Measurement of Health Equity in Public Health Investigations:

Genetic Ancestry, Social Determinants of Health, and Race/Ethnicity

Stuart K. Shapira, MD, PhD
National Center on Birth Defects and Developmental Disabilities
CDC

CDC

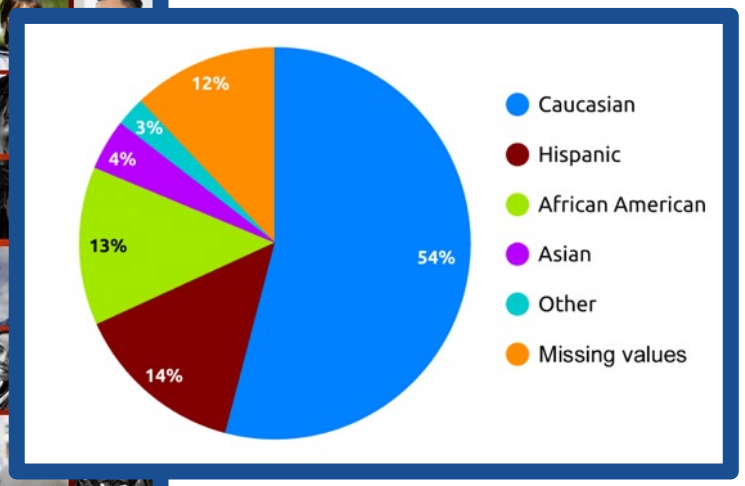
Oak Ridge Institute For Science and Education

Race/Ethnicity and Health Outcomes

- Epidemiologic analyses, clinical studies, and randomized controlled trials often include race/ethnicity to identify health disparities or differences in outcomes
- Identifying health disparities may lead to improved interventions or therapeutic options that address disparities and improve health

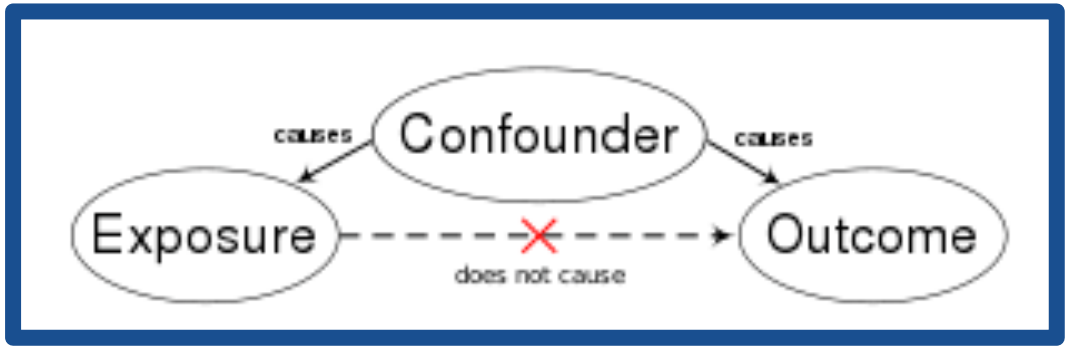
Race and Ethnicity

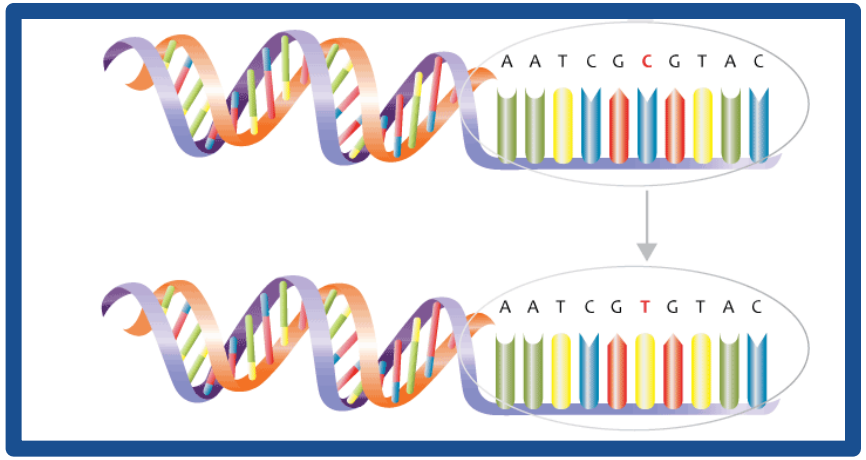
- Race and ethnicity are social constructs
- Race describes population groups assumed to have different biogeographical ancestries or genetic makeup
- Ethnicity reflects geographical origins, historical influences, as well as shared customs, beliefs, and traditions among populations that may or may not have common genetic origins



Race and Ethnicity

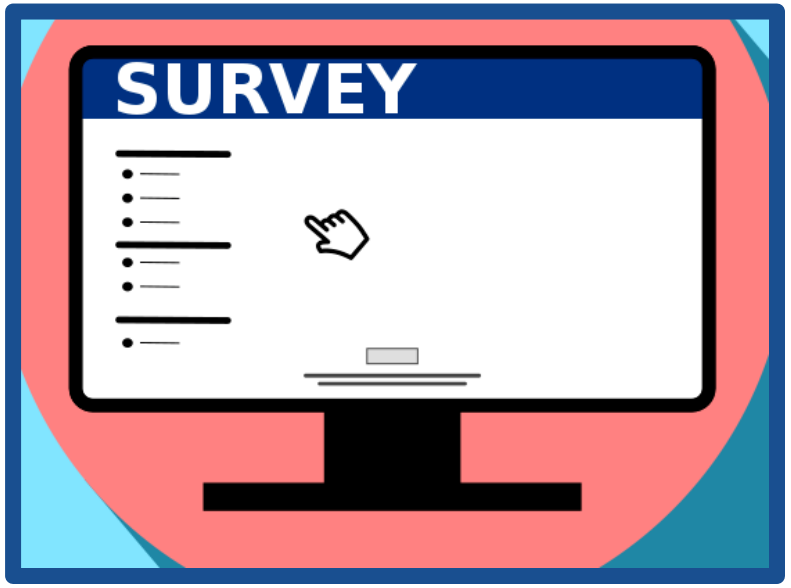
- Why collect and analyze race and ethnicity?
  - Adjustment for potential confounding
  - Assessment of disparities in health outcomes, although interpretation of differences is challenging given myriad of potential explanations
    - Socioeconomic status and access to care
    - Environmental exposures and stressors
    - Behavioral factors
    - Structural racism
    - Host genetics (genomic variants)





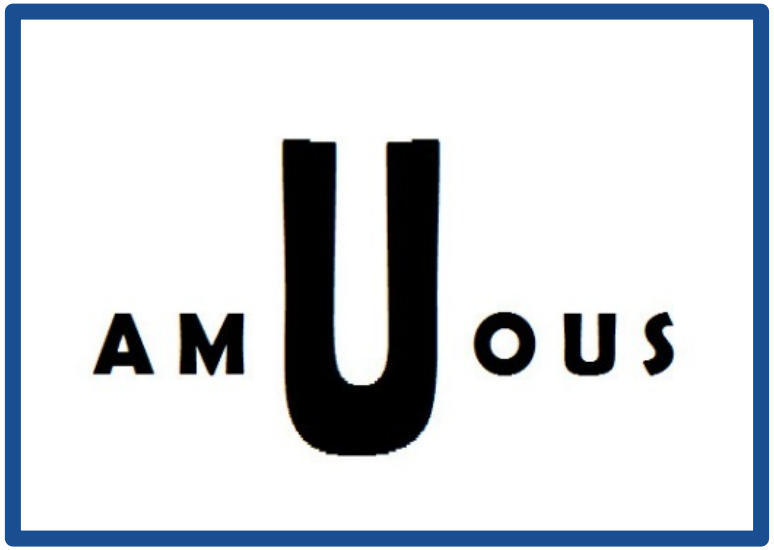
Race and Ethnicity

- Who reports race and ethnicity?
  - Self-report (e.g., in surveys)
    - Often considered as “gold standard”
  - Identification by third parties (common in administrative databases)
    - Hospital and clinic staff
    - Funeral directors may classify people based on appearance, name, language spoken by family, etc.



Race and Ethnicity

- Is self-reported race/ethnicity a gold standard?
  - More accurate than reported by third parties, but can be ambiguous
  - People can identify with multiple groups
  - Official race/ethnicity categories may not correspond to how people self-identify and some people may not identify with any of the official categories (e.g. Middle Eastern/North African)
  - Reported identity may vary depending on the context
  - Self-reported identity may change over time (e.g., reporting to U.S. census)

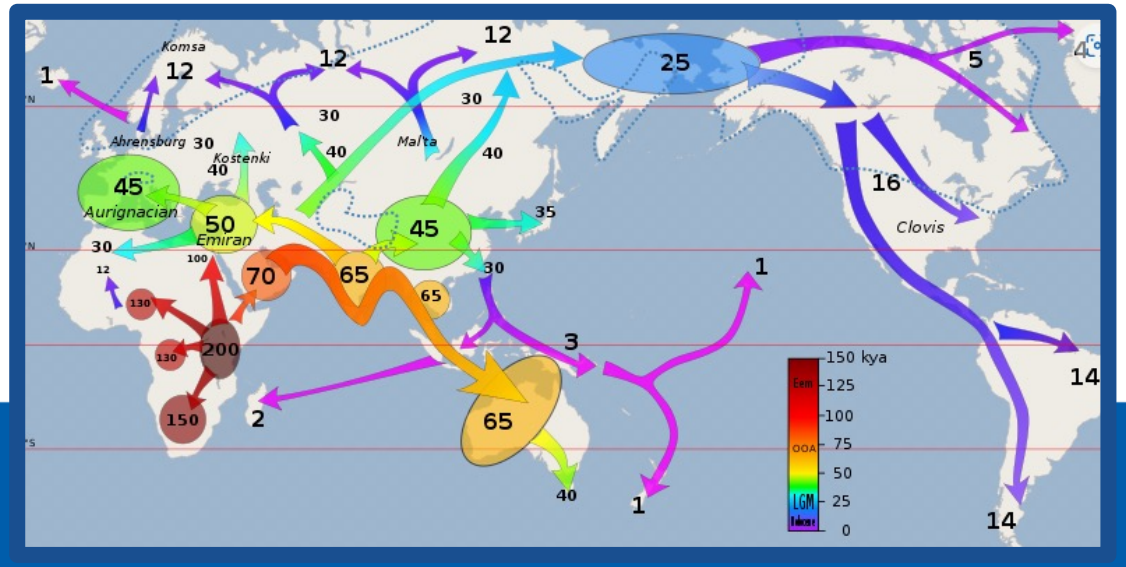


Race/Ethnicity and Health Outcomes

- Is self-reported race/ethnicity the optimal construct to assess differences in biologic outcomes and interventions?
- Or does genetic ancestry and evaluation of genetic admixture provide more useful information about the risk for diseases, disorders, and biologic outcomes?

Genetic Ancestry

- Modern humans left Africa about 60-70 thousand years ago in waves of migrations
- Local communities were formed, separated by geographic, linguistic, cultural, and social barriers
- Genetic variation, genetic drift*, and natural selection led to patterns of human variation resulting in genetic ancestry and population structure carried in the genomes of individuals and groups



*<u>Genetic Drift</u>: Change in frequency of a gene variant in a population due to random chance.

Genetic Ancestry and Admixture

- **Genetic ancestry is a genetics concept that describes the architecture of genome variation between populations**
- **Genetic ancestry refers to the part of the world from which an individual's ancestors originally came, so it might also be thought of as geographic ancestry**
- **Genetic admixture occurs when previously diverged or isolated genetic ancestral lineages mix (i.e., have children together)**
- **Individuals and groups in admixed populations can have quite different proportions of ancestral DNA** (Example: Percent European ancestry contribution to African American genomes in Gullah-speaking Sea Islanders (SC)—3.5%; and in Seattle—35%)

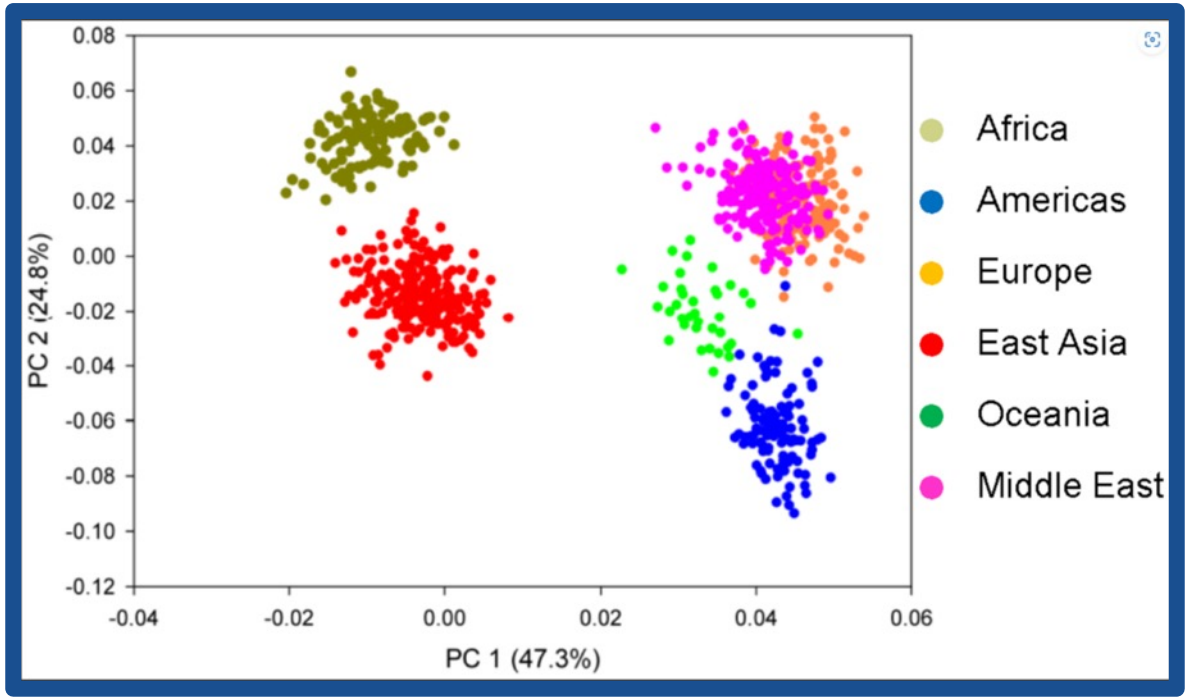
Genetic Ancestry and Informative Markers

- Genetic ancestry informative markers (AIMs) are sets of genetic variants for particular DNA sequences that occur in different frequencies in populations from different regions of the world

Example: Variant is G/A in the *SLC24A5* gene
yellow=%G
blue=%A
This SNP is commonly included in AIM panels

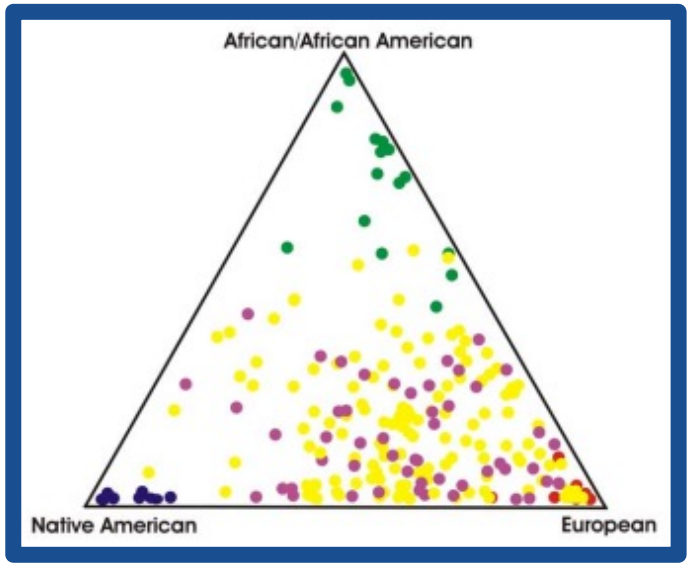
Genetic Ancestry and Informative Markers

- **An individual's genetic variants are compared using statistical methods to markers with previously analyzed genomic reference sets from individuals with fairly well-known ancestral history**



Genetic Ancestry and Informative Markers

- AIMs are used to estimate the geographical origins of an individual's ancestors, usually expressed as proportion of one's ancestry that comes from different continental regions
- Admixture can also be estimated for groups in a study

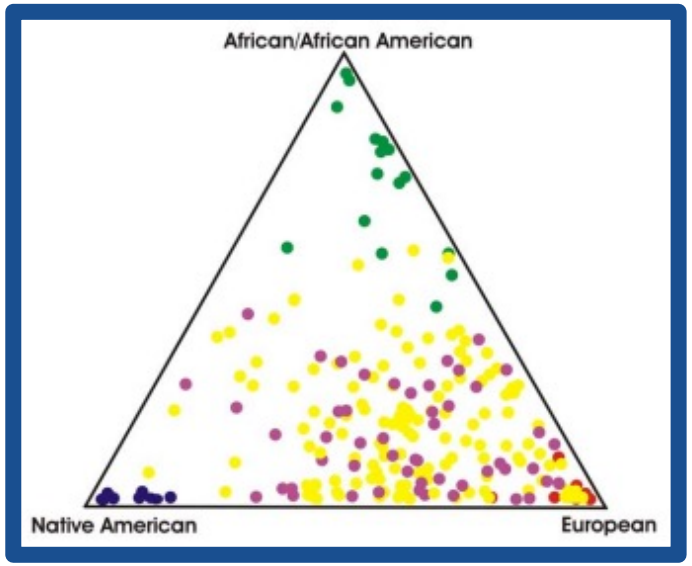


Cases=magenta; Controls=yellow
Average admixture among cases: 0.18 African American; 0.54 European; 0.28 Native American
Average admixture among controls: 0.18 African American; 0.54 European; 0.28 Native American

Tarazona-Santos E., et al. Population Genetics of GYPB and Association Study between GYPB*S/s Polymorphism and Susceptibility to *P. falciparum* Infection in the Brazilian Amazon. *Plos One*. 2011;6(1):e16123.

Genetic Ancestry and Informative Markers

- Care must be taken in utilizing genetic ancestry in case-control studies
  - Risk for population stratification bias when cases and controls are sampled from genetically different underlying populations
  - Therefore, any observed associations are more likely due to sampling differences rather than due to the disease of interest



Cases=magenta; Controls=yellow
Average admixture among cases: 0.18 African American; 0.54 European; 0.28 Native American
Average admixture among controls: 0.18 African American; 0.54 European; 0.28 Native American

Tarazona-Santos E., et al. Population Genetics of GYPB and Association Study between GYPB*S/s Polymorphism and Susceptibility to *P. falciparum* Infection in the Brazilian Amazon. *Plos One*. 2011;6(1):e16123.

Race, Ethnicity, and Genetic Ancestry

- **Do self-reported race and ethnicity correlate with genetic ancestry?**
- Lee YL, et al. Comparing Genetic Ancestry and Self-reported Race/Ethnicity in a Multiethnic Population in New York City. *J Genet*. 2010;89(4):417-423.

--Estimated proportions of genetic admixture in a genetically diverse population with 35 AIMs

--Among those self-reported as Black or White, the imputed ancestry proportions were 77.6% African and 75.1% European, respectively

--Among those self-reported as Hispanic, imputed ancestry proportions were 29.2% European, 26.0% African, and 44.8% Native American

--Conclusions:

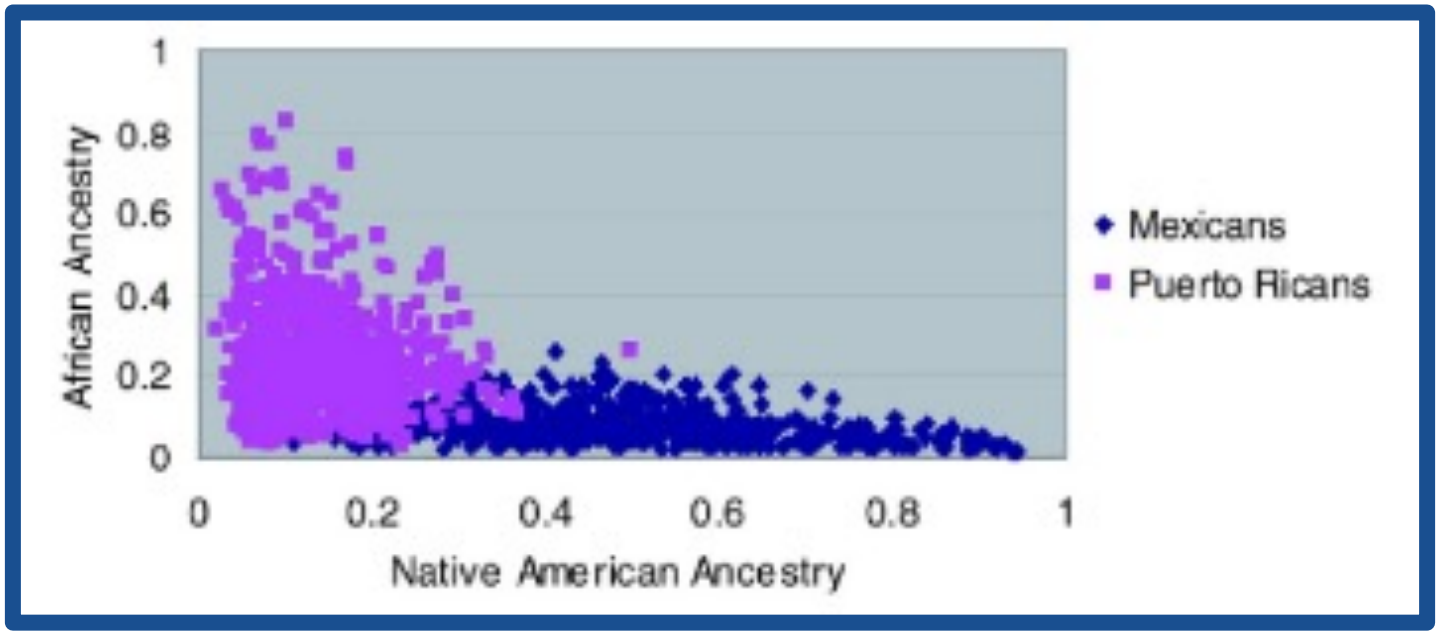
--Reliable estimates of ancestry proportion in individuals self-reported as White or Black

--Multiracial makeup of Hispanic participants reflected their heterogeneous origins

--Use of a single "Hispanic" category may be insufficient for characterizing genetic background in epidemiologic or other analyses

Race, Ethnicity, and Genetic Ancestry

- Do self-reported race and ethnicity correlate with genetic ancestry?



--Use of a single "Hispanic" category may be insufficient for characterizing genetic background in epidemiologic or other analyses

Race, Ethnicity, and Genetic Ancestry

- **Do self-reported race and ethnicity correlate with genetic ancestry?**
- Banda Y, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics*. 2015;200:1285-1295.

--103,006 participants in the Kaiser Northern California GERA reported any of 23 race/ethnicity/nationality categories that were collapsed into 7 major racial/ethnic groups

--144,799 high-performing SNPs were used in analysis of population structure and admixture

--Self-reported White: broad spectrum of genetic ancestry ranging from northern to southern Europe and the Middle East; except for Ashkenazi Jews, there weren't distinct clusters

--Self-reported East Asian: genetic ancestry correlated with nationality, although 10% were admixed with European ancestry

--Self-reported African American: most had some degree of European genetic ancestry (average of 26%)

Race, Ethnicity, and Genetic Ancestry

- **Do self-reported race and ethnicity correlate with genetic ancestry?**
- Banda Y, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics*. 2015;200:1285-1295.

--Self-reported Hispanic: most had varying proportions of Native American, European, and African genetic ancestry

--Conclusion: By and large, there was high correspondence between self-reported race/ethnicity and genetic ancestry, although there was a broad range of genetic ancestry among African Americans and Hispanics

Race, Ethnicity, and Genetic Ancestry

- **Does genetic ancestry vs. self-reported race/ethnicity matter in studies of biologic outcomes?**
- Louwers YV, et al. The Impact of Self-Reported Ethnicity Versus Genetic Ancestry on Phenotypic Characteristics of Polycystic Ovary Syndrome. *J Clin Endocrinol Metab*. 2014;99(10):E2107-E2116.

--1499 patients with polycystic ovary syndrome (PCOS) from the Netherlands comprised 11 self-reported racial/ethnic groups and genetic analysis revealed 8 genetic ancestry clusters

--Age, self-reported race/ethnicity, and genetic ancestry cluster were included as independent variables in regression models of PCOS characteristics

--Association between self-reported race/ethnicity and genetic ancestry was moderate

--However, final models included mainly genetic ancestry clusters because they explained a larger proportion of phenotype variability for amenorrhea, total follicle count, BMI, SHBG, and DHEA sulfate; insulin levels were predominantly explained by genetic ancestry alone

--Conclusion: Self-reported race/ethnicity is not a perfect proxy for genetic ancestry in patients with PCOS, so using genetic ancestry data can avoid PCOS-relevant misclassification

Race, Ethnicity, and Genetic Ancestry

- **Does genetic ancestry vs. self-reported race/ethnicity matter in studies of biologic outcomes?**
- Lee KK, et al. Association of Genetic Ancestry and Molecular Signatures with Cancer Survival Disparities: a Pan-cancer Analysis. *Cancer Res*. 2022;82(7):1222-1233.

--8010 study participants had self-reported race/ethnicity as either White, Black, Asian, or Hispanic; concordance with genetic ancestry was high >90% for all groups, except Hispanic (68% concordance)

--Univariable and multivariable cancer survival analyses were performed, modeling the effects of self-reported race/ethnicity and genetic ancestry on survival and survival time for 33 cancers

--Observable differences in cancer survival disparities between self-reported race/ethnicity and genetic ancestry with significant differences for 4 cancer types involving breast, head/neck, kidney, and skin.

--Uncovered various molecular mechanisms through which genetic ancestry might impact cancer survival disparities, revealing potential population-specific therapeutic targets for groups disproportionately burdened by cancer

Socioeconomic Deprivation

- Multidimensional concept referring to the relative disadvantage an individual or a social group experiences in terms of access and control over economic, material, or social resources and opportunities
- Complex construct, difficult to operationalize because it involves many aspects that pertain to the economic and material domain, such as employment, income, education, occupation, housing, as well as the social domain, such as social position, family support, and social integration
- At the level of the individual, either the classic triad of education, occupation and income and/or subjective measures of an individual's own perceived position on the social ladder, are used to operationalize a person's socioeconomic status
- Composite indices of relative deprivation have also been described

Race, Ethnicity, Socioeconomic Deprivation and Genetic Ancestry

- **What is the role of socioeconomic deprivation in studies of biologic outcomes?**
- Marino-Ramirez L, et al. Effects of Genetic Ancestry and Socioeconomic Deprivation on Ethnic Differences in Serum Creatinine. *Gene*. 2022;837:146709.

--Inclusion of race/ethnicity in the equations to estimate kidney glomerular filtration rate (eGFR) from serum creatinine is controversial

--35,590 participants in the UK Biobank were grouped by self-reported race/ethnicity, genetic ancestry, and socioeconomic deprivation (SED) using a composite metric that included unemployment, non-car ownership, non-home ownership, and household overcrowding

--Multivariable modeling tested for associations between self-reported race/ethnicity, genetic ancestry, SED, and serum creatinine levels

--Overall, genetic ancestry explained significantly more of the variation in creatinine levels than self-reported race/ethnicity; genetic ancestry explained more of the variation in serum creatinine than SED, suggesting that racial/ethnic differences in creatinine are possibly shaped much more by genetic than social factors

Race, Ethnicity, Socioeconomic Deprivation and Genetic Ancestry

- **What is the role of socioeconomic deprivation in studies of biologic outcomes?**

- Nagar SD, et al. Comparing Genetic and Socioenvironmental Contributions to Ethnic Differences in C-reactive Protein. *Front Genet*. 2021;12:738485.

--C-reactive protein (CRP), a blood biomarker for inflammation, is associated with response to infection, risk for a number of complex common diseases, and psychosocial stress; Blacks have higher average CRP than Whites

--Study of UK Biobank participants evaluated contributions of genetic ancestry, socioenvironmental factors and inflammation-related heath conditions to self-reported racial/ethnic differences in CRP levels

--Significant associations observed between CRP, self-reported race/ethnicity and genetic ancestry; associations almost completely attenuated in models including socioenvironmental factors and inflammation-related health conditions; BMI, smoking, and SED all have high relative effect on CRP; results indicate that socioenvironmental factors contribute more than genetics to CRP racial/ethnic differences

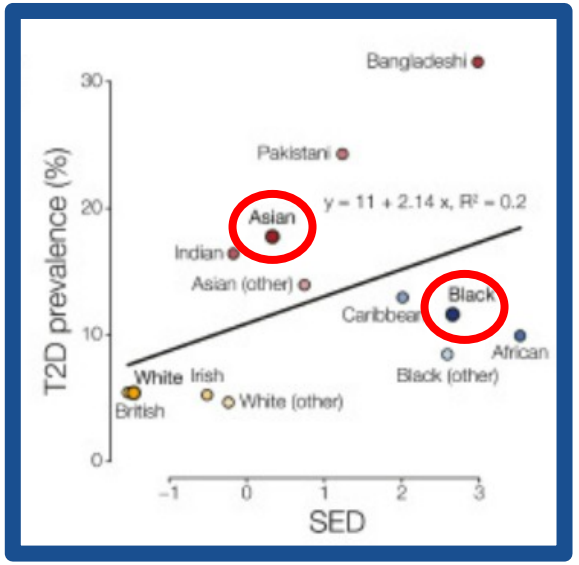
Race, Ethnicity, Socioeconomic Deprivation and Genetic Ancestry

- **What is the role of socioeconomic deprivation in studies of biologic outcomes?**
- Nagar SD, et al. Socioeconomic Deprivation and Genetic Ancestry Interact to Modify Type 2 Diabetes Ethnic Disparities in the United Kingdom. *EClinicalMedicine*. 2021;37:100960.

--Type 2 diabetes (T2D) is a complex disease that disproportionately impacts minority racial/ethnic groups in the UK; socioeconomic deprivation (SED) is widely considered a potential explanation for the racial/ethnic disparities

--Case and control cohorts from the UK Biobank were utilized to model the relationship between SED, genetic ancestry and T2D status using multivariable logistic regression

--Self-reported Asian race/ethnicity had highest prevalence of T2D, followed by Black and White groups; the disparity was not completely explained by SED (Blacks had higher average SED and lower T2D prevalence compared to Asians with lower average SED but higher T2D prevalence)



Race, Ethnicity, Socioeconomic Deprivation and Genetic Ancestry

- **What is the role of socioeconomic deprivation in studies of biologic outcomes?**

- Nagar SD, et al. Socioeconomic Deprivation and Genetic Ancestry Interact to Modify Type 2 Diabetes Ethnic Disparities in the United Kingdom. *EClinicalMedicine*. 2021;37:100960.

--However, SED and non-European genetic ancestry were significantly associated with the observed T2D disparities; genetic ancestry and SED showed significant interaction effects on T2D with SED being a greater risk factor for the non-European ancestry groups

--The approach to form discrete non-overlapping genetic ancestry groups was useful to help disambiguate socioenvironmental factors from genetic factors that might contribute to observed T2D racial/ethnic disparities

--The significant interactions between SED and genetic ancestry underscore how the effects of environmental risk factors can differ among genetic ancestry groups, suggesting the possible need for group-specific interventions

Summary

- Studies that incorporate self-reported race/ethnicity can identify health disparities and differences in health outcomes
- However, understanding the basis of disparities may require more in-depth assessment of genetic and socioenvironmental factors
- Genetic ancestry, as a proxy for genetic factors, may be useful for teasing out relationships with biologic outcomes
- Assessing both socioeconomic factors and genetic factors and their interaction has the potential to inform targeted group-specific interventions to help decrease health disparities


EDICINE
a.bacall
"Your weight problem is partly genetic and partly Boston Cream pie."

Stuart K. Shapira, MD, PhD
Associate Director for Science
National Center on Birth Defects and Developmental Disabilities
Centers for Disease Control and Prevention
E-mail: sshapira@cdc.gov

The findings and conclusions in this presentation have not been formally disseminated by the Centers for Disease Control and Prevention and should not be construed to represent any agency determination or policy.