

# Using Population Descriptors in Genetics and Genomics Research

*A New Framework For An Evolving Field*

*Aravinda Chakravarti, Committee Co-chair*

*Charmaine Royal, Committee Co-chair*

**Presentation to NHGRI Council**

**May 15, 2023**

# Statement of Task

The committee convened to review and assess existing methodologies, benefits, and challenges in the use of race and ethnicity and other population descriptors in genomics research. The committee's work included:

- Assessing use of race, ethnicity, and genetic ancestry in the basic science of genetics and genomics, health risk as a function of our genomes, and health disparities
- Developing approaches to advance the appropriate use of population descriptors and proposing best practices
- Discussing obstacles to implementation and adoption of best practices for population descriptor use
- Proposing potential implementation strategies to help enhance the adoption of best practices by the research community
- *Out of scope:* use of race and ethnicity in clinical care and biomedical research generally; racism in science and genomics; providing policy recommendations to NIH and government agencies

# Committee Membership

**Aravinda Chakravarti, Ph.D.**  
**(Co-Chair) *NAM/NAS***

New York University Grossman  
School of Medicine

**Charmaine Royal, Ph.D. (Co-Chair)**

Duke University

**Katrina Armstrong, M.D. *NAM***

Columbia University

**Michael Bamshad, M.D.**

University of Washington & Seattle  
Children's Hospital

**Luisa Borrell, Ph.D., D.D.S., M.P.H.**

City University of New York, NY

**Katrina Claw, Ph.D.**

University of Colorado Denver – Anschutz  
Medical Campus

**Clarence Gravlee, Ph.D.**

University of Florida

**Mark Hayward, Ph.D.**

The University of Texas at Austin

**Rick Kittles, Ph.D.**

Morehouse School of Medicine

**Sandra (Soo-Jin) Lee, Ph.D.**

Columbia University

**Andrés Moreno-Estrada, M.D.,  
Ph.D.**

Centro de Investigación y de Estudios  
Avanzados del Instituto Politécnico  
Nacional

**Ann Morning, Ph.D.**

New York University

**John Novembre, Ph.D.**

University of Chicago

**Molly Przeworski, Ph.D. *NAS***

Columbia University

**Dorothy Roberts, J.D. *NAM***

University of Pennsylvania

**Sarah A. Tishkoff, Ph.D. *NAM/NAS***

University of Pennsylvania

**Genevieve Wojcik, Ph.D.**

Johns Hopkins Bloomberg School of  
Public Health

# What is a Population Descriptor?

... a *concept* of difference or a classification scheme that categorizes individuals into groups or “populations” based on a perceived characteristic or dimension of interest. The salience of a given population descriptor has varied from place to place and over time.

Of particular interest to genetic studies are **Descent-associated Population Descriptors** which are descriptions of populations whose members are thought to share some characteristic deriving from their common origin. Human beings across the globe have devised a family of descent-associated categorization systems (e.g., *clan, caste, tribe, ancestry, ethnicity, indigeneity, race, etc.*).

# Problems with the Current Use of Population Descriptors

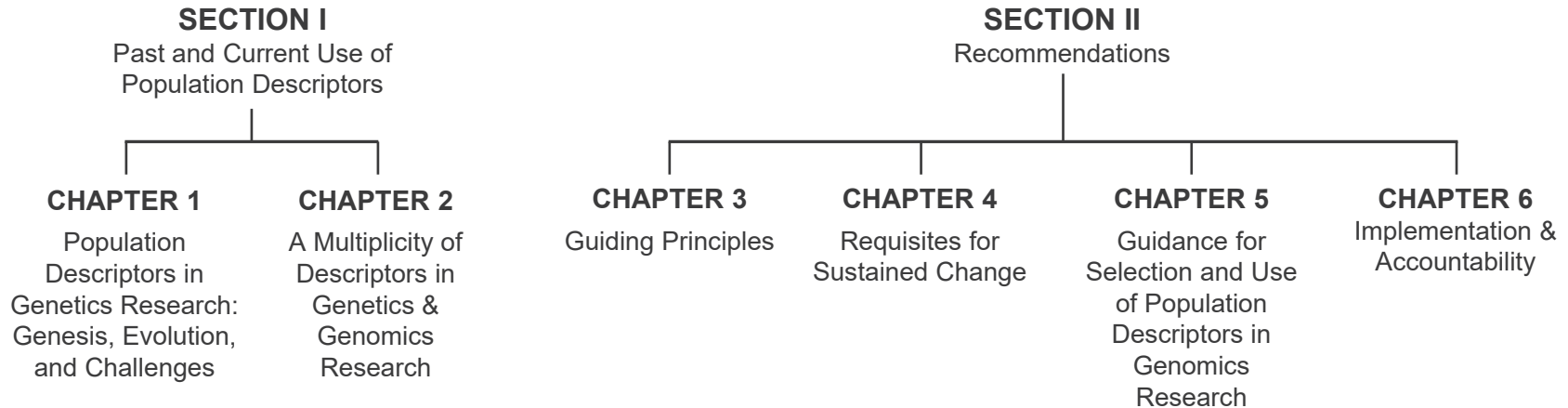
- Typological thinking, reinforcing the misconception that humans can be objectively grouped into discrete, innate categories
  - *Example:* continued use of continental ancestry groups
- Continuing use of race, a scientifically invalid measure of human genetic differences
  - Use of socially constructed groupings (e.g., OMB categories) can be harmful and erroneous for human genetics research because human genetic differences are distributed in complex ways that do not necessarily align with such social identities
- The failure to realize that population differences reflect differences in environmental exposures as well as genetic variant frequencies
- The variable and inconsistent use of population descriptors such as race and ethnicity within and across studies

# Why is this Report Needed Now?

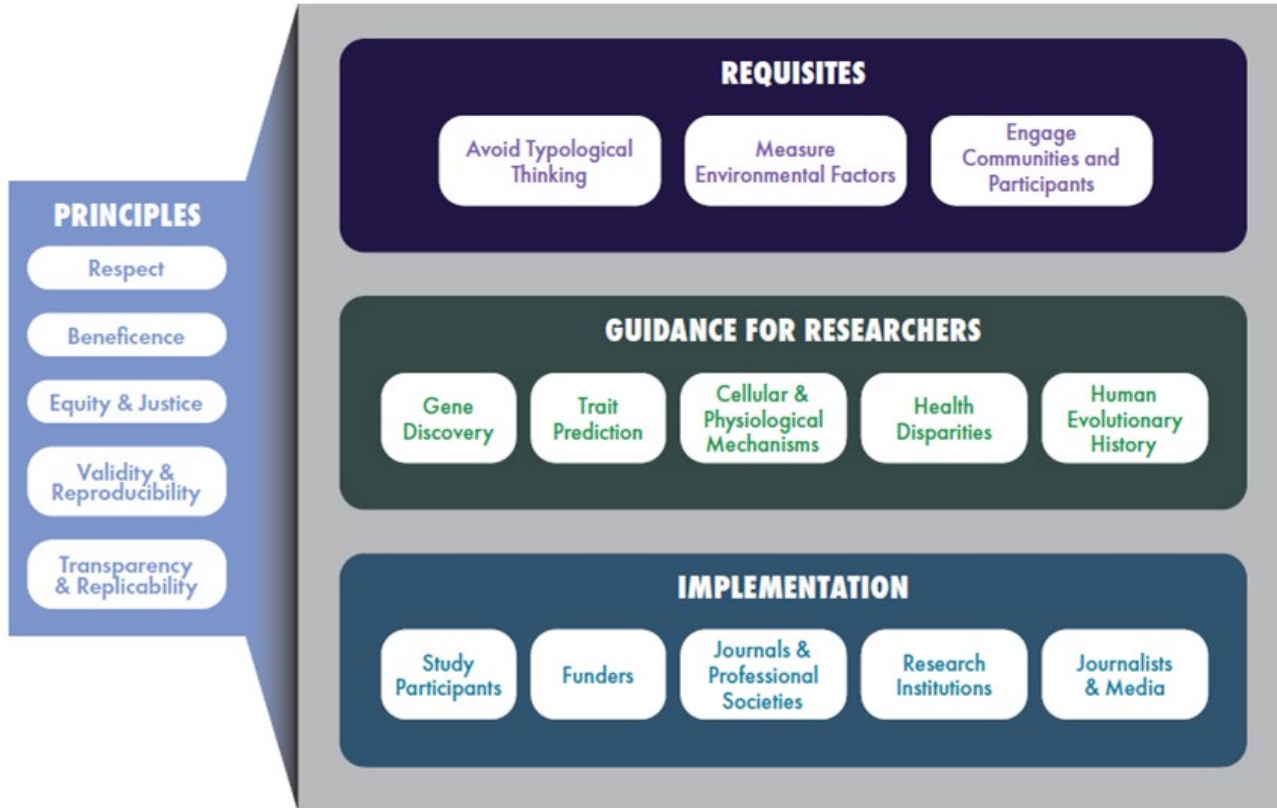
- There is an urgent need to address issues surrounding race and racism in science and society
- Genomics data have become more accessible and widespread across biomedical research
- Current large-scale genomics studies aim to include more diverse groups of people
- Population descriptors require change owing to research participants' evolving personal preferences of group affiliation
- More advanced methods of analyzing genetic variation among humans have been developed
- Genetics investigators need to be clearer about how and why they selected certain population descriptors
- Previous efforts have had limited success in changing the use of population descriptors because recommendations and their implementation plans were not comprehensive

# Using Population Descriptors in Genetics & Genomics Research

## A New Framework for an Evolving Field



# What Makes this Report Unique?





# Guiding Principles

Respect

Beneficence

Equity & Justice

Validity &  
Reproducibility

Transparency &  
Replicability

## Why guiding principles?

- The field is dynamic
- It is not possible to provide specific practices for every situation

## The guiding principles ...

- undergird all the recommendations
- offer direction for decision-making when a situation falls outside the specific recommendations and best practices
- apply broadly throughout the field

***Purpose: To foster ethically and empirically sound best practices for supporting trustworthy research***

# Overview of Recommendations

The committee developed 13 recommendations that fall into three categories

## Requisites

- **Recommendations 1-5**
- For a general audience
- Overarching approaches important for the long-term success of this effort

## Guidance for Researchers

- **Recommendations 6-8**
- 16 best practices for different types of genomics studies
- For researchers using genetics and genomics data

## Implementation & Accountability

- **Recommendations 9-13**
- For selected key players in the research ecosystem
- To support researchers implementing these recommendations and best practices

# Requisites to Sustain Change

## Avoid typological thinking

- There is a misconception that humans can be grouped into discrete, innate biological categories
- Patterns of human genetic variation are complex
- Researchers should avoid the inaccurate assumptions of typological thinking (e.g., homogeneity of groups, hierarchy)
- **See Recommendations 1-3**

## Measure environmental factors

- Virtually all phenotypes result from interplay between genetic and environmental factors
- Descent-associated population descriptors are not good proxies for environmental effects
- Researchers should use variables that more precisely capture the information needed
- **See Recommendation 4**

## Engage communities and participants

- Misperceptions about human genetic variation and group identities can have negative impacts on individuals and communities and can impede research
- Research teams should include experts in community engagement to integrate perspectives of communities throughout the research process
- **See Recommendation 5**

# Guidance for Researchers

Researchers should tailor their use of population descriptors to the type and purpose of the study

- There are many types of genetics and genomics studies
- There is no one-size-fits-all solution
- Researchers are decision-makers about how population descriptors are used in research. The report charges researchers to be active participants in deciding whether to use population descriptors and, if so, which ones
- Researchers should be transparent and report their decisions about population descriptors and group labels
- **See Recommendations 6-8**

## *Types of Genomics Studies*

Gene Discovery –  
Mendelian

Trait Prediction –  
Mendelian

Gene Discovery –  
Complex Traits

Trait Prediction –  
Complex Traits

Cellular &  
Physiological  
Mechanisms

Health Disparities

Human  
Evolutionary  
History

# Examples of Guidance for Researchers

**Race** should not be used except for a subset of health disparities studies

**Genetic similarity** is a preferred descriptor in most cases


## DEFINITIONS

**Race:** a sociopolitically constructed system for classifying and ranking human beings according to subjective beliefs about shared ancestry based on perceived innate biological similarities.


**Genetic similarity:** quantitative measure of the genetic resemblance between individuals that reflects the extent of shared genetic ancestry.




































## LEGEND

 Preferred population descriptor(s)

 Should not be used

 In some cases; refer to Ch. 5 text and the decision tree in Appendix D

 Descriptors could be used if appropriate proxies for environmental, not genetic, effects

GENOMICS STUDY TYPE	Race	Ethnicity/Indigeneity	Geography	Genetic Ancestry	Genetic Similarity	Notes
1: Gene Discovery - Mendelian Traits						Similarity suffices as a genetic measure; at fine-scale, other variables may be useful
2: Trait Prediction - Mendelian Traits						No population descriptors may be necessary for analysis
3: Gene Discovery - Complex Traits						Similarity suffices as a genetic measure
4: Trait Prediction - Complex Traits						Similarity suffices as a genetic measure
5: Cellular and Physiological Mechanisms						No population descriptors may be necessary for analysis
6: Health Disparities with Genomic Data						Not all health disparities studies rely on descent-associated population groupings, so none may be necessary for analysis
7: Human Evolutionary History						Reconstructing genetic ancestry may be of central interest

# Implementation & Accountability

The research ecosystem has many players who individually and collectively share responsibility for making changes and helping researchers implement these recommendations

Study participants

Funders of genomics  
research

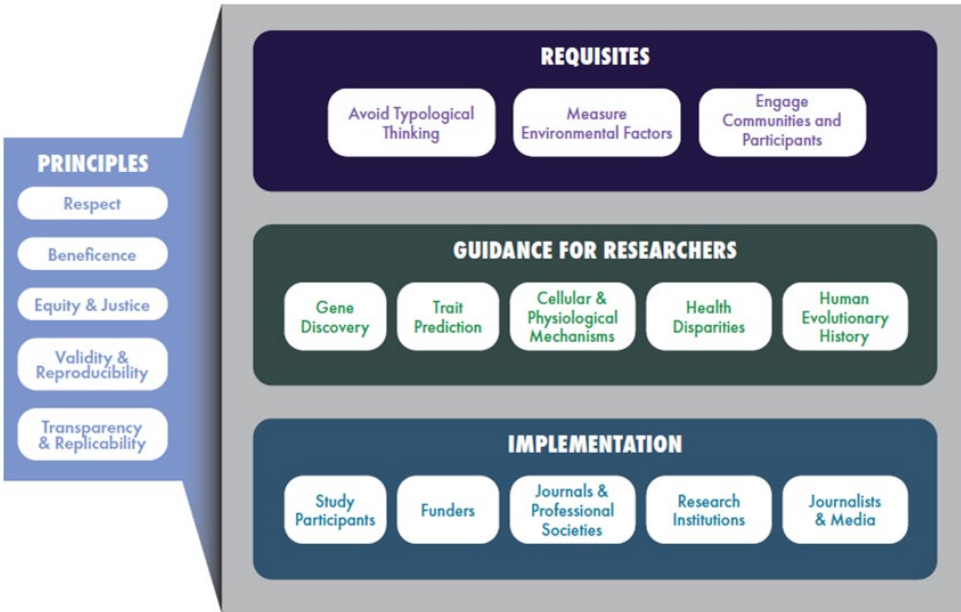
Research institutions

Journals &  
professional societies

Journalists & media

**See Recommendations 9-13**

# Parting Thoughts



- “Scientists must get the descent-associated concepts right—that is, have a clear understanding of what these descriptors represent and a rigorous rationale for using them—before selecting the appropriate group categories and labels to work with” – *Chapter 1*
- The practice of genetics today requires knowledge and expertise in many fields besides genetics & genomics – thus this report is for all researchers who use genetics & genomics data
- It will take a concerted effort by all relevant parties to achieve the desired goal of properly using and reporting population descriptors reliably and consistently

# Resources

- A free PDF of the report is available at <http://www.nap.edu/>
- Related materials are available through the study page, including



- Report Highlights
- Recommendations
- Scrolling Page with Interactive for helping investigators decide on population descriptor use in genomics studies

**Thank You!**

