

Biased European Database Participation

Why is this a HUGE problem for
precision medicine?

If you choose two people at random, who would have the most genomic diversity between them?

1. One individual of European Ancestry, one individual of African Ancestry
 2. One individual of European Ancestry, one individual of European Ancestry
 3. One individual of African Ancestry, one individual of African Ancestry

How much genomic diversity would be between these individuals?

Recap of “Understanding Race” in Genomics

- At any given site in the human genome (Lewontin, 1972):
 - 85.4% of total variation in all humans occurs **within** a population
 - 6.3% of total variation in all humans occurs **between** historically defined ‘races’
 - 8.3% of total variation in all humans occurs **between populations within a ‘race’**
- Race is not a genetically (or biologically) based metric; it is a socially constructed one.
 - However, most health professionals think that ‘race’ can provide useful clinical information by proxy – although most of them agree that ancestral population is most relevant for diagnosis
- Ancestry is a broad term that assigns genomic variation at many sites to:
 - population genetic forces: migration (and isolation), mutation, recombination, natural selection

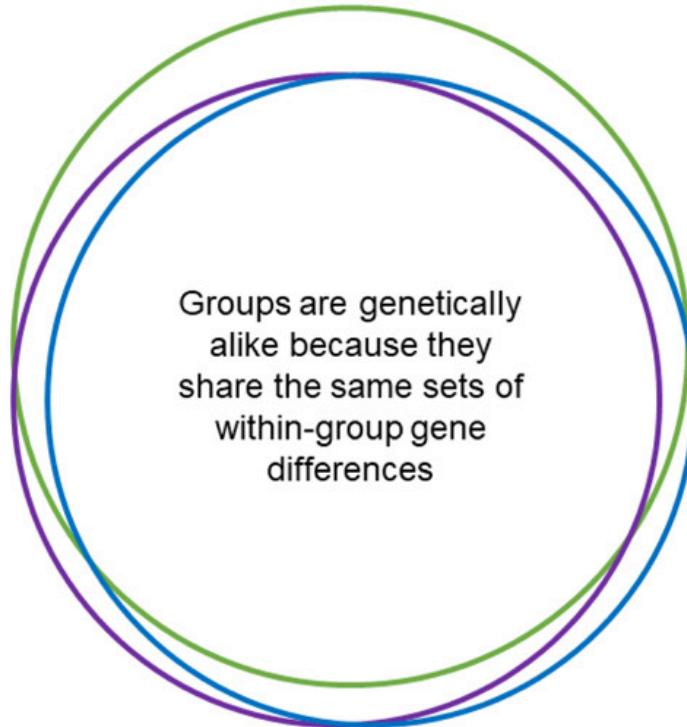
Toward a more humane genetics education: Learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations

█ African ancestry

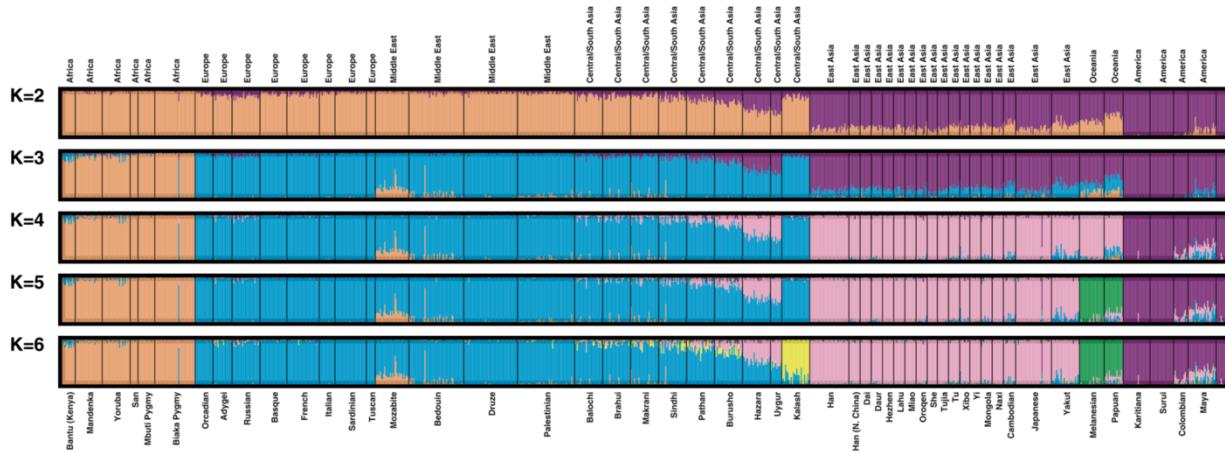
█ East-Asian ancestry

█ European ancestry

○ Each circle represents the total number of alleles (gene variants) in the human genome found within a particular group



Donavon et al (2019)

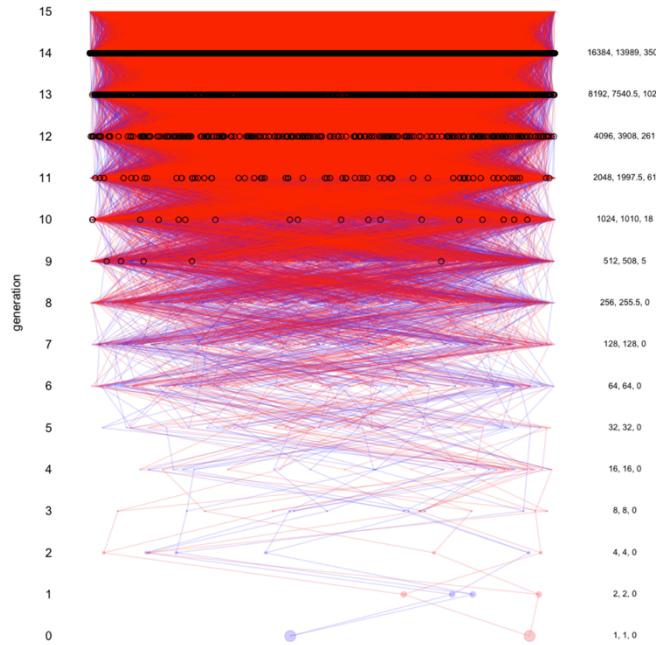


Rosenberg et al (2002)

Your pedigree is not always your ancestry: impact on GWAS

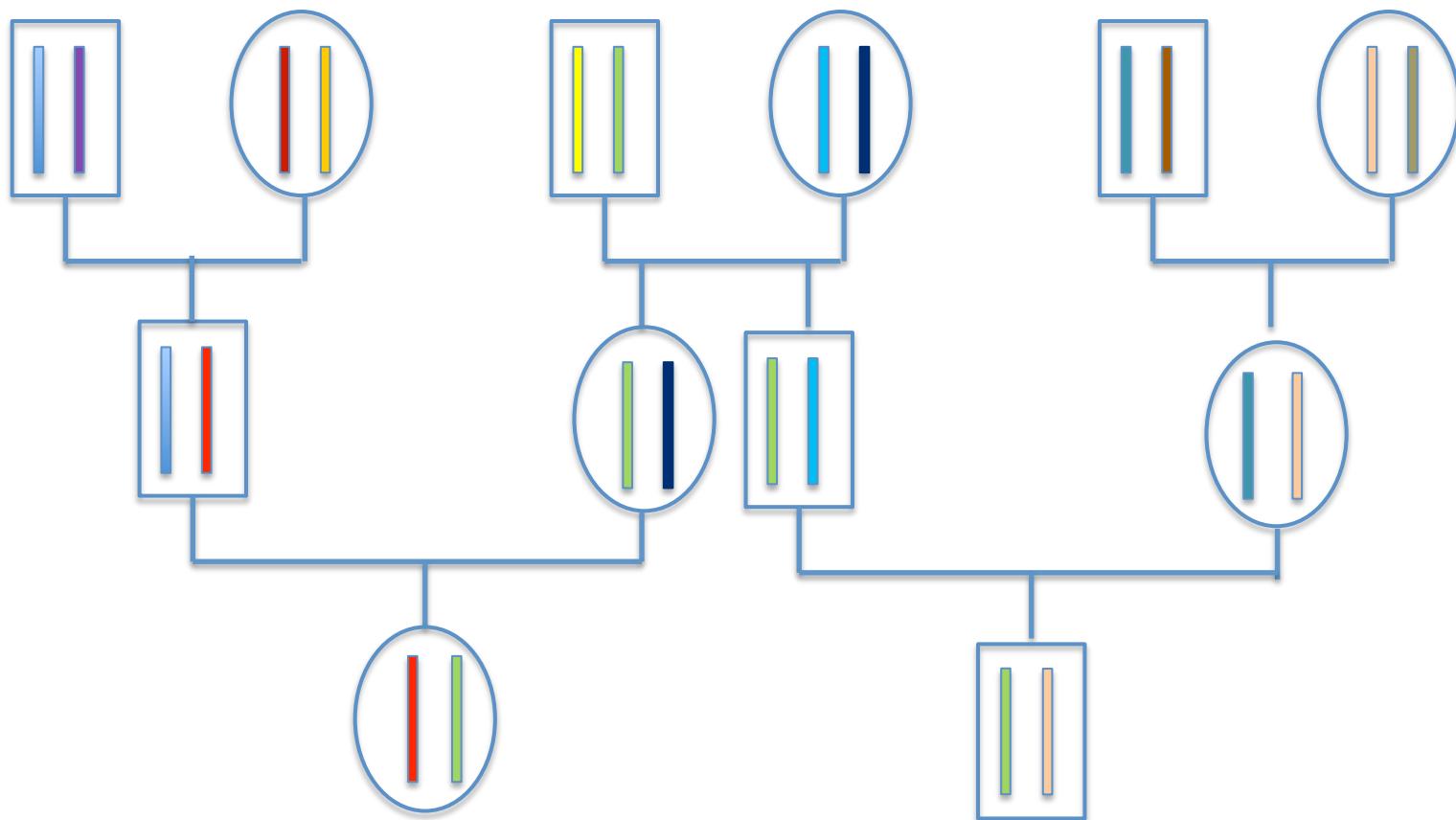
Number of ancestors grows so quickly that is almost unavoidable that we share an ancestor:

- 600 years ago ($2^{20} > 1,000,000$ ancestors)
- 1000 years ago ($2^{33} > 1,000,000,000$ ancestors way more Europeans than were alive 1000 years ago)

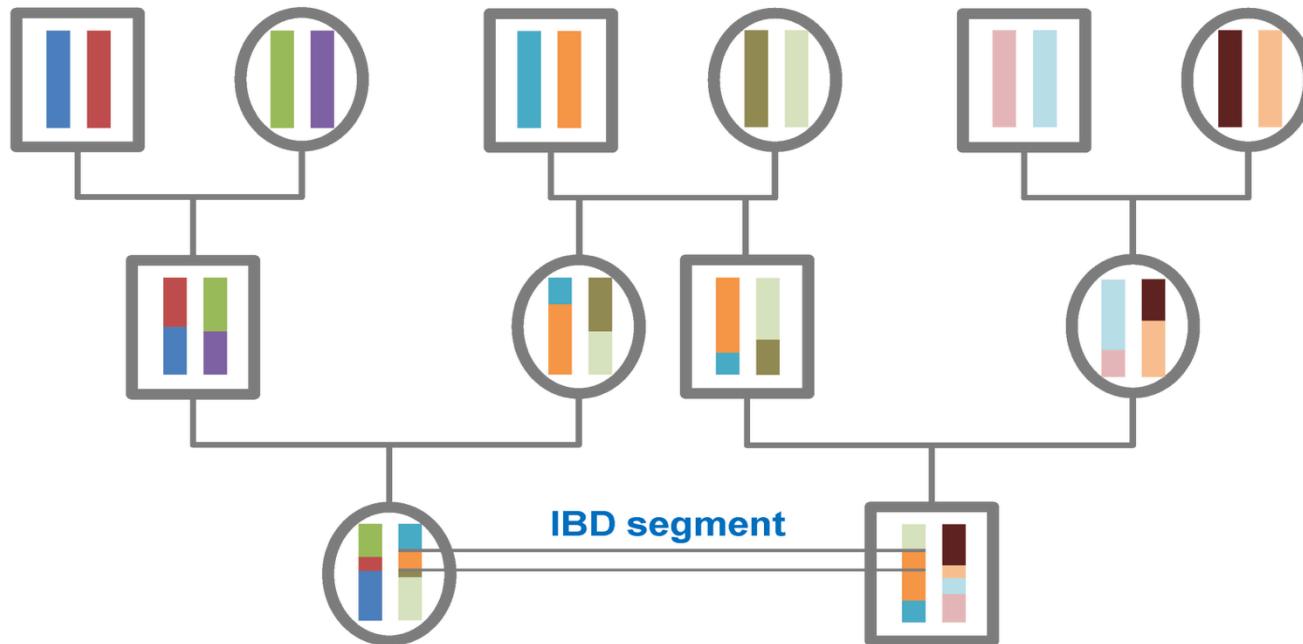


What is Identity-by-Descent (IBD)?

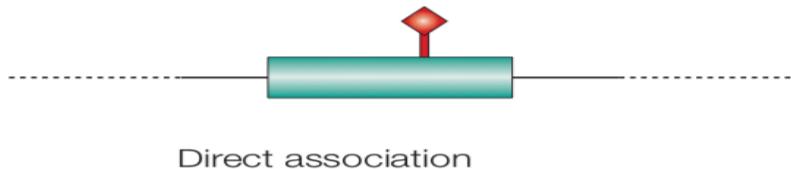
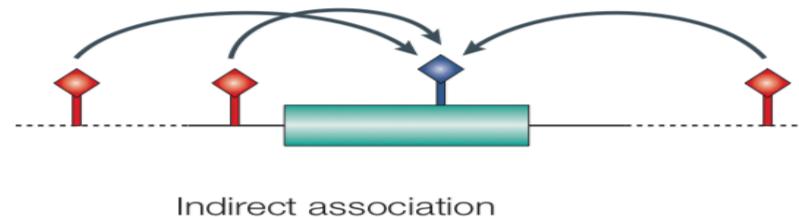
What we think it is when we begin to learn about genetics:



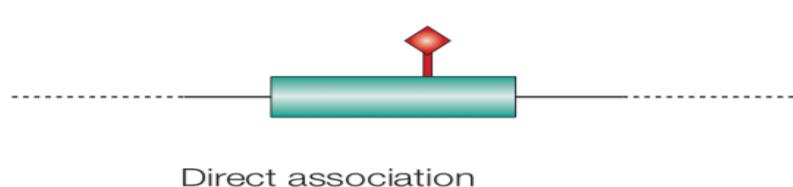
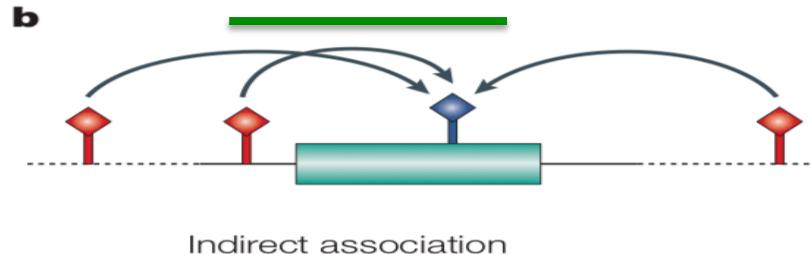
What I.B.D. actually is...



- Neighbouring “chunks” of each chromosome can have **DIFFERENT PEDIGREES/Ancestries**
 - Linkage Disequilibrium is when two adjacent sites have the same pedigree
- Differences in LD between ancestral populations has **PROFOUND** implications for GWAS

a**b**

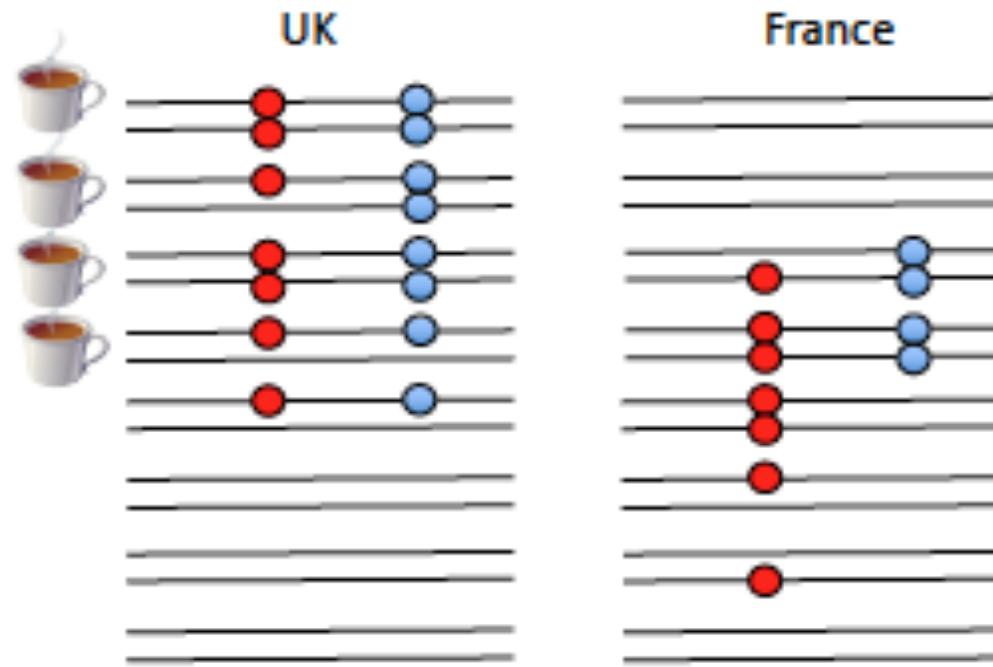
East-Asian Linkage Disequilibrium

a**b**

African Linkage Disequilibrium

Polygenic risk score

- Coop, Graham. (2019). Reading tea leaves? Polygenic scores and differences in traits among groups.



Polygenic risk scores are a summary statistic that sums all the variants impacting a disease as discovered via GWAS.

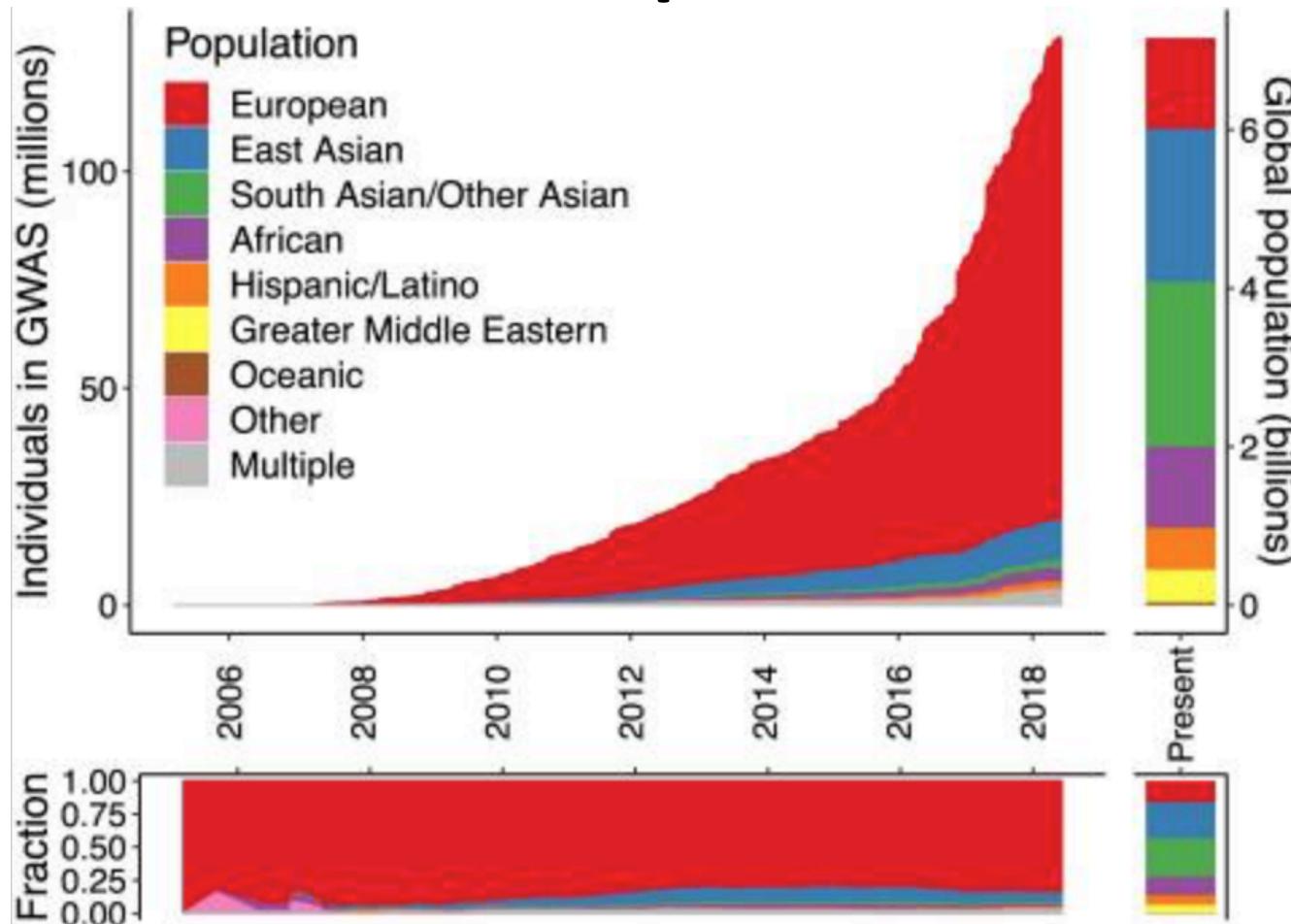
So.... Knowing that neighbouring sites in your genome can have different ancestries that are not the same as your pedigree....

How would you recommend Kitt and her mom, Eartha, self-identify for healthcare providers?



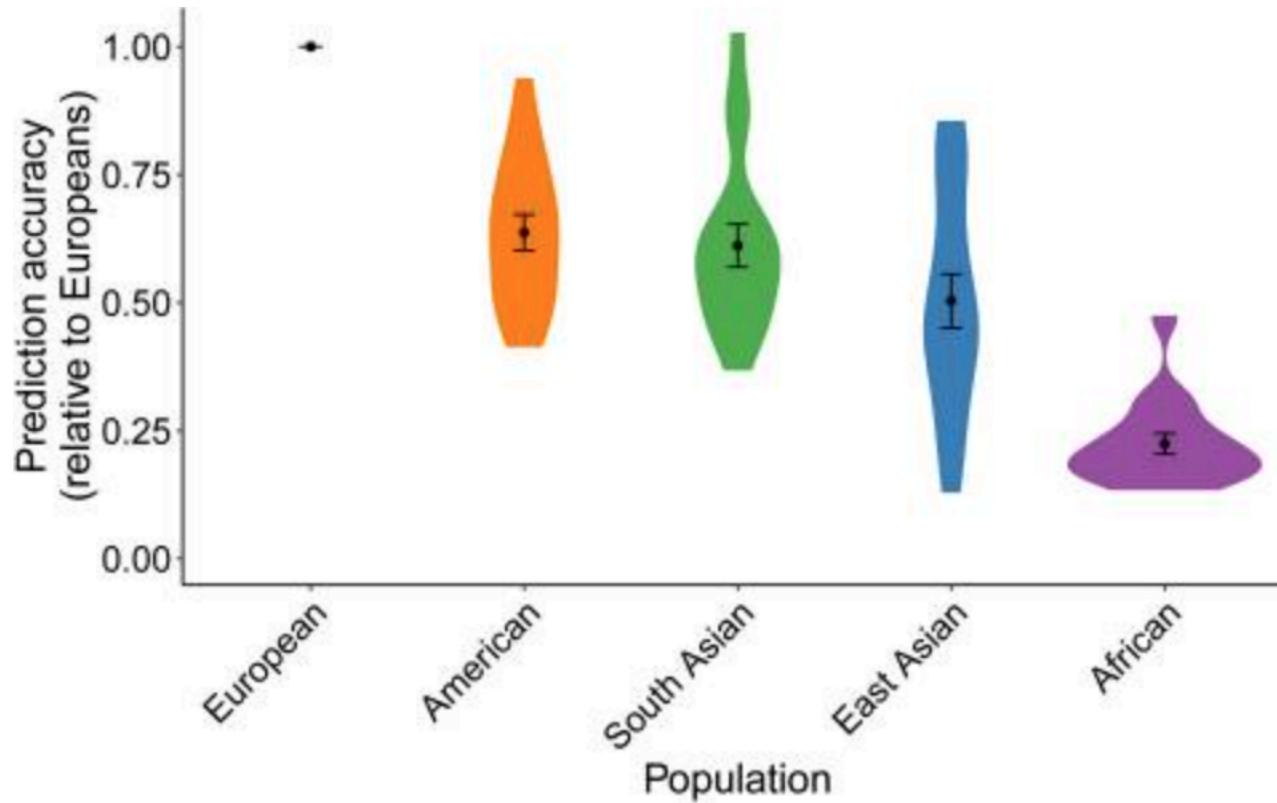
Equity in Genomics?

>80% of participants in genetic databases are European

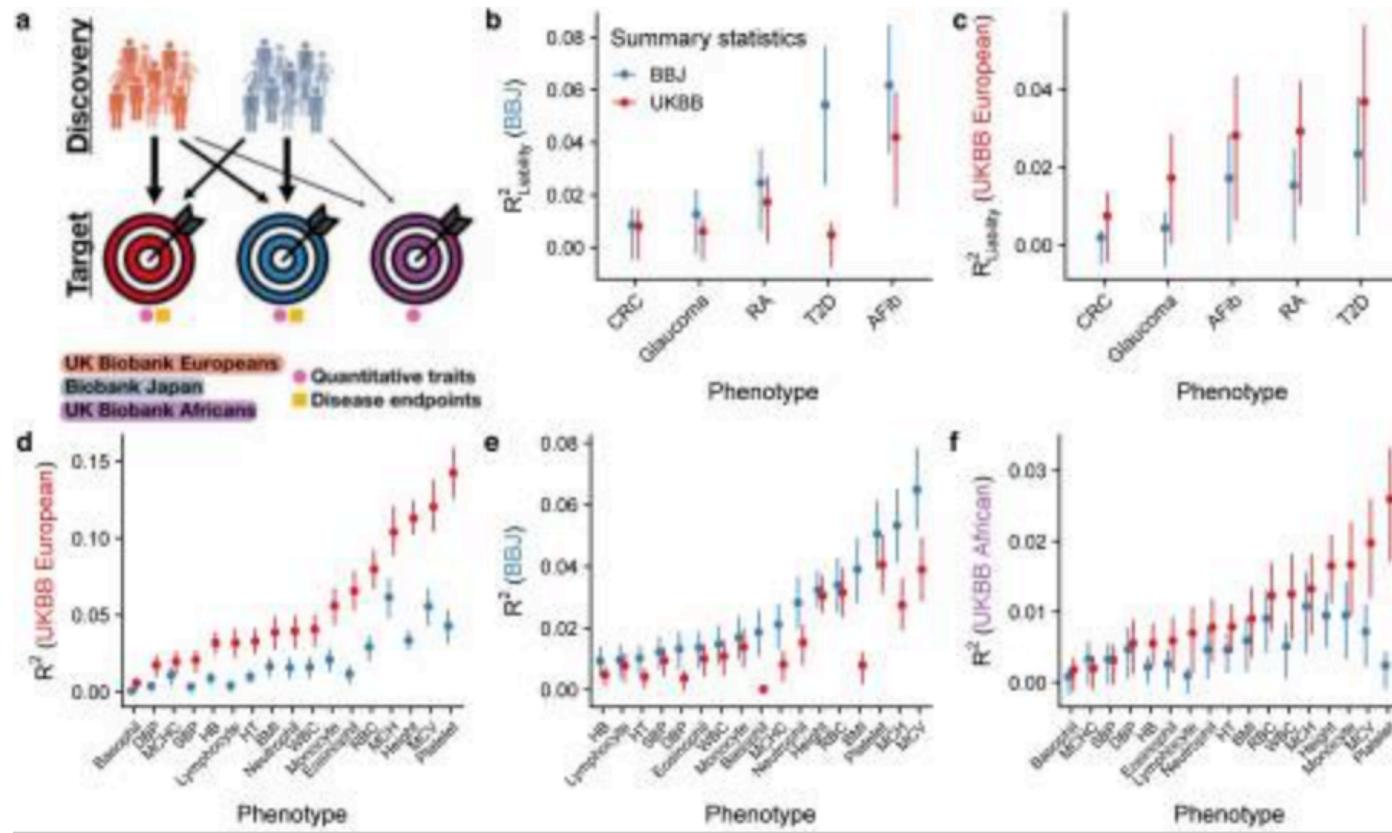


Martin et al (2019). Clinical use of current polygenic risk scores may exacerbate health disparities.
Nature Genetics, 51, 584-591 (2019)

Polygenic Risk Scores aren't portable between different ancestries



Polygenic Risk Scores aren't portable between different ancestries



PRS overview challenges

- AJHG paper by John Novembre, Catherine Stein,Sarah Tishkoff
 - Paper is in the kaplan folder (where this presentation is located)
- “Addressing the challenges of polygenic scores in human genetic research” – outlines the basic issues (and explains fine mapping as determining causative variants)

Using cross ancestry (Japanese, Finnish, European) to create PRS!

Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction

Abstract

Atrial fibrillation (AF) is a common cardiac arrhythmia resulting in increased risk of stroke. Despite highly heritable etiology, our understanding of the genetic architecture of AF remains incomplete. Here we performed a genome-wide association study in the Japanese population comprising 9,826 cases among 150,272 individuals and identified East Asian-specific rare variants associated with AF. A cross-ancestry meta-analysis of >1 million individuals, including 77,690 cases, identified 35 new susceptibility loci. Transcriptome-wide association analysis identified *IL6R* as a putative causal gene, suggesting the involvement of immune responses. Integrative analysis with ChIP-seq data and functional assessment using human induced pluripotent stem cell-derived cardiomyocytes demonstrated ERRg as having a key role in the transcriptional regulation of AF-associated genes. A polygenic risk score derived from the cross-ancestry meta-analysis predicted increased risks of cardiovascular and stroke mortalities and segregated individuals with cardioembolic stroke in undiagnosed AF patients. Our results provide new biological and clinical insights into AF genetics and suggest their potential for clinical applications.

Sections

[Abstract](#)

[Main](#)

[Results](#)

[Discussion](#)

[Methods](#)

[Data availability](#)

[References](#)

Figure



How do we address this serious deficiency?

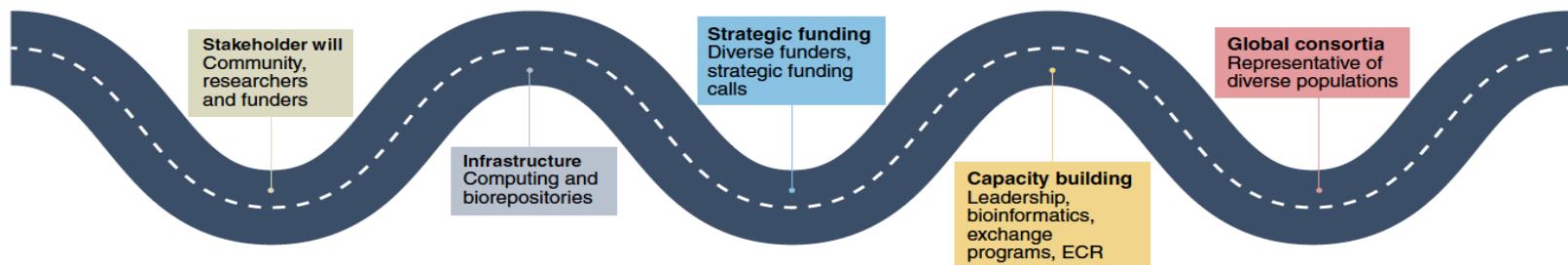


Fig. 3 | Roadmap showing the key pillars for setting up and sustaining diverse global genomic studies. ECR, early career researcher.

Fatumo et al (2022). A roadmap to increase diversity in genomic studies. Nature Medicine

Genetics is not enough to address health disparities

1. Barriers to participation:

- <https://www.nytimes.com/2022/06/08/magazine/eugenics-movement-america.html>
- Medical Apartheid: <https://www.penguinrandomhouse.com/books/185986/medical-apartheid-by-harriet-a-washington/>

2. Other upcoming issues of race and ancestry in precision medicine:

- AI and algorithms in healthcare (and others) have racial bias
- Gut Biome cultural factors contributions to overall health that we don't fully understand
- Environment – pollution, community: education levels, access to medical treatments. So called “ZNA” – zipcode