromagnetic wavelengths. You have many beautiful colored objects that reflect various wavelengths and produce lovely colors in the eye. But your collection is not very diverse. You have a lot of objects that reflect long wavelengths, the wavelengths that produce the color red. You have objects that are crimson, vermilion, shading into orange, and even some orangey-yellows. But you long to have a broader collection of wavelengths that represents the visible light spectrum of colors. If you wanted more wavelength diversity in your collection, you would do well to put out an advertisement for objects that are green, blue, and purple. Diversity in these broad color categories will be a good proxy for diversifying your collection with objects that reflect different wavelengths than do the objects in your current collection. Color categories, which are themselves socially constructed and do not map onto any physically real, discrete categories, can be a proxy for some underlying physical reality.

You can see the analogy. Genomic researchers have genomic samples mostly from individuals of proximate European ancestry. Many of these people will socially identify as white, as well. We know that allele frequencies among humans vary clinally by geography, with genetic distance correlating with geographic distance.[34] Thus, this narrow, geographically defined group represents only a limited set of human genomes. If you want to diversify your research genomically, people's ancestry can tell us which geographic region some or most of their proximate progenitors came from.[35] Race can be a rough guide to a person's proximate ancestry because it has been socially constructed to categorize people by visible traits that roughly correlate to having proximate ancestry from particular geographic regions. Thus, because of how race and ancestry are socially constructed, racial diversity can be a proxy for obtaining that geographically based genomic diversity. If you want more diversity in your mostly "white" genomic samples, you would do well to recruit for people of African, Asian, and Indigenous American ancestries, and so on. Race and ancestry of individuals, socially constructed categories, can be helpful, in this context, for indicating something biologically real. Namely, if you have more racial diversity on the whole among your samples, you should get more genetic diversity. But that does not mean that race and ancestry groupings are biologically real.

Like the color categories we've imposed upon the light spectrum, the race categories we've imposed upon geographical human populations are crude and arbitrary. We could have carved up the spectrum differently—for example, why not have a unique name for the yellow-orange of a marigold, why not carve up the blues into more discrete categories of aqua, cobalt, and periwinkle? Likewise, we could have carved up human populations differently. Why not, for instance, have more fine-grained categories for African populations, given all the genetic diversity in Africa?[36] But these categories can still do some work. Arbitrary though they are, we know that conventionally blue objects will reflect shorter wavelengths. Arbitrary though it is, we know that if genomic research focuses mostly on individuals who identify as white, it is lacking the genomic diversity that can be found in a more diverse sample of people who identify as Black, Pacific Islander, Asian, or Indigenous American.

So racial diversity can be a proxy for the purpose of getting more genetic diversity in our genomic research, just like color can help the fictional wavelength collector diversify their collection. But the fact that racial diversity is a proxy does not mean that race is biologically real, just as color diversity as a proxy for wavelength diversity does not mean that color categories map onto discrete features of the real light spectrum.

The point of the spectrum analogy is to simplify the issue at hand and put it in other terms in order to try to make sense of an otherwise novel and complicated phenomenon. But that simplification comes at a cost. The real pictures, for both

real colors and the relation between race categories and genetic diversity, are far more complex. A simple analogy has its limits.

Complicating this analogy in accordance with reality, however, can be instructive. The spectrum itself is a simplification of color. We get pure colors represented on the spectrum, saturated colors like true yellows and greens. In real life, the color of objects is very rarely pure. Most real colors are mixes of the purer, more saturated colors, just as Mark Fedyk noted in his chapter that most gold in the world is alloyed. The color of a real apple is not vivid, saturated, pure red but rather is a brownish, grayish red. In reality, a real apple reflects back all of the spectrum wavelengths, just in different proportions, so that red wavelengths are dominant.

Something similar can be said of humans. Real humans are not representations of "pure" ancestral populations from which they came. There are no such things. Real humans are the products of complex human breeding histories. We all have genetic ancestors who came from many different places. Most of the alleles found in different frequencies in different parts of the world are present across the globe. The simplifying analogy of the color spectrum is inapt in at least one way to represent clinal human genetic diversity because any particular individual probably has genes that represent a crisscrossing, complex ancestral lineage that does not easily allow us to order individuals clinally along one dimension. Human breeding patterns and migration are dynamic; our ancestral populations did not stay in just one place, nor did they remain isolated from one another. Our genes reflect this dynamic, intermixing history.

Once we move to the more complex understanding of color, we can see the issue. A pure color spectrum can organize color linearly because it focuses on only one dimension of color: *hue*. Hue is the main local color of an object—for example, blue. But colors have two other main properties. *Saturation* is the purity of the color: is it a vivid, true blue with little else mixed in, or is it a desaturated, muted slate (a blue with gray in it)? *Value* is the depth of the color, how much black or white is mixed in: is it a dark navy or a pastel sky blue? Color theorists have endeavored for centuries to organize all the variation within colors in a way that could reflect these dimensions, coming up with complicated forms that relate all the colors in three dimensions. Move beyond the simple one-dimensional color spectrum, and this task proves quite difficult. Look at Munsell's color system.

Human genetic diversity is even more complicated. Humans have many gene variants, and although these vary clinally across global geographic distance, they vary "nonconcordantly," meaning that they do not covary together geographically.[37] If color diversity is difficult to represent with just three dimensions, organizing human genomic diversity with many more dimensions that can combine in multiple permutations is near impossible. So there is a limit to the spectrum analogy. In its simple form, it functions to show that diversity in socially constructed, discrete categories can serve as a proxy for diversity in some underlying, clinal physical reality. But more detailed, precise inferences from color to wavelength, or

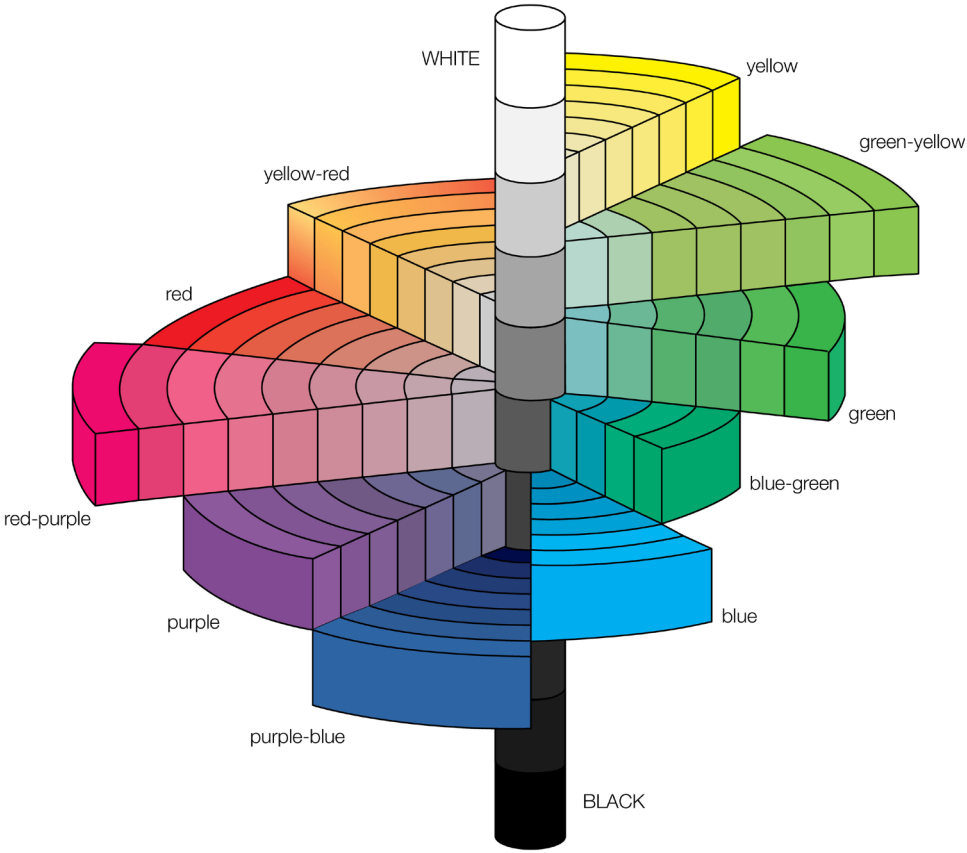

FIGURE 2.2. The Munsell color tree showing Albert H. Munsell's organization of colors by hue, value, and saturation, demonstrating the complexity of organizing spectral phenomena along more than one dimension of measurement. Universal Images Group North America LLC / Alamy Stock Photo.

race to genotype, are blocked when we add complexity to the model in accordance with messy reality.

RACE IN THE CLINICAL SETTING

We see the limit to the inference when we move race to the clinical setting. From the fact that racial diversity can serve as a proxy for genetic diversity among a large population of people, one might infer that the race of one individual can be a guide to their underlying genotype. But this a fallacious inference.

Back to our analogy. Say you are a comprehensive color collector, and you have objects representing the vast array of the pure spectrum. But you do not have any

objects of the specific wavelength 578 nm. This wavelength falls into the green band on the spectrum. So you hoard a large set of green objects. Now, while this would be a better method for narrowing your search for 578 nm—say, as compared to scavenging for red objects—it is quite crude. You are not guaranteed to get 578 nm if you search for green. Green is quite broad a category to be a reliable proxy for something as specific as 578 nm. Importantly, if you have a particular green object in front of you, you cannot assume it is 578 nm.

The point strengthens when you consider the complexities we add to this analogy. Consider now that this is a real, colored object, not one merely representative of the pure color spectrum. You can't assume this green object doesn't have the wavelengths of other colors in it, since real green-colored objects have a complex mix of all the wavelengths. In fact, 578 nm could be present in any of your non–green-colored objects. In brief, with real colors, you can't infer from the presence of some color that you have either the presence or the absence of a particular wavelength.

Likewise, insofar as socially identified race is a rough proxy for people's genetic ancestry, then racial diversity can be a good proxy for getting more genomic diversity in your research.[38] But it won't guarantee you the presence or absence of any particular gene variants at the more fine-grained level, just as seeking green objects in no way guarantees that you will get the 578 nm wavelength. While certain gene variants are more frequent in certain global populations than in others, they may be present in lower frequencies all around. And within a population with a higher frequency of an allele, there will be some individuals who do not have it. Thus, focusing on an individual's race is not helpful in guessing which gene variants they may or may not have. One is not rationally licensed to move from the idea that ancestral background and race are proxies for genetic diversity within a sample with many people in it to inferring anything about the genes of an individual person in front of oneself from the way they look or how they identify.

Take the following example, in which race is used in the clinical setting as a proxy for the kinds of genes a person can have. Cystic fibrosis is a monogenic, recessive disease that primarily affects the lungs, resulting in excess production of mucus, difficulty breathing, lung infections, and hence shortened lifespan. It is commonly seen as a white disease, and one that affects Askhenazi Jewish people in particular. Dorothy Roberts tells the story of a two-year-old African American girl who presented in the emergency room with respiratory issues.[39] She had ongoing respiratory issues for years, until at age eight, a new doctor looked at her lung scan, not knowing her race, and accurately diagnosed her with cystic fibrosis. The child's race obscured the possibility of accurate diagnosis for her clinicians, who did not consider the possibility that people who were not socially identified as white could have cystic fibrosis. For this error, she went undiagnosed and untreated for a deadly lung disease for years. Race-based medicine runs this dangerous risk of licensing the assumption that a gene variant cannot be present in a person because of their race.

There are several reasons why race fails to be a good proxy in the clinical setting. First, our racial categories are too broad. Recall that there is more genetic diversity in the group of people with recent African ancestry than in any other continentally defined group. Recruiting people who self-identify as Black is a good way to cast a wide net for genetic diversity. But we cannot infer, with the appropriate level of accuracy, that any particular Black-identified individual has a particular gene variant or trait. This may seem obvious. Yet the mistake is repeatedly made. Consider the claim that 40 percent of people of African ancestry are slower metabolizers of antidepressants, which is used as grounds in the clinical context for giving anyone of African ancestry—usually via self-report or clinician report of the patient as Black—a different dose than one would give a white person.[40] Yet according to this statistic, fewer than half of people of African ancestry have the trait. One using this racial heuristic in clinical treatment is undertreating more than half of their Black patients. And some non-Black patients may be overtreated since some percentage of them are presumably slow metabolizers of antidepressants.

Another reason race is a poor proxy in clinical practice—setting aside the point that racialized groups are too broad—is the arbitrariness of social rules for the assignment of race, which obscures the reality of "racial mixing." In the United States, for instance, someone who has half recent African ancestry and half recent European ancestry may identify as Black or African American due to the historical rule of hypodescent and the contemporary political understanding of racial group assignments. But this person has recent ancestry from at least two different, continentally defined ancestral groups. They may identify as Black in the clinical setting; they may be identified by the clinician as Black based on appearance and the historical, social rules for designating "mixed" ancestry. Both they and their clinician would be ignoring half of their ancestry if they are defined this way. These worries are especially sharp for diasporic Africans and Latinx individuals, given these groups' rich, diverse ancestry. Thus, even if ancestry is ultimately what matters and race is a crude proxy for ancestry, our social construction of race gives simple, typically discrete racial assignments to people with complex ancestral histories. This complexity is erased in the clinical encounter when a person self-reports their race or a doctor infers it based on their appearance. Racial determinations in the clinical setting are typically based on self-report.[41] Self-report of a politically created category is a very poor proxy for an individual's genetic profile.[42] Alternatively, a clinician surmising a patient's race based on their appearance, last name, or other features is not reliable either. None of these features is a reliable indicator of a person's complex ancestry. Further, although race is already a poor category for making these kinds of inferences, it is still less useful given the increased mixing of people who are differently raced. This is particularly concerning since the population of people who self-report being of two or more races is growing.[43]

Race is also a problematic category in the clinical setting because of its overtly social and political construction, which brings together people with diverse ancestry. Take, for example, the Asian race in the US context. As discussed earlier, "Asian American" is a political category intentionally created during the civil rights movement to unify people with ancestry from the Asian continent who are small minorities of the US population. Grouping together gave these subpopulations of Asia critical mass and relatively stronger political power in the United States. But *Asian*—derived from this conception of Asian American—in the US context includes people from South Asia, Southeast Asia, and East Asia, people who could not be grouped together as one homogeneous genetic population (nor would this be true for any of these constituent subcategories). Yet Asian is used as a racial category in medicine. For example, in the United States, spirometers, which measure lung function, are routinely "race-corrected" for Black and Asian patients.[44] I've already discussed the problem of using "Black" as a biologically meaningful category. Likewise, what started as a category for an explicit political power movement among Asian Americans has been biologized by the medical establishment as an indicator of innate biological response in the clinical setting. There is no scientific warrant for this.

The use of race in the clinical setting suggests that "racially profiling doctors" have internalized crude race realism in making their assumptions about patients. Were crude race realism true, it would better allow the inference from individual to gene or trait because race realism is the view that races are discrete and essentialist. So being of race *X* means having the features that people of race *X* have. This kind of race realism is false. We have no justification for sliding back into it in medical practice. In addition to the dangers of misdiagnosis, this practice sends the message that crude racialist races are real.

But the statistical notion of race, endorsed by some scientists and doctors, does not license the inference either. At best, among a group of people similarly racialized, we see an increase in some clinically relevant alleles in the group. But one is guilty of committing the ecological fallacy when one moves from this group-level statistic to inference about individual risk. Higher incidence of *Y* among a defined population does not mean an individual member of the group has a higher risk of *Y*. This is starkly the case when the criterion for grouping itself is not medically or biologically meaningful.

One might interject and claim that these alleged statistical associations between race and certain gene variants could be meaningfully deployed in medicine, even if they are not perfect. But that is too hasty. Determining whether race is a good proxy for genes requires settling a value-based assumption. Our tolerance for a proxy's accuracy can vary from context to context. The context tells us how sensitive and specific the proxy must be for our purposes. A highly sensitive test gives us a high rate of true positives. A highly specific test gives us a high rate of true

negatives. We must evaluate the risks if our proxy is not precise and weigh them against the benefits of having a proxy with its particular level of accuracy. Context tells us what risk tolerance we should have.

For instance, race may be an appropriate proxy for genomic research recruitment. In recruiting for genomic research, all we need to do is cast a very wide net in order to include many people from many different geographical areas. The risks to individuals in doing so are relatively minimal.[45] We are including them in a genomic sample database but not otherwise interacting with them. Yet, in the clinical setting, the lack of precision in inferring genes from race has real, tangible risks, including misdiagnosis and inaccurate drug dosing. The case of the Black child with cystic fibrosis is instructive. The risks of misdiagnosis using race as a proxy for genotype are high—they are life or death. In this context (and for the other reasons already mentioned), race is not a good predictor of genotype. This then limits the use of race as a proxy for genomic diversity in genomic medicine.

CONCLUSION: RACIAL DIVERSITY CAN BE A LIMITED PROXY FOR GENOMIC DIVERSITY

Human genetic diversity changes gradually across the globe, with genetic differences correlating with geographical distance. Since the US race construct is based on taxonomizing people by relatively recent geographic place of origin for proximate ancestors, this common race construct is a rough proxy for large, continental ancestral place of origin. For this reason, race might be a good proxy for casting a wide net to increase genetic diversity in genomic research. For reasons of justice, we should care about and advocate for racial diversity in genomic research. All people should be able to benefit from the medical findings of these studies. But advocating for racial diversity in genomic research does not require a commitment to biological race realism. Racial diversity, relying on race as a socially constructed category, can serve as a proxy for genomic diversity; more racially diverse people included among genomic samples should correlate with more genomic diversity. But we should also be very aware of the limitations of this relationship. We'll need racial diversity to develop genome-based, precision medicine equitably; but race should not serve as a proxy for making genetic inferences about individuals in the clinical setting.

There is another way that race can be relevant to health outcomes, and that is through socially and environmentally mediated processes. Racism, differential access to health care, exposure to pollutants, and so on have deep and lasting health outcomes and are differentially distributed by race. My hope is that the preoccupation with the relationship between genes and race does not obscure this more promising avenue for understanding racial disparities in health.

NOTES

1. "GWAS Diversity Monitor," accessed August 24, 2023, https://gwasdiversitymonitor.com.

2. Carlos D. Bustamante, Esteban Gonzalez Burchard, and Francisco M. De La Vega, "Genomics for the World," *Nature* 475, no. 7355 (2011): 163–65, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3708540/; Alice Popejoy and Stephanie M. Fullerton, "Genomics Is Failing on Diversity," *Nature* 538, no. 7624 (2016): 161–64, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5089703/; Jonas Korlach, "We Need More Diversity in Our Genomic Databases," *Scientific American: Voices*, December 4, 2018, https://blogs.scientificamerican.com/voices/we-need-more-diversity-in-our-genomic-databases/.

3. See Michael Montoya, *Making the Mexican Diabetic: Race, Science, and the Genetics of Inequality* (Oakland: University of California Press, 2011); Dorothy Roberts, "Debating the Cause of Health Disparities: Implications for Bioethics and Racial Equality," *Cambridge Quarterly of Healthcare Ethics* 21 (2012): 332–41; Jay S. Kaufman, Lena Dolman, Dinela Rushani, and Richard S. Cooper, "The Contribution of Genomic Research to Explaining Racial Disparities in Cardiovascular Disease: A Systematic Review," *American Journal of Epidemiology* 181, no. 7 (2015): 464–72.

4. Hispanic/Latinx people, often taken to be a race by Americans, are considered an ethnicity in this taxonomy, although the contemporary racialization of this group is yet one more example of the social construction of race in real time.

5. National Academies of Sciences, Engineering, and Medicine (NASEM), *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field* (Washington, DC: National Academies Press, 2023), https://nap.nationalacademies.org/catalog/26902/using-population-descriptors-in-genetics-and-genomics-research-a-new.

6. NASEM, *Using Population Descriptors*, 7.

7. Roberta Millstein, "Thinking about Populations and Races in Time," *Studies in History and Philosophy of Biological and Biomedical Sciences* 52 (2015): 5–11. Geneticists are not consistent in their definition and use of the term *population*, however.

8. Joshua Glasgow, "Is Race an Illusion or a (Very Basic) Reality?," in *What Is Race? Four Philosophical Views*, by Joshua Glasgow, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer (New York: Oxford University Press, 2019), 111–49.

9. Arjun K. Manrai et al., "Genetic Misdiagnoses and the Potential for Health Disparities," *New England Journal of Medicine* 375 (2016): 655–65, https://www.nejm.org/doi/full/10.1056/NEJMsa1507092; Popejoy and Fullerton, "Genomics Is Failing on Diversity."

10. For claims in this paragraph, see Bustamante, "Genomics for the World."

11. Bustamante, "Genomics for the World."

12. Bustamante, "Genomics for the World"; Serena Tucci and Joshua M. Akey, "The Long Walk to African Genomics," *Genome Biology* 20 (2019): 1–3. Humans also migrated back to Africa, increasing genetic diversity in Africa in yet another way.

13. James Gallagher, "'Huge Advance' in Fighting World's Biggest Killer," *BBC*, March 17, 2017, https://www.bbc.co.uk/news/health-39305640; George Adigbli, "Race, Science and (Im)precision Medicine," *Nature Medicine* 26 (2020): 1675–76, https://www.nature.com/articles/s41591-020-1115-x.

14. Bustamante, "Genomics for the World."

15. Popejoy and Fullerton, "Genomics Is Failing on Diversity." This finding is for whole genome and exome sequencing, which has slightly better but still inequitable diversity among its sample participants compared to GWAS, which scans the genomes for select variants.

16. Bustamante, "Genomics for the World."

17. The genetic clustering data is from Noah A. Rosenberg et al., "Genetic Structure of Human Populations," *Science* 298, no. 5602 (2002): 2381–85. Notably, Rosenberg et al. do not call these clusters races, writing in a follow-up publication: "Our evidence for clustering should not be taken as evidence of our support for any particular concept of 'biological race'"; Rosenberg et al., "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure," *PLoS Genetics* 1, no. 6

(2005): e70, 668. In contrast, the following authors think these clusters are evidence of biological races. See Jiannbin Lee Shiao, Thomas Bode, Amber Beyer, and Daniel Selvig, "The Genomic Challenge to the Social Construction of Race," *Sociological Theory* 30, no. 2 (2012): 67–88; Quayshawn Spencer, "How to Be a Biological Racial Realist," in *What Is Race? Four Philosophical Views*, by Joshua Glasgow, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer (New York: Oxford University Press, 2019), 73–110.

18. Joan H. Fujimura, Deborah A. Bolnick, Ramya Rajagopalan, Jay S. Kaufman, Richard C. Lewontin, Troy Duster, Pilar Ossorio, and Jonathan Marks, "Clines without Classes: How to Make Sense of Human Variation," *Sociological Theory* 32, no. 3 (2014): 208–27; Ann Morning, "Does Genomics Challenge the Social Construction of Race?," *Sociological Theory* 32, no. 3 (2014): 189–207.

19. Fujimura et al., "Clines without Classes"; Michael James, "Race," *Stanford Encyclopedia of Philosophy* (2020), https://plato.stanford.edu/entries/race/.

20. Fujimura et al., "Clines without Classes"; Morning, "Does Genomics Challenge the Social Construction of Race?"

21. Shiao et al., "Genomic Challenge to the Social Construction of Race," 70. This false separation of science from systems of power and politics is alarming and inaccurate.

22. See Morning, "Does Genomics Challenge the Social Construction of Race?," 90.

23. For a detailed discussion of social constructionism in both its political and its cultural varieties, see Sally Haslanger, "Tracing the Sociopolitical Reality of Race," and Chike Jeffers, "Cultural Constructionism," both in Joshua Glasgow, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer, *What Is Race? Four Philosophical Views* (New York: Oxford University Press, 2019).

24. Among philosophers, there is another camp. *Eliminativists* about race argue that since race is a biological concept and there is no meaningful biological reality to race, we should eliminate our race concepts from scientific (and other) discourse entirely. Instead, we should speak of *racialized groupings*, which are real, but explicitly socially constructed. The differences between eliminativists and social constructionists need not concern us here. The eliminativist can substitute claims in this paper about race as claims about racialized groupings. For more on eliminativism, see Glasgow, "Is Race an Illusion or a (Very Basic) Reality?"

25. "Measuring Race and Ethnicity across the Decades: 1790–2010," US Census Bureau, accessed August 24, 2023, https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html; Anna Brown, "The Changing Categories the U.S. Census Has Used to Measure Race," *Pew Research Center*, 2020, accessed February 8, 2024, https://www.pewresearch.org/fact-tank/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race.

26. Anna Purna Krishnamurthy, "In 1968, These Activists Coined the Term 'Asian-American'—and Helped Shape Decades of Advocacy," *Time*, May 22, 2020, https://time.com/5837805/asian-american-history/.

27. Neda Maghbouleh, Ariela Schacter, and René D. Flores, "Middle Eastern and North African Americans May Not Be Perceived, Nor Perceive Themselves, to Be White," *Proceedings of the National Academy of Sciences* 119, no. 7 (2020): e2117940119.

28. Bustamante, "Genomics for the World."

29. There is a large literature in philosophy on the nature of color and color perception that I am setting aside in order to offer an accessible analogy for my purposes here.

30. Adam Rogers, *Full Spectrum: How the Science of Color Made Us Modern* (New York: Houghton Mifflin Harcourt, 2021), 2.

31. Rogers, *Full Spectrum*, 38.

32. Rogers, *Full Spectrum*, 53–54.

33. Rogers, *Full Spectrum*, 151.

34. John H. Relethford, "Biological Anthropology, Population Genetics, and Race," in *The Oxford Handbook of Philosophy and Race*, ed. Naomi Zack (New York: Oxford University Press, 2017), 160–69, 163.

35. Ancestry itself is socially constructed. We all have ancestors in every generation of humans. Choosing as the relevant ancestry ones that pertain to the continental geographic regions is both a spatial choice

(which regions, how large or small?) and a temporal one (a particular snapshot in time where the relevant groups can be tied, roughly, to specific locations). These choices are informed by our socially constructed conception of what the races are.

36. Geneticists do this to some extent, specifying populations within Africa. They also frequently lapse into using the broader continental groupings.

37. Fujimura et al., "Clines without Classes."

38. Although we each have many ancestries, not one.

39. Dorothy Roberts, "What's Race Got to Do with Medicine?," *TED Radio Hour*, NPR, February 10, 2017, https://npr.org/transcripts/514150399.

40. Sally Satel, "I Am a Racially Profiling Doctor," *New York Times*, May 5, 2002, https://www.nytimes.com/2002/05/05/magazine/i-am-a-racially-profiling-doctor.html. Satel is a publicly vocal proponent of what she calls "racial profiling in medicine," and for that reason she may be seen as representing an extremist view. But her thinking is representative of a large swath of clinicians and, more generally, medical practices that use race as a proxy in the clinical setting. For a recent assessment of race-based medical practices, see Jessica P. Cerdeña, Emmanuella Ngozi Asabor, Marie V. Plaisime, and Rachel R. Hardeman, "Race-Based Medicine in the Point-of-Care Clinical Resource UpToDate: A Systematic Content Analysis," *eClinicalMedicine* 52 (2022), https://thelancet.com/journals/elinm/article/PIIS2589-5370(22)00311-X/fulltext. For a discussion and criticism of race-based clinical algorithms, see A. Vyas, Leo G. Eisenstein, and Davis S. Jones, "Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms," *New England Journal of Medicine* 383 (2020): 874–82. They state: "Most race corrections implicitly, if not explicitly, operate on the assumption that genetic difference tracks reliably with race."

41. Youssef Roman, "Race and Precision Medicine: Is It Time for an Upgrade?," *Pharmacogenomics Journal* 19 (2019): 1–4, https://www.nature.com/articles/s41397-018-0046-0.

42. Michael Root, "Race in the Biomedical Sciences," in *The Oxford Handbook of Philosophy and Race*, ed. Naomi Zack (New York: Oxford University Press, 2017), 463–73.

43. Roman, "Race and Precision Medicine."

44. Lundy Braun, "Race, Ethnicity, and Lung Function: A Brief History," *Canadian Journal of Respiratory Therapy* 51, no. 4 (2015): 99–101.

45. That's not to say there are no ethical issues for GWAS. See Stephen J. O'Brien, "Stewardship of Human Biospecimens, DNA, Genotype, and Clinical Data in the GWAS Era," *Annual Review of Genomics and Human Genetics* 10 (2009): 193–209, https://doi.org/10.1146/annurev-genom-082908-150133l; Jantina de Vries, Susan J. Bull, Ogobara Doumbo, Muntaser Ibrahim, Odile Mercereau-Puijalon, Dominic Kwiatkowski, and Michael Parker, "Ethical Issues in Human Genomics Research in Developing Countries," *BMC Medical Ethics* 12 (2011): 5, https://doi.org/10.1186/1472-6939-12-5.