The distinct impacts of race and genetic ancestry on health

Genetic ancestry affects human health, but this is distinct from the impact of race, a social construct that has its foundations in systemic racism. These terms need to be better defined and understood in medical research to achieve health equity.

Timothy R. Rebbeck, Brandon Mahal, Kara N. Maxwell, Isla P. Garraway and Kosj Yamoah

There is substantial controversy about the meaning and value of the concept of race and related terms, such as racialism, racialization and racial formation[1,2]. These terms are often used to categorize individuals based on an ill-defined amalgamation of social, cultural and/or physical characteristics. The controversy arises from the implication that biological, ancestral or genetic components are defined by social concepts of race (and vice versa), despite the lack of clear, evidence-based actions for individuals or populations characterized by race. Consequently, the value of race in medical research, clinical practice and public health policy is under scrutiny due to a growing body of evidence that suggests that the use of race in medicine can cause more harm than good[3].

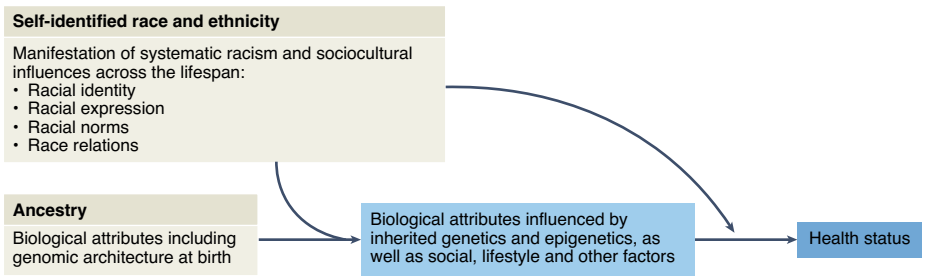

Fig. 1 | Influences of race and genetic ancestry on human health. Self-identified race and ethnicity and genetically inferred ancestry both impact on health, but in distinct ways.

Distinguish race from genetic ancestry

As the healthcare field attempts to better understand the complex interaction of ancestral, genetic and sociocultural factors on the biology underlying human health and disease, it is important to precisely define the distinct concepts of ancestry and race (Box 1 and Fig. 1). The term ancestry has various nuanced contextual meanings, including genealogical ancestry (an individual's family tree), genetic ancestry (an individual's genome) and genetic similarity (common genetic variation in populations)[4]. To consider ancestry in the context of precision medicine, genetically inferred ancestry, measured by inherited genetic variation, correlates with an individual's genealogical ancestry and may impact overall health, longevity, and disease susceptibility and severity (Box 1). Genetically inferred ancestry varies across geography, changes slowly over generations and with admixture, and can be measured by reference panels to infer an individual's continental or subcontinental genetic origins[5].

Genetically inferred ancestry must be distinguished from race, which has historically demarcated the relative degrees of power and privilege in society that often persist today. Derived from the racism that justified the institution of chattel slavery, the concept of race is propagated by laws and social norms that have codified discriminatory practices in criminal justice, political representation, governance structures, housing, employment, education and healthcare. Some propose that the term race be abandoned in favor of 'racism', as the systemic nature of racism is the central contributor to the poor health of some groups[6]. Others argue that race is a biological construct[7], while racial formation theory posits that race is both biological and social[8].

The American Association of Physical Anthropologists has stated[9] that: "Race does not provide an accurate representation of human biological variation and that humans are not divided biologically into distinct continental types or racial genetic clusters. Instead, the Western concept of race must be understood as a classification system that emerged from, and in support of, European colonialism, oppression, and discrimination. It thus does not have its roots in biological reality, but in policies of discrimination." Race (and ethnicity) are therefore socially and culturally defined identifiers selected by the individual (Box 1), with 'self-identified race and ethnicity' a useful term to include a wide range of sociocultural influences across the lifespan. Self-identified race and ethnicity is neither static nor discrete, but represents an amalgamation of social dimensions that may interact to impact health (Fig. 1), including identity (how individuals perceive themselves), expression (how individuals present themselves), norms (social expectations of race- and/or ethnicity-specific behaviors), and relations (how people react to different races and/or ethnicities).

Genetically inferred ancestry and self-identified race and ethnicity are not orthogonal concepts. Genetically inferred ancestry correlates with historical manifestations of racial identity, expression or norms, whereas self-identified race and ethnicity is influenced by individual phenotypes, such as skin tone, facies and hair types, that often overlap with genetically inferred ancestry. Sociocultural characteristics, lifestyle and behavior associated with self-identified race and ethnicity may influence the disease process by modifying gene expression. Thus, health and disease are influenced by the intersection of features that are associated with genetically inferred ancestry and self-identified race and ethnicity.

Although highly correlated, these variables should not be conflated or interpreted as having a fixed meaning. The variation inherent to both concepts is large, so discrete labels such as Black, African American, white, Hispanic and

Box 1 | Actions to tackle health disparities

Health disparities can be tackled by understanding the separate roles of self-identified race and ethnicity, and genetically inferred ancestry.

Researchers should recognize that both genetic and non-genetic factors can impact disease biology and contribute to disparities. Disparities can be defined by genetically inferred ancestry, as well as by social constructs such as self-identified race and ethnicity. Race and ancestry concepts, although frequently overlapping, should not be conflated to avoid propagating racism.

Any mention of ancestry or race should include precise definitions, with a description of why these concepts are appropriate for the research question. Assumptions and limitations associated with these concepts should be clearly understood and described.

Both sociocultural and ancestral diversity in research participants and populations is critical to ensure that interventions will optimally reduce disease risk and disparities for all populations.

Interventions must be developed in a manner that allows equitable access by all groups. Development of new tools or technologies that do not consider underlying socioeconomic differences in risk, access or usage across all groups may create or exacerbate disparities.

The interpretation and use of health-relevant data is not the purview of any one group. The generation and use of any data that compare ancestrally or socially defined groups must define the assumptions that were used to generate these definitions. However, these assumptions must acknowledge and respect the values and preferences of the group being studied, and ideally to involve the co-production of data generation and research results, to avoid potential stigmatization, misuse and misinterpretation of the data.

Similarly, research that uses genetic or ancestry data to compare or describe populations that are historically underrepresented in research and medicine must be explicitly considered within the context of systemic racism.

Asian are likely to be oversimplified when attempting to interpret their meaning and relationship to health and disease. However, practical considerations may require simplified concepts to be used to achieve specific research, implementation, policy or other goals. For example, the US Office of Management and Budget employs minimal definitions for race and ethnicity. The mandate to use and report these categories in federally funded research[10] has further entrenched simplistic racial categories in the scientific literature. When used, care should be taken to avoid overstating conclusions and drawing assumptions that reinforce historic racial stereotypes, prejudice and stigmatization of these groups.

Address health disparities

Human disease is the consequence of a multitude of genetic, environmental, cultural, socioeconomic and other factors. The degree to which these factors intersect to contribute to health disparities is largely unknown. Consequently, the use of self-identified race and ethnicity or genetically inferred ancestry in medical research should be well justified. For example, a study of the distribution of pathogenic variants in diverse populations to improve the application of precision therapeutics should involve a different set of definitions and metrics to a study attempting to understand access to healthcare. The former should apply reference panels of genetic variants to determine genetically inferred ancestry, while the latter should use standardized categories of self-identified race and ethnicity. A clear a priori definition and rationale for the use of self-identified race and ethnicity or genetically inferred ancestry metrics should be provided, with appropriate discussion regarding the assumptions and limitations of the chosen concepts.

Precision medicine offers unique opportunities to improve health outcomes, but the Eurocentric focus of such efforts could exacerbate disparities. Manrai et al.[11] reported rare genetic variants associated with cardiomyopathies that were proven as non-pathogenic only when African ancestral populations were included in the analysis, demonstrating that diverse data benefit both majority and minority populations. Discovery research similarly benefits from maximal variation, as genetic contributors to health disparities cannot be addressed without representation of diverse populations.

Although self-identified race and ethnicity groups reflect social constructs, some biological and genomic differences are likely to be enriched in certain groups, which may have clinical implications for disease prevention and management. For example, prostate cancer displays a more than two-fold increase in mortality in African American men compared to European American men. In addition to differences in screening, treatment and care access, differences in prostate tumor mutational signatures and distribution of actionable mutations (that is, those that inform precision oncology and have prognostic or treatment implications) were also observed[12–14]. These data suggest that molecular features may inform our understanding of aggressive disease and provide pathways to precision screening and treatment approaches. As more data emerge, however, critical questions remain about what self-identified race and ethnicity represents when it comes to influences on disease incidence and/or aggressiveness, such as the impact of non-genetic exposures, lifestyles and lived experience (including stress from racism) on tumor initiation and aggressiveness. Knowledge of genomic and sociocultural determinants that influence disease etiology is necessary to develop and deploy interventions that will mitigate disparities.

Treat underlying causes

Both genetic variation and sociocultural factors contribute to disease etiology and health disparities, and so interventions are needed that optimally target these underlying causes. Interventions can be tailored around the characteristics of the individual or group, using multilevel risk-stratified intervention methods that reclassify risk at a baseline assessment or risk-adaptive intervention methods that allow for updated dynamic stratification based on continuously changing characteristics[15]. These approaches require knowledge of the specific risk source, the magnitude of any effects, and an understanding of the setting in which an intervention may be applied.

While determinants of health and disparities are often multifactorial, they may be predominantly biological or social in nature. For example, cancer development in women who have inherited *breast cancer gene 1* or *2* (*BRCA1*/*BRCA2*) pathogenic variants is overwhelmingly genetic. Thus, the most appropriate means to limit mortality in these individuals is to address the genetic underpinnings of disease. It is appropriate to recommend that women undertake extreme measures such as salpingo-oophorectomy to address their exceptionally high cancer risk. Knowledge of common or founder pathogenic variants

Table 1 | Key concepts of race and ancestry

	Race	Ancestry
Synonyms	Racialism, racialization, racial formation	Genealogical ancestry, genetic ancestry, genetic similarity
Proposed terminology and definition	Self-identified race and ethnicity: a personal identifier based on shared culture, history, background, physical features and/or lived experiences	Genetically inferred ancestry: inherited DNA that comprises an individual's genome inclusive of common and rare variations that can be linked to populations with common geographic origins
Determined by	The individual	The individual's genome
Correlates relevant to health and disease	Systemic racism; historical marginalization and discrimination; socioeconomic status; healthcare access; cultural beliefs, norms and preferences	Population and evolutionary genetics, likely interacting with non-genetic influences

that identify specific populations at risk (such as Ashkenazi Jews, Icelanders and others) has facilitated genetic testing and disease management. With the evidence that *BRCA1*/*BRCA2* pathogenic variants may have different frequencies in different self-identified race and ethnicity groups[16], molecularly driven precision approaches may be required to ultimately reduce cancer risks and disparities.

In contrast, when disease etiology is primarily a function of exposure, behavior or sociocultural factors, interventions may be best approached with limited or no consideration of ancestry or genetics. For example, the epidemic of menthol cigarette use in African Americans is the result of economic, social and policy decisions. This disparity, with a root cause of systemic racism, may be best addressed through policy change.

Most health disparities cannot be effectively addressed until the relative magnitude of importance of genetic, ancestral, sociocultural, lifestyle and other factors is understood. Knowledge of the major contributors of disease etiology, and how these differ by genetically inferred ancestry and/or self-identified race and ethnicity, will identify the most promising interventional approaches. Heterogeneity of disease risk and treatment response are also poorly understood. Population strata may exist in which a more biologically driven intervention is warranted, while stratum behavioral or policy interventions will be more effective in another population. For example, while many smokers benefit from behavioral interventions to achieve smoking cessation, a subset of smokers have genomic traits that impact biology and render them amenable to nicotine replacement or other medical therapies. A tailored behavioral approach for some, and a tailored biological approach for others, may be warranted.

Novel tools and technologies also hold great promise to improve screening, diagnosis or treatment, but the impact of these technologies is often not equally applied to all groups. For example, mammography for breast cancer early detection is widely available yet has not benefitted all populations equally. Mammographic features of African American women differ from those of European American women and access is unequal across self-identified race and ethnicity groups. Thus, breast cancer mortality in European American women has dropped with widespread use of mammography, while breast cancer mortality in African American women has not[17]. A combination of factors that differ by self-identified race and ethnicity may be at play, including ancestry, age, lifestyle and screening access. Importantly, research focused on genomics and novel technologies must not divert resources away from approaches that address the consequences of systemic racism. Rather, knowledge of genomics and classification tools should complement existing population health and policy-based strategies — as long as they are accessible to all.

Systemic racism

Metrics such as self-identified race and ethnicity and genetically inferred ancestry are frequently used to categorize individuals, but are often poorly defined and are likely too simplistic to achieve meaningful clinical utility when evaluating etiologies of health disparities. The lack of clear and consistent definitions of concepts may therefore cause more harm than benefit. Given that the underlying concepts of race and ancestry are often ambiguous and highly nuanced, it is critical that they are well defined and appropriately used (Table 1).

There is ample evidence that persistent societal inequities are the primary cause of racial disparities, but ancestral genetic factors should still be evaluated. Indeed, a combination of elements likely determine disease aggressiveness, risk of progression and other outcomes. Studies of health disparities must consistently consider the effects of systemic racism manifest in our social, educational, economic, policy and health systems acting on individuals who may also have variable genetic propensities for health and disease. More precise knowledge of specific genetic and non-genetic factors that underlie risk and disease biology will aid in the discovery of novel interventions and improve health equity for all. ❐

Timothy R. Rebbeck[1]✉, Brandon Mahal[2], Kara N. Maxwell[3], Isla P. Garraway[4,5] and Kosj Yamoah[6]

[1]*Harvard TH Chan School of Public Health and Dana-Farber Cancer Institute, Boston, MA, USA.* [2]*University of Miami Sylvester Cancer Center, Miami, FL, USA.* [3]*Department of Medicine and Abramson Cancer Center, University of Pennsylvania School of Medicine and Michael Crescenz VA Medical Center, Philadelphia, PA, USA.* [4]*Department of Urology and Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, USA.* [5]*VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA.* [6]*H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA.*

✉*e-mail: timothy_rebbeck@dfci.harvard.edu*

Published online: 09 May 2022

https://doi.org/10.1038/s41591-022-01796-1

References

1. Goldberg, D. *Ethn. Racial Stud.* **15**, 543–569 (1992).
2. Barot, R. et al. *Ethn. Racial Stud.* **24**, 601–618 (2001).
3. Hsu, C. Y. et al. *N. Engl. J. Med.* **385**, 1750–1760 (2021).
4. Mathieson, I. et al. *PLoS Genet.* **16**, e1008624 (2020).
5. Mersha, T. B. et al. *Hum. Genomics* **9**, 1 (2015).
6. Braveman, P. et al. *Front. Public Health* **9**, 689462 (2021).
7. Templeton, A. *Stud. Hist. Philos. Biol. Biomed. Sci.* **44**, 262–271 (2013).
8. Omi, M. et al. *Racial Formation in the United States* 3rd edn. (Routledge, 2014).
9. Fuentes, A. et al. *Am. J. Phys. Anthropol.* **169**, 400–402 (2019).
10. Office of the Assistant Secretary for Planning and Evaluation. *HHS-DATA* https://aspe.hhs.gov/collaborations-committees-advisory-groups/hhs-data/hhs-data-council-introduction/dc-archive/policy-statement-inclusion-race-ethnicity-dhhs-data-collection-activities (2021).
11. Manrai, A. K. et al. *N. Engl. J. Med.* **375**, 655–665 (2016).
12. Awasthi,, S. et al. *Clin. Cancer Res.* **27**, 320–329 (2021).
13. Mahal, B. A. et al. *N. Engl. J. Med.* **383**, 1083–1085 (2020).
14. Cooperberg, M. R. et al. *Eur. Urol.* **74**, 444–452 (2018).
15. Rebbeck, T. R. et al. *Cancer Discov.* **8**, 803–811 (2018).
16. Friebel, T. M. et al. *Hum. Mutat.* **40**, 1781–1796 (2019).
17. DeSantis, C. E. et al. *CA Cancer J. Clin.* **69**, 438–451 (2019).

Acknowledgements

This work was supported in part by Department of Defense award CDMRP-PC181013 and Prostate Cancer Foundation (K.Y.); National Cancer Institute grant P20-CA233255

(K.Y. and T.R.R.); Jean Perkins Foundation, STOP Cancer Foundation, National Cancer Institute and Specialized Programs of Research Excellence grant P50 CA092131 and Prostate Cancer Foundation Challenge Award 17CHAL04 (I.P.G.); and National Institutes of Health (K08CA215312), Burroughs Wellcome Foundation (1017184), Basser Center for BRCA at the University of Pennsylvania, and the Prostate Cancer Foundation (K.M.).

Author contributions

All authors were involved in the development of the concepts presented in this Comment and contributed to writing and editing of the content.

Competing interests

T.R.R.'s wife is a consultant to AstraZeneca. All other authors declare no competing interests.