

# The quagmire of race, genetic ancestry, and health disparities

Giorgio Sirugo,<sup>1,2</sup> Sarah A. Tishkoff,<sup>3,4</sup> and Scott M. Williams<sup>5</sup>

<sup>1</sup>Institute of Systems Pharmacology and Translational Therapeutics, <sup>2</sup>Division of Translational Medicine and Human Genetics, and <sup>3</sup>Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>4</sup>Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>5</sup>Departments of Population and Quantitative Health Sciences, and Genetics and Genome Sciences, Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio, USA.

The debate on human categorization based on “race,” its relationship to genetic variation among populations, and its scientific legitimacy has been raging for decades. Yet, there is still no clear consensus on the utility of “race” as a useful descriptor, especially in the realm of human health and disease (1). In the 18th and early 19th centuries, Carl Linnaeus and Johann Blumenbach used “race” to taxonomically classify human types (2). Linnaeus defined four human races based, largely, on skin color that matched continental population differences; Blumenbach similarly described races based mainly on physical traits, though he stated that “innumerable varieties of mankind run into each other by insensible degrees.” Blumenbach elaborated an incorrect Eurocentric theory of human origins, and traced the “ancestral white race” to the Caucasus, hypothesizing that all other races derived (and degenerated aesthetically) from a “Caucasian” one (2). Hence, “Caucasian” is a historical residue with a clear implication of hierarchy among groups that has no valid place in the scientific literature. In general, these “racial” classifications were viewed as intertwined with intellectual and/or moral ranking, with various degrees of complexity; the Comte de Buffon, a contemporary of Linnaeus (and young Blumenbach), appropriately interpreted variation in skin color as correlated with climate, but he also believed in European superiority over Africans that he explained as a product of different environmental conditions (2). Although skin color was used as a primary criterion to assign “race,” we now know this classification is only “skin deep” because skin color is an adaptive trait in response

to high UV exposure, with the most darkly pigmented people originating from across the globe in areas close to the equator (3). Indeed, populations from Africa, often referred to as a single race based on skin color, have higher levels of genetic diversity than that observed among all other global populations (4). Nevertheless, skin color-based human “typology” continues to be used and has largely contributed to falsely premised, biological racist theories with tragic consequences.

Remarkably, these historically defined human “races” capture much less genetic divergence than what is observed in populations of great apes (5). The relatively high levels of human genetic similarity in global populations results from their recent common ancestry from a source population of *H. sapiens* who left Africa 50,000 to 100,000 years ago. Importantly, there is less human genetic variation between geographic populations (~7%–12% globally) than among individuals from the same population (~88%–93%), and even small interbreeding populations display considerable amounts of genetic heterogeneity (6). Furthermore, at the genome level, on average humans differ from the reference genome(s) by only a few million base pairs, representing greater than 99% identity (6). Even though regional or population-specific variation exists, especially for rare variants, it is minimal compared with the amount of variation shared between populations (6).

## Variation in trait prevalence among populations

Ethnicity, which corresponds with common history, culture, faith, rituals, and

language, is often a more useful classifier than “race.” However, neither “race” nor ethnicity necessarily reflects genetic ancestry, which is defined as genetic similarities derived from common ancestors (7). Further, common diseases with differences in prevalence among ethnic groups can have both genetic and environmental risk factors. For example, American Indian/Alaska Native adults have the highest type 2 diabetes mellitus (T2DM) prevalence compared with other ethnic groups (14% of all Native Americans are affected), and the Pima Indians of Arizona, specifically, have the highest prevalence in the world (up to 50% of the Pima population is affected) (8). In contrast, the Pima of Mexico have T2DM prevalence under 10%, demonstrating that environmental factors can be strong determinants of T2DM even in genetically predisposed individuals (9). Another example of the effect of the interaction of genes and environment on disease risk was demonstrated in a study that examined the prevalence of asthma in Puerto Ricans, who have varying amounts of Native American, African, and European ancestry. For those who had a lower socioeconomic status, European ancestry was associated with increased risk of asthma, and African ancestry with decreased risk (10). However, in those who had higher socioeconomic status, European ancestry was associated with decreased risk of asthma, and African ancestry with increased risk.

Variation in prevalence of both common and Mendelian disease among populations can be due to differences in risk allele frequencies caused by random founder events and/or by natural selection. For example, STAC3 disorder, caused by mutations in the SH3 and cysteine-rich domain 3 (*STAC3*) gene, is associated with risk of malignant hyperthermia during anesthesia; it has a prevalence of 1:5000

**Authorship note:** GS, SAT, and SMW contributed equally to this work. The order of authorship is alphabetical.

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Copyright:** © 2021, American Society for Clinical Investigation.

**Reference information:** *J Clin Invest*. 2021;131(11):e150255. <https://doi.org/10.1172/JCI150255>.

among the Lumbee of North Carolina but is extremely rare in other populations (11). Populations may also differ with respect to variability in therapeutic response to drugs; in Asians, a lower dose of rosuvastatin (a statin used to lower LDL-C levels) has been recommended by the FDA because pharmacokinetic studies have shown a nearly two-fold increase in average drug circulating levels in individuals of Asian ancestry as compared with those of European ancestry. This observation is important because side effects associated with statin blood levels range from relatively common myalgia to rare instances of rhabdomyolysis. Variants in two genes (solute carrier organic anion transporter family member 1B1 [*SLCO1B1*] and ATP-binding cassette subfamily G member 2 [*ABCG2*]) that encode drug transporter proteins play key roles in rosuvastatin absorption, distribution, and elimination, and differ in frequency between Europeans and Asians. Therefore, *SLCO1B1* and *ABCG2* genotypes can better predict appropriate therapeutic dosing than ancestry alone (12). A consequence of using broad-based “racial” classification to determine treatment with rosuvastatin is that clinicians may be hesitant to prescribe it to people of Asian ancestry who might benefit from higher doses to prevent and treat atherosclerotic cardiovascular disease. In addition, “Asian,” as a classifier, encompasses many ethnicities with differences in atherosclerotic cardiovascular disease risk that can be due to a variety of nongenetic factors, including access to specialty lipid clinics, socioeconomic and education factors, and comorbidities (13).

### Impact of “race” and genetic ancestry on health disparities

Researchers conducting genome-wide association studies (GWAS) often use principal component or similar analyses to group people by broad geographic ancestry (referred to as “global ancestry”) and to control for population substructure, which can cause false positive (or negative) results, if not corrected. Genetically inferred clusters often, but not always, correlate with commonly used “racial” classifications based on broad geographic origin, although many individuals (especially those who are admixed) do not neatly cluster into a group. Individuals who are

admixed may have different ancestry at specific regions of the genome (referred to as “local ancestry”) despite similar global ancestries. For example, African Americans, on average, have approximately 80% West African ancestry and approximately 20% European ancestry (though this varies among individuals and by geographic region in the United States) but they may have 100% European, 100% African, or mixed ancestry at particular loci that affect disease (4, 14). Thus, “global genetic” ancestries may not correspond with genetic risk for disease at any particular locus. A risk allele in an individual who self identifies as “African American” and with high percentage of African ancestry can derive from a European ancestor, while a risk allele inherited from an African ancestor may occur in an African American individual with mostly European ancestry. Genetic ancestry and underlying patterns of genetic diversity can only affect disparity of disease through the portion of the genome that differs among populations and that associates with disease. Hence, “racial” classifications may not capture genetic differences that associate with disease risk. Variants associating with diseases will not, in most cases, have any relationship to “race” as socially defined, and hence, using this categorization can be misleading.

Although “race” is often used in diagnostic processes, it is not clear whether or how it affects risk or outcomes for many diseases. In addition, medical interventions are sometimes designed using “race” as a criterion, even though there is limited evidence to support its use (15, 16). The potential consequences of using “race” in diagnostic algorithms are non-negligible, including the possibility of misdiagnosis and inappropriate therapeutic treatment. A good example comes from nephrology, where calculations based on serum creatinine levels are used to infer the estimated glomerular filtration rate (e-GFR). Data indicate higher serum creatinine concentrations, on average, in individuals classified as African American versus European American; given the varying levels of African ancestry in self-identified African Americans, application of “race”-based algorithms may result in some people having a “race”-corrected e-GFR value that indicates better renal function than they have (15). The validity and use of this

“race”-based correction has been questioned as it can delay treatment in people with African ancestry (17), who have higher rates of kidney failure than people of other ancestries (African Americans represent 13% of the US population but >30% of patients with end-stage renal disease). On the other hand, lack of appropriate adjustments based on ancestry could lead to overtreatment. Ultimately, precision medicine based on individual genetic risk factors should supersede simple racial classifications. For example, apolipoprotein L1 (*APOLI*) genotypes (variants G1 and G2) associate with multiple fold (2-fold to 89-fold in certain patient subgroups) increased risk of several types of kidney disease (18). These variants are relatively common in some populations of West African descent, including African Americans, but are much rarer in most non-African descent populations and in populations from other regions of Africa; this frequency difference can plausibly be explained by a selective advantage in individuals with the risk alleles due to protection from infection by co-occurring trypanosomes causing sleeping sickness, a parasitic disease with high mortality, if untreated (19).

Another example of how natural selection may impact allele frequency distributions that have clear clinical implications is the impact of glucose-6-phosphate dehydrogenase (G6PD) deficiency on glycated hemoglobin levels (HbA1c), an indicator of blood glucose over time and a diagnostic criterion for T2DM. G6PD deficiency is common in populations historically endemic for malaria because it confers protection from the disease (20). In African Americans, the X-linked G6PD G202A variant, rs1050828, causes enzymatic deficiency (T-allele, frequency >10% in West Africans but <0.1% in European-descent individuals), and is associated with a decrease in HbA1c that could result in underdiagnosis of T2DM in approximately 650,000 people of African ancestry in the United States (21). Other mutations from East Asia that cause G6PD deficiency have also been associated with altered HbA1c levels (22). The differences here are not based on “race” per se, but on specific genetic variants that differ in frequency among populations.

Although “race” is not a useful classifier for describing population genetic differences, it can be useful for identify-

ing sociodemographic factors that influence health disparities. For example, “race” may influence chronic stress due to racial discrimination as well as differences in health care access, both of which can affect disease risk. A case in point comes from observations that preterm birth is approximately 50% more common in African American compared with European American women (23). However, women of African ancestry living in Canada show lower preterm birth rates than women of African ancestry living in the United States (8.9% vs. 12.7%), likely due to environmental risk factors, including differences in access to health care in the United States. The impact of environmental factors is further reinforced by the observation that foreign-born women of African ancestry in the United States have better birth outcomes than those who are native born (24).

## Conclusions

Our goal in this Viewpoint has been to explore how “race” is used and misused in medicine. We have argued that although “race” and genetic ancestry both can play a role in disease disparity, they are often independent of each other. “Race” may be seen as a proxy of some genetic differences (but we argue it is a poor one, as clearly shown in admixed populations), but in other situations it may be a risk factor independent of the genetic variation that exists because it is a proxy for racial discrimination and social inequities. Although many current disparities are driven by sociodemographic factors, there are some that are driven by genetic factors, and distinguishing the environmental and genetic factors influencing health disparities (and their interactions) is an important challenge for improving public health. As researchers and clinicians we need to heed the population geneticist Theodosius Dobzhansky (25): “diversity is an observable fact of nature, while equality is an ethical commandment” and inequality is “not biologically given but ... a societally imposed prescription.” We need to fill the existing gap in current knowledge of human genetic diversity (26), and understand if, and when, genetic diversity plays a role in disease risk differences as opposed to inequalities defined by socially determined “race.” Then we need to strive

to remove those inequalities in order to achieve health equity.

## Acknowledgments

GS was supported by NIH award U01HG010219. SAT was supported by NIH grants 1R35GM134957 and R01AR076241, American Diabetes Association grant 1-19-VSN-02, and a grant from the Chan Zuckerberg Foundation (CZF2019-002449). SMW was supported by NIH grant 5R01LM010098 and the March of Dimes Prematurity Research Center – Ohio Collaborative. We thank Dana Crawford, Reed Pyeritz, and Hua Tang for commenting on earlier versions of the manuscript.

Address correspondence to: Giorgio Sirugo, Division of Translational Medicine and Human Genetics, University of Pennsylvania School of Medicine, Department of Systems Pharmacology and Translational Therapeutics, 11-155 Smilow Center for Translational Research, 3400 Civic Center Blvd, Building 421, Philadelphia, Pennsylvania 19104-5158, USA. Phone: 215.746.0358; Email: giorgio.sirugo@penmedicine.upenn.edu. Or to: Sarah Tishkoff, Departments of Genetics and Biology, Center for Global Genomics and Health Equity, University of Pennsylvania, 415 Curie Blvd, Philadelphia, Pennsylvania 19104, USA. Phone: 215.746.2670; Email: tishkoff@penmedicine.upenn.edu. Or to: Scott Williams, Department of Population and Quantitative Health Sciences, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. Phone: 216.368.5659; Email: smw154@case.edu.

1. Yudell M, et al. Science and society.

Taking race out of human genetics. *Science*. 2016;351(6273):564–565.

2. Fredrickson GM, Camarillo A, eds. *Racism: A Short History*. Princeton University Press; 2002.

3. Crawford NG, et al. Loci associated with skin pigmentation identified in African populations. *Science*. 2017;358(6365):eaan8433.

4. Tishkoff SA, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009;324(5930):1035–1044.

5. Templeton AR. Biological races in humans. *Stud Hist Philos Biol Biomed Sci*. 2013;44(3):262–271.

6. Genomes Project C, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.

7. Mathieson I, Scally A. What is ancestry? *PLoS Genet*. 2020;16(3):e1008624.

8. Ravussin E, Bogardus C. Energy expenditure in the obese: is there a thrifty gene? *Infusionstherapie*. 1990;17(2):108–112.
9. Schulz LO, et al. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care*. 2006;29(8):1866–1871.
10. Choudhry S, et al. Dissecting complex diseases in complex populations: asthma in Latino Americans. *Proc Am Thorac Soc*. 2007;4(3):226–233.
11. Stamm DS, et al. Native American myopathy: congenital myopathy with cleft palate, skeletal anomalies, and susceptibility to malignant hyperthermia. *Am J Med Genet A*. 2008;146A(14):1832–1841.
12. Wu HF, et al. Rosuvastatin pharmacokinetics in Asian and White subjects wild type for both OATP1B1 and BCRP under control and inhibited conditions. *J Pharm Sci*. 2017;106(9):2751–2757.
13. Amrock SM, et al. Health disparities among adult patients with a phenotypic diagnosis of familial hypercholesterolemia in the CASCADE-FH patient registry. *Atherosclerosis*. 2017;267:19–26.
14. Bryc K, et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci U S A*. 2010;107(2):786–791.
15. Vyas DA, et al. Hidden in plain sight — reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874–882.
16. Borrell LN, et al. Race and genetic ancestry in medicine — a time for reckoning with racism. *N Engl J Med*. 2021;384(5):474–480.
17. Diao JA, et al. Clinical implications of removing race from estimates of kidney function. *JAMA*. 2021;325(2):184–186.
18. Friedman DJ, Pollak MR. APOL1 nephropathy: from genetics to clinical applications. *Clin J Am Soc Nephrol*. 2021;16(2):294–303.
19. Nadkarni GN, et al. Worldwide frequencies of APOL1 renal risk variants. *N Engl J Med*. 2018;379(26):2571–2572.
20. Tishkoff SA, et al. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*. 2001;293(5529):455–462.
21. Wheeler E, et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med*. 2017;14(9):e1002383.
22. Leong A, et al. Association of G6PD variants with hemoglobin A1c and impact on diabetes diagnosis in East Asian individuals. *BMJ Open Diabetes Res Care*. 2020;8(1):e001091.
23. Burrell HH, et al. The promise and pitfalls of precision medicine to resolve black-white racial disparities in preterm birth. *Pediatric Research*. 2020;87(2):221–226.
24. Elo IT, et al. Variation in birth outcomes by mother’s country of birth among non-Hispanic black women in the United States. *Matern Child Health J*. 2014;18(10):2371–2381.
25. Dobzhansky T. *Genetic Diversity and Human Equality*. Basic Books Inc; 1973.
26. Sirugo G, et al. The missing diversity in human genetic studies. *Cell*. 2019;177(1):26–31.