Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research

Prof. Melinda Mills, University of Oxford
Director, Leverhulme Centre for Demographic Science

The National Academies of Sciences, Engineering, Medicine, 04 April 2022

COMMUNICATIONS BIOLOGY

REVIEW ARTICLE

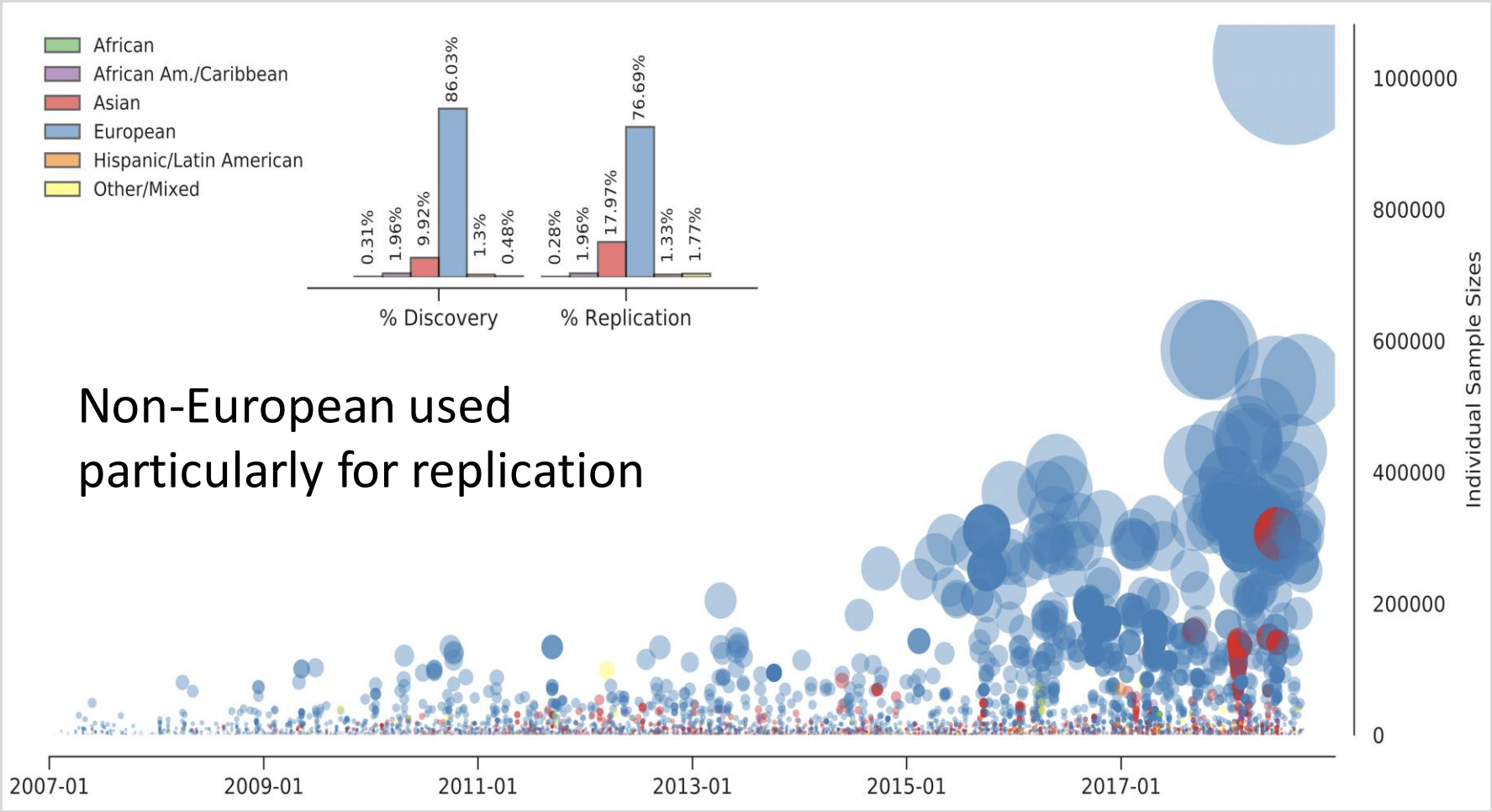
https://doi.org/10.1038/s42003-018-0261-x OPEN

A scientometric review of genome-wide association studies

Melinda C. Mills [1] & Charles Rahal [1]

This scientometric review of genome-wide association studies (GWAS) from 2005 to 2018 (3639 studies; 3508 traits) reveals extraordinary increases in sample sizes, rates of discovery and traits studied. A longitudinal examination shows fluctuating ancestral diversity, still predominantly European Ancestry (88% in 2017) with 72% of discoveries from participants recruited from three countries (US, UK, Iceland). US agencies, primarily NIH, fund 85% and women are less often senior authors. We generate a unique GWAS H-Index and reveal a tight social network of prominent authors and frequently used data sets. We conclude with 10 evidence-based policy recommendations for scientists, research bodies, funders, and editors.

80-90% of genetic discovery European ancestry



Mills & Rahal (2019) https://www.nature.com/articles/s42003-018-0261-x

Defining race, ethnicity, ancestry

Race/ethnicity socially constructed not biological category – different from ancestry in genetics

J Health Soc Behav. 2021 Jun 8;221465211018682. doi: 10.1177/00221465211018682.
Online ahead of print.

Reconstructing Sociogenomics Research: Dismantling Biological Race and Genetic Essentialism Narratives

Pamela Herd [1], Melinda C Mills [2], Jennifer Beam Dowd [2]

Affiliations + expand

PMID: 34100668 DOI: 10.1177/00221465211018682

Abstract

We detail the implications of sociogenomics for social determinants research. We focus on education and race because of how early twentieth-century scientific eugenic thinking facilitated a range of racist and eugenic policies, most of which helped justify and pattern racial and educational morbidity and mortality disparities that remain today, and are central to sociological research. Consequently, we detail the implications of sociogenomics research by unpacking key controversies and opportunities in sociogenomics as they pertain to the understanding of racial and educational inequalities. We clarify why race is not a valid biological or genetic construct, the ways that environments powerfully shape genetic influence, and risks linked to this field of research. We argue that sociologists can usefully engage in genetics research, a domain dominated by psychologists and behaviorists who, given their focus on individuals, have mostly not examined the role of history and social structure in shaping genetic influence.

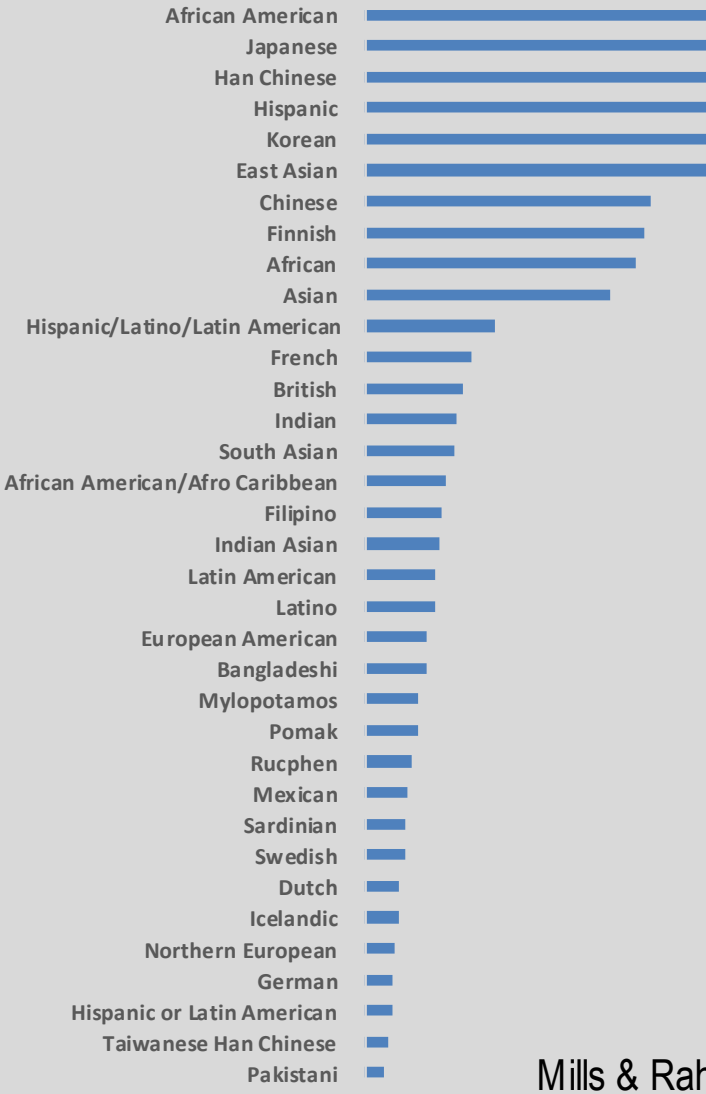
Keywords: education; gene-environment interactions; race; sociogenomics.

UNIVERSITY OF OXFORD

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

Top terms used in GWAS

Below terms after European (N=5,890 studies) removed, Mills & Rahal 2019



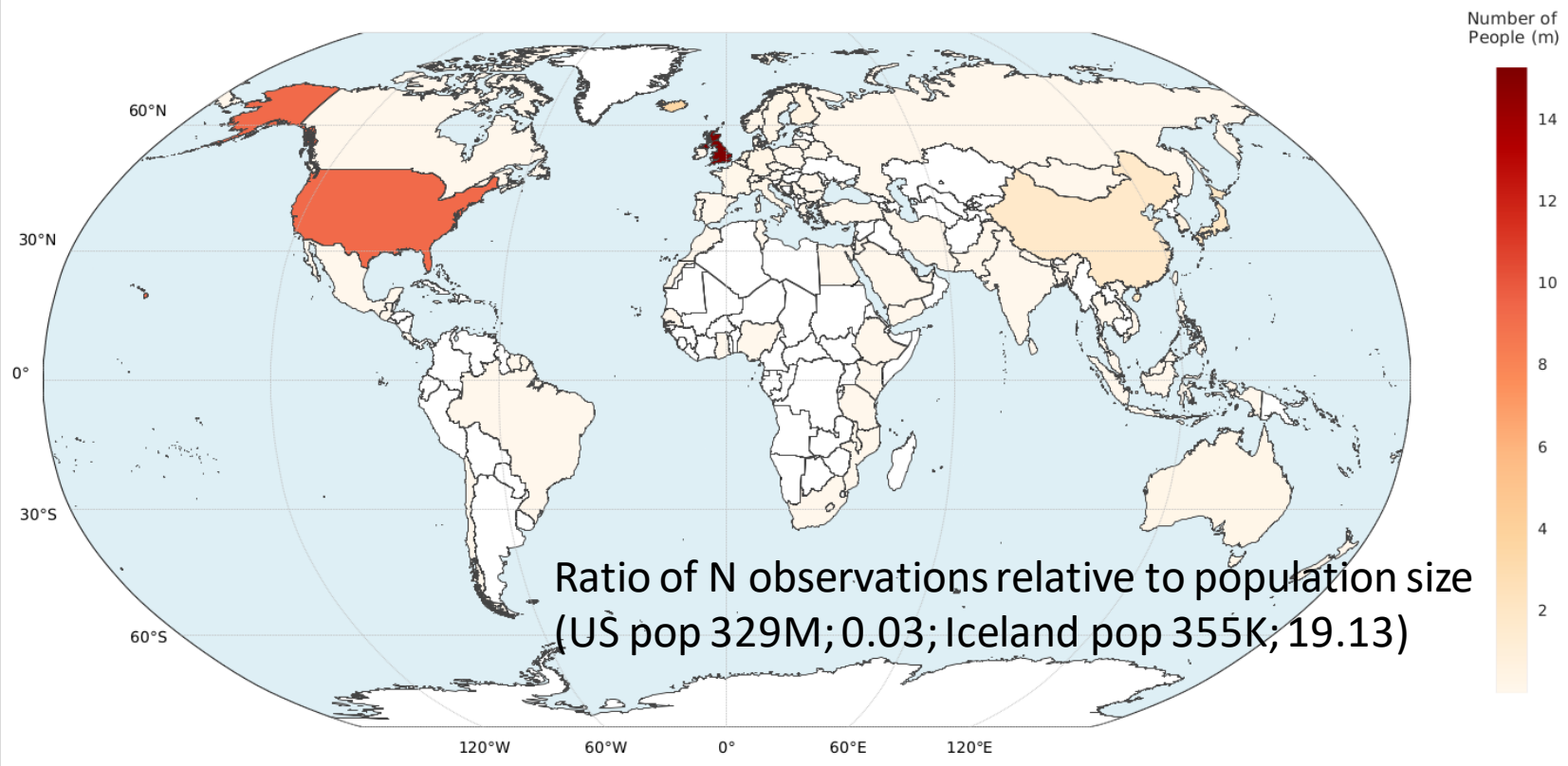
212 'Ancestry' terms
European in 5,890 studies

Fluid ancestry categories, but also:
*__nationality/geography__ (Finnish, French, British)
*__geography__ (Mylapotamos, Sardinian)
*__political__ (Han Chinese –Taiwanese Han Chinese)

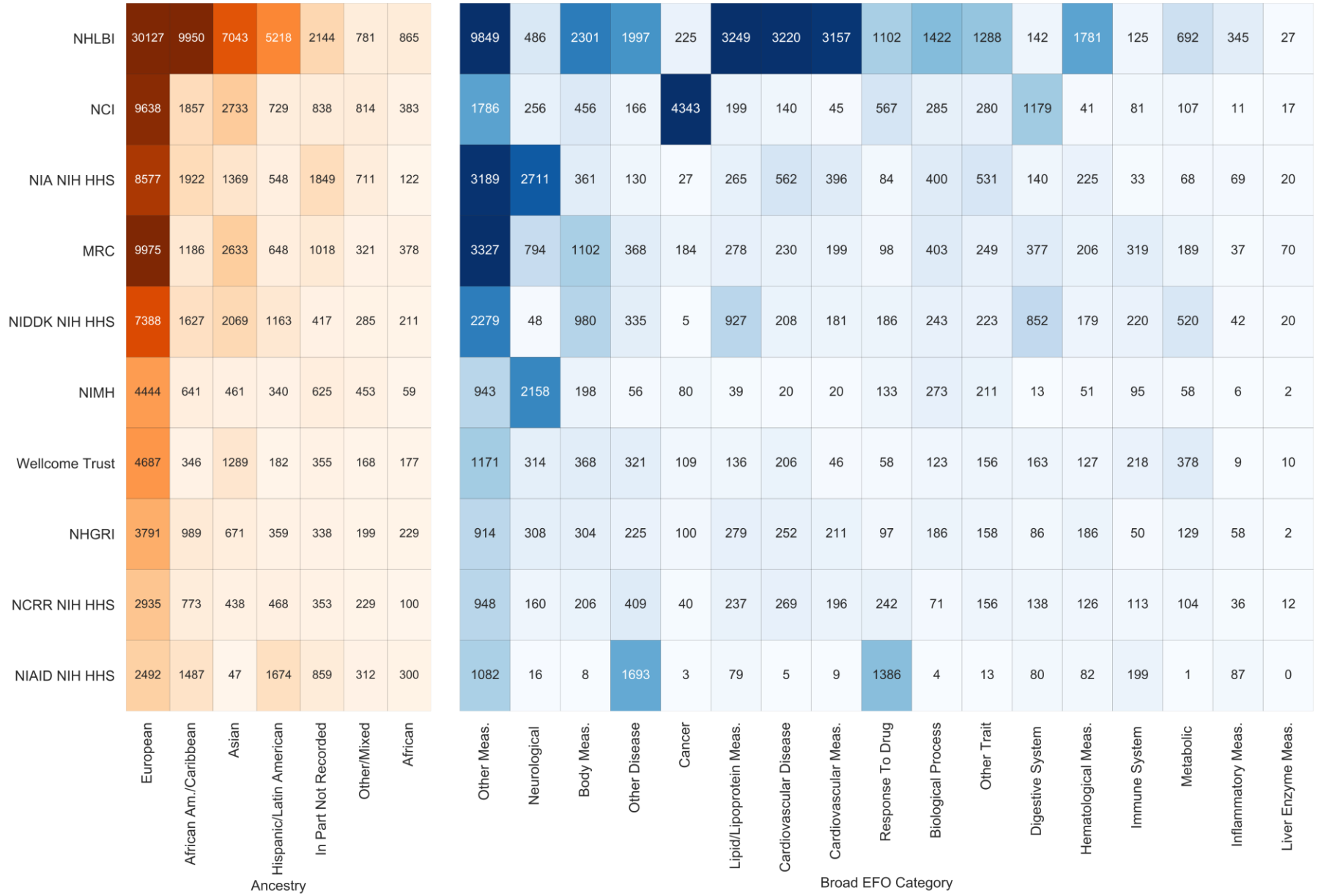
Prominent groupings:
*African American
*Hispanic/Latino/Latin American or just Latino
*African American/Afro Caribbean
*Indian Asian
*Northern European

Mills & Rahal (2019) https://www.nature.com/articles/s42003-018-0261-x

Country of Recruitment – 72% come from 3 countries (UK, US, Iceland)



Mills & Rahal (2019) https://www.nature.com/articles/s42003-018-0261-x



Mills & Rahal (2019) https://www.nature.com/articles/s42003-018-0261-x

UNIVERSITY OF OXFORD

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

gwasdiversitymonitor.com

correspondence

The GWAS Diversity Monitor tracks diversity by disease in real time

To the Editor — The Genome-wide association study (GWAS) is a primary tool for the discovery of associations between genetic variants and complex phenotypes, cataloged by the National Human Genome Research Institute–European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog, which currently contains information on more than 4,346 published studies across more than 4,933 diseases and traits. Although there has been a considerable

2019. Our cumulative estimate at the time of writing currently stands at 88.45%, even despite the recent launch of initiatives such as H3Africa, the African Genome Variation Project and GenomeAsia 100k. Geographic and demographic diversity is also limited, and other estimates suggest that 72% of participants are recruited from just three countries (the United States, the United Kingdom and Iceland)[1].

The transferability of GWAS results

age or sex, and socioeconomic status of individuals[6]. With the move toward the use of PRSs derived from GWAS for clinical applications[7], most PRSs derived from GWAS would exacerbate existing global health inequalities[8]. Substantially more genetic variation exists in non-European populations, and this variation can provide a rich resource for finding new genetic associations (Supplementary Note 1).

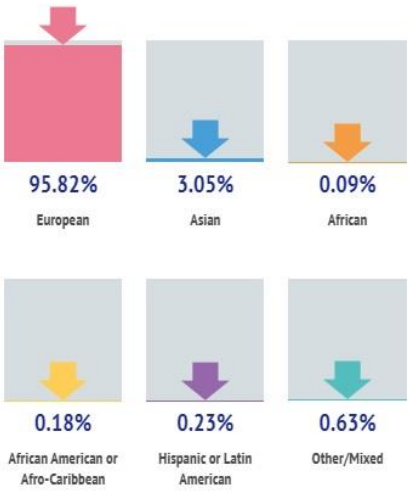
GWAS often fail to identify variants that

Mills, M.C. & C. Rahal (2020), *Nature Genetics, https://rdcu.be/b2BX4*

gwasdiversitymonitor.com

Total GWAS participants diversity

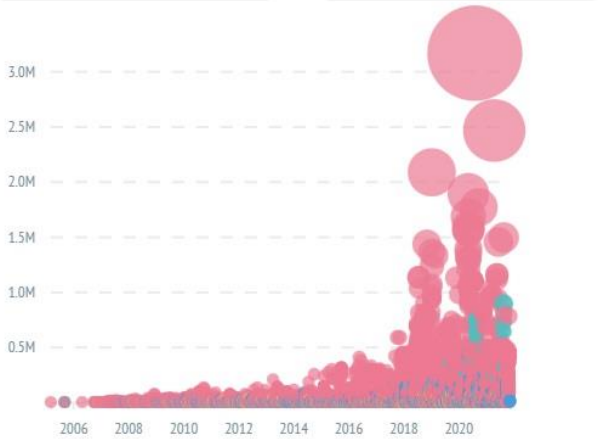
Version 1.0.0. Last check for data: 2022-01-18 09:34:08 .



Ancestry over time by parent term

Discovery Stage

All parent terms OR Search for one or more trai



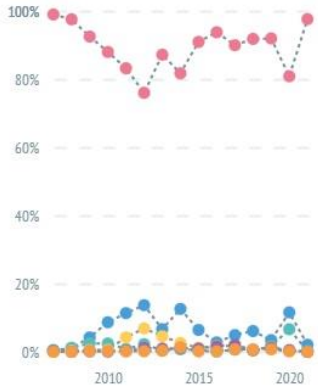
VIEW ALL

- European
- Asian
- African
- African American or Afro-Caribbean
- Hispanic or Latin American
- Other/Mixed

Participants across all parent terms

Discovery Stage

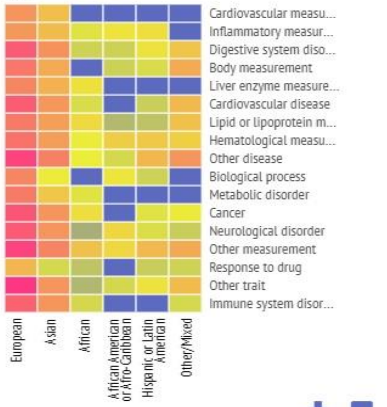
All ancestries Include not recorded



Parent term by 'broader' ancestry

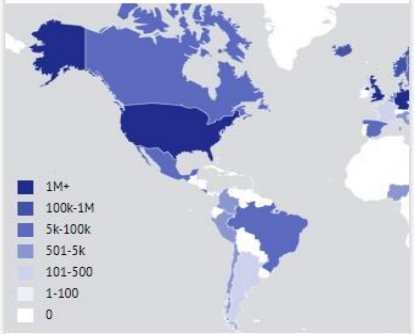
By Participants - Discovery Stage

2021



Participants by country (all parent terms)

Both Stages



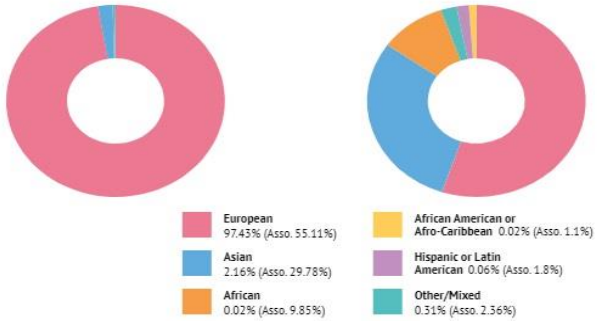
2021

Participants by ancestry

Discovery Stage

Click to hide associations discovered

All parent terms



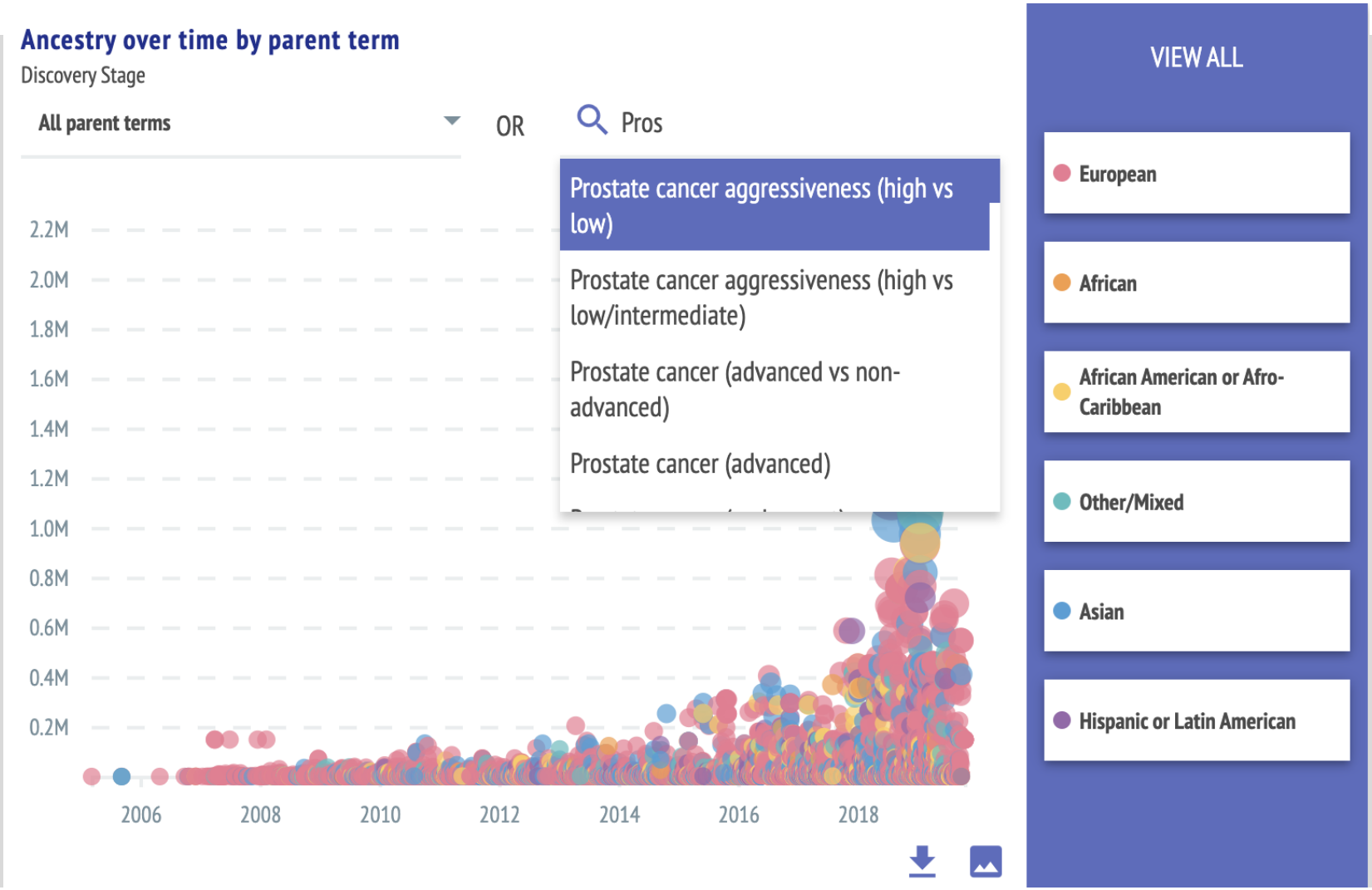
2021

Count of all associations discovered

Discovery Stage

Hide

Real-time search 5500+ phenotypes by ancestry



Mills, M.C. & C. Rahal (2020) The GWAS Diversity Monitor tracks diversity by disease in real time, *Nature Genetics*

Top 10 datasets

1. European ancestry
2. Mostly industrialized countries (NL, US, UK,Germany) – share similar disease prevalence and population profiles
3. Older populations
4. Sex ratio imbalance, more women
5. Non-representative samples (UKBB)

- genetic associations are modifiable,
- bias GWAS estimates towards over-represented group; association observed in one study dependent on exposure-outcome relationship in discovery & target population, Keyes & Westreich 2019)

Coded largest 1,250 largest GWAS as of August 2018 to generate list most used datasets
Full list of 2,000+ data used:
https://github.com/crahal/GWASReview/blob/master/tables/Manually_Curated_Cohorts.csv

Mills & Rahal (2019) https://www.nature.com/articles/s42003-018-0261-x

nature human behaviour

ARTICLES

DOI: 10.1038/s41562-017-0195-1

Hidden heritability due to heterogeneity across seven populations

Felix C. Tropf[1*], S. Hong Lee[2], Renske M. Verweij[3], Gert Stulp[3], Peter J. van der Most[4], Ronald de Vlaming[5,6], Andrew Bakshi[7], Daniel A. Briley[8], Charles Rahal[1], Robert Hellpap[1], Anastasia N. Iliadou[9], Tõnu Esko[10], Andres Metspalu[10], Sarah E. Medland[11], Nicholas G. Martin[11], Nicola Barban[1], Harold Snieder[4], Matthew R. Robinson[7,12] and Melinda C. Mills[1]

Meta-analyses of genome-wid torical time periods and popul the heritability estimates obta zle. Using seven sampling popu sampling populations and tim populations. We show that the for education, 40% for age at more likely to reflect heteroger These findings have substantia behavioural phenotypes and th

eLIFE elifesciences.org

RESEARCH ARTICLE

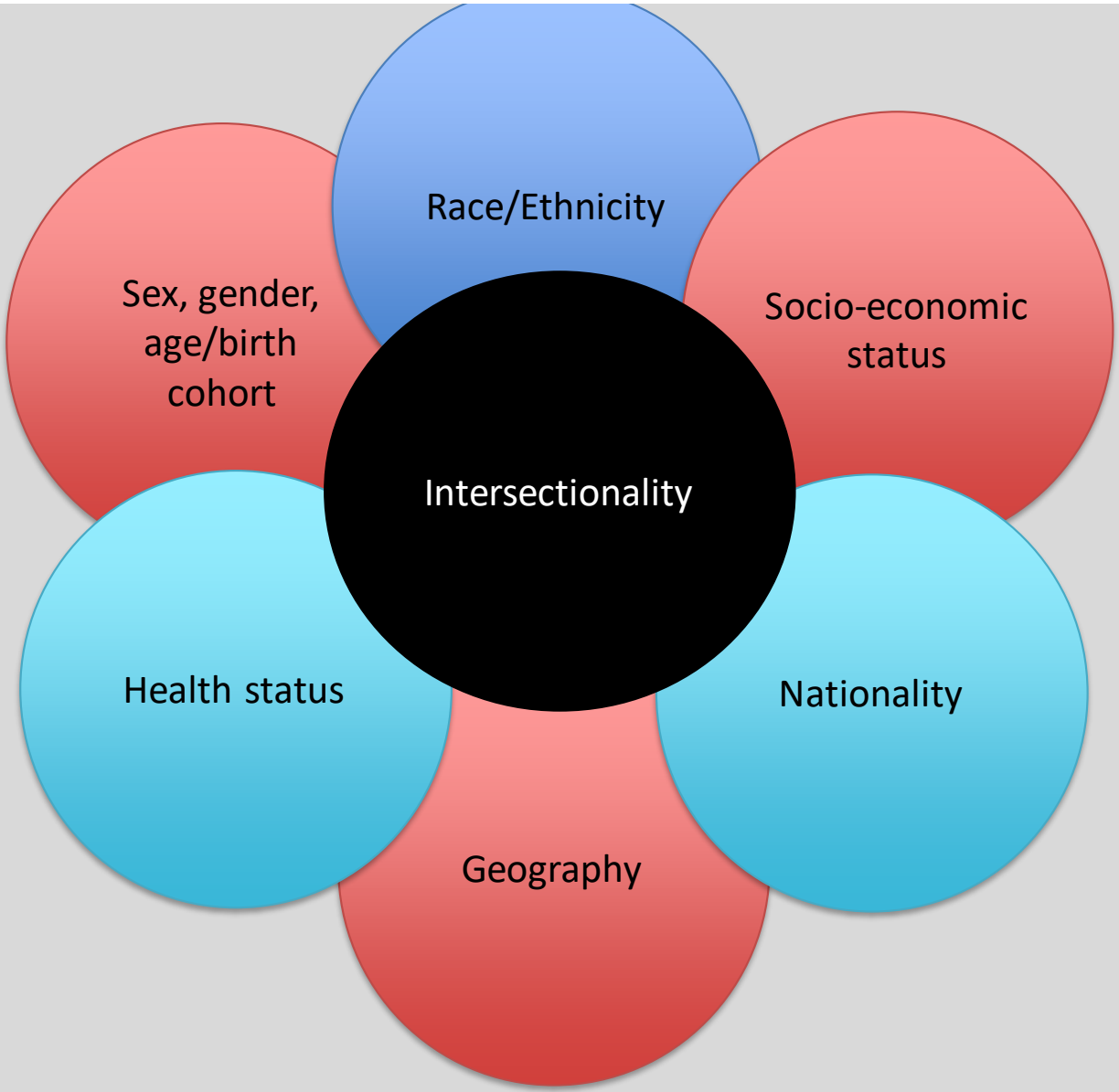
Variable prediction accuracy of polygenic scores within an ancestry group

Hakhamanesh Mostafavi[1†‡*], Arbel Harpak[1†*], Ipsita Agarwal[1], Dalton Conley[2,3], Jonathan K Pritchard[4,5,6], Molly Przeworski[1,7*]

[1]Department of Biological Sciences, Columbia University, New York, United States; [2]Department of Sociology, Princeton University, Princeton, United States; [3]Office of Population Research, Princeton University, Princeton, United States; [4]Department of Genetics, Stanford University, Stanford, United States; [5]Department of Biology, Stanford University, Stanford, United States; [6]Howard Hughes Medical Institute, Stanford University, Stanford, United States; [7]Department of Systems Biology, Columbia University, New York, United States

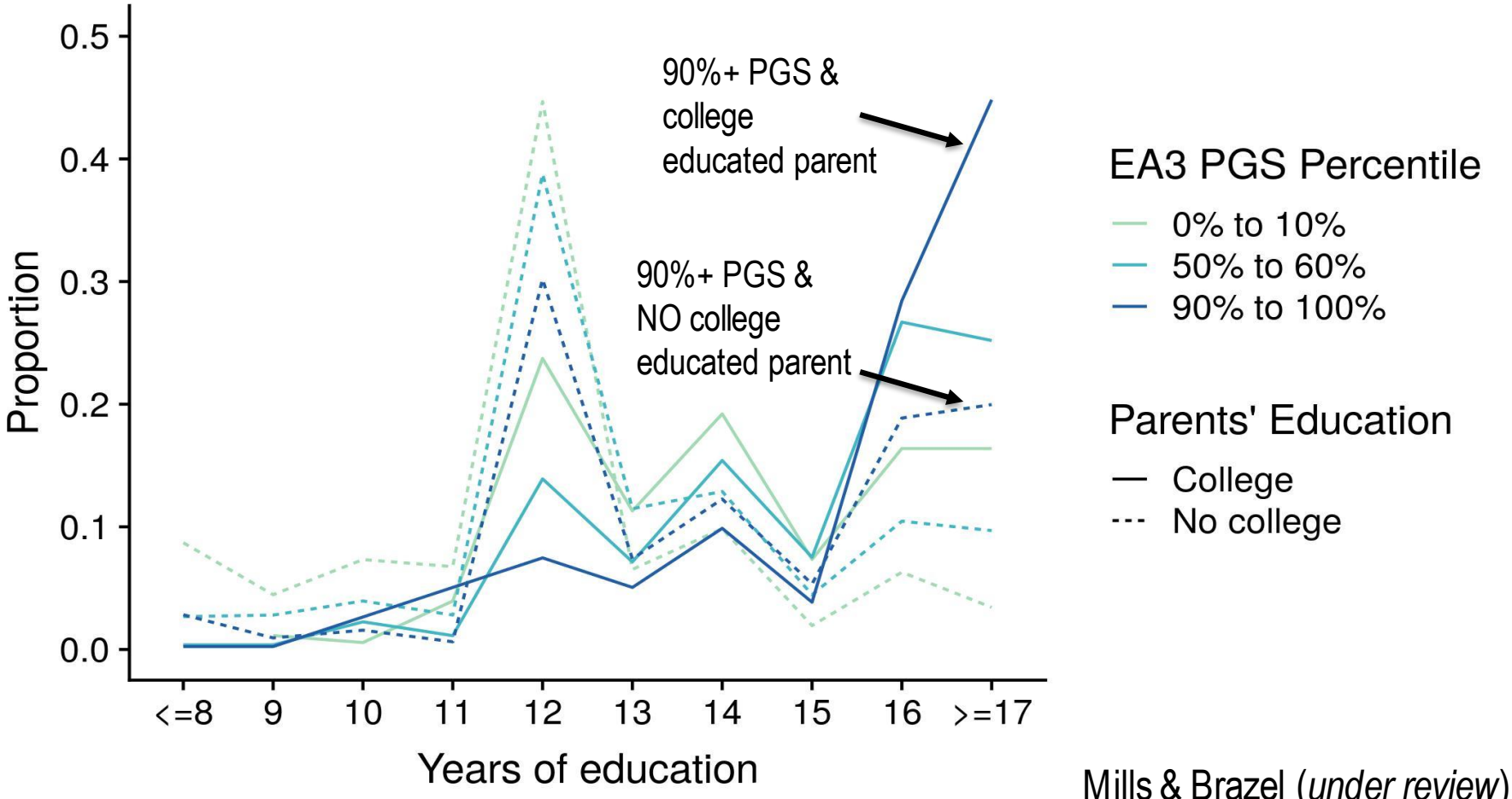
Abstract Fields as diverse as human genetics and sociology are increasingly using polygenic scores based on genome-wide association studies (GWAS) for phenotypic prediction. However,

Cumulative disadvantage & intersectionality



Highest PGS group (90-100%) highest education levels, but……

GxE: children with at least **one college-educated parent have substantially higher educational attainment** in same PGS range than those without a college-educated parent



Mills & Brazel (*under review*)

Future recommendations

Mills & Rahal (2019)

1. prioritize the inclusion of **multiple types of diversity** (socioeconomic status, sex)
2. **careful interpretation** of genetic differences
3. **participant and researcher** involvement
4. reduce **inequalities in authorship** and investigators
5. **reform incentive structures** role of authorship, data ownership, and dating sharing
6. **coordinated governance, guidance** from multiple stakeholders
7. **monitoring** with funding consequences
8. **utilize influence** for the good of more people