

Loss of prediction accuracy across genetic ancestries

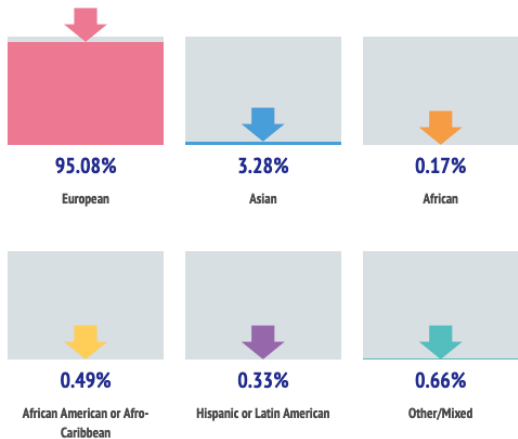
Valentin Hivert

**Postdoctoral Researcher – Visscher Group
Institute for Molecular Bioscience
The University of Queensland**

Credit: Slides from Dr Loic Yengo

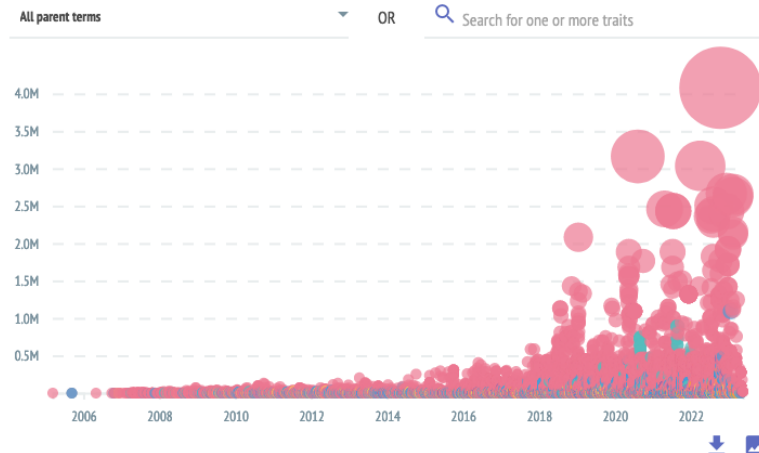
Total GWAS participants diversity

Version 1.0.0. Last check for data: 2023-06-12 00:17:38.



Ancestry over time by parent term

Discovery Stage

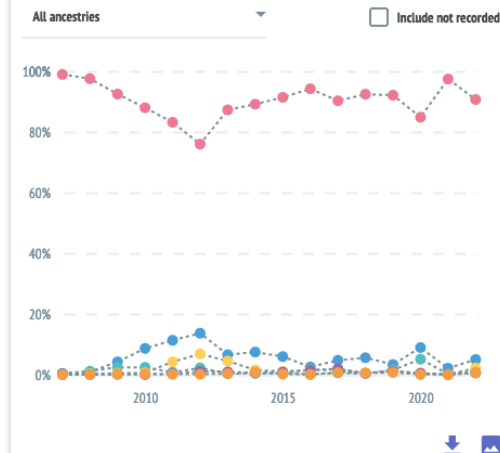


VIEW ALL

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

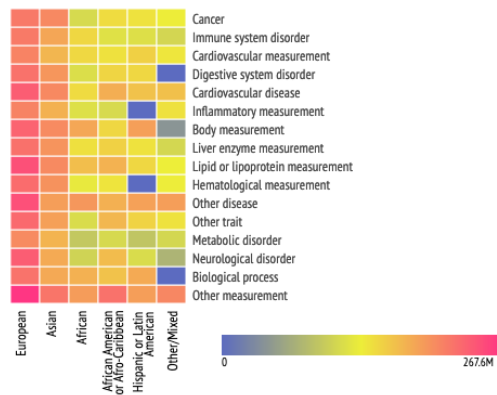
Participants across all parent terms

Discovery Stage



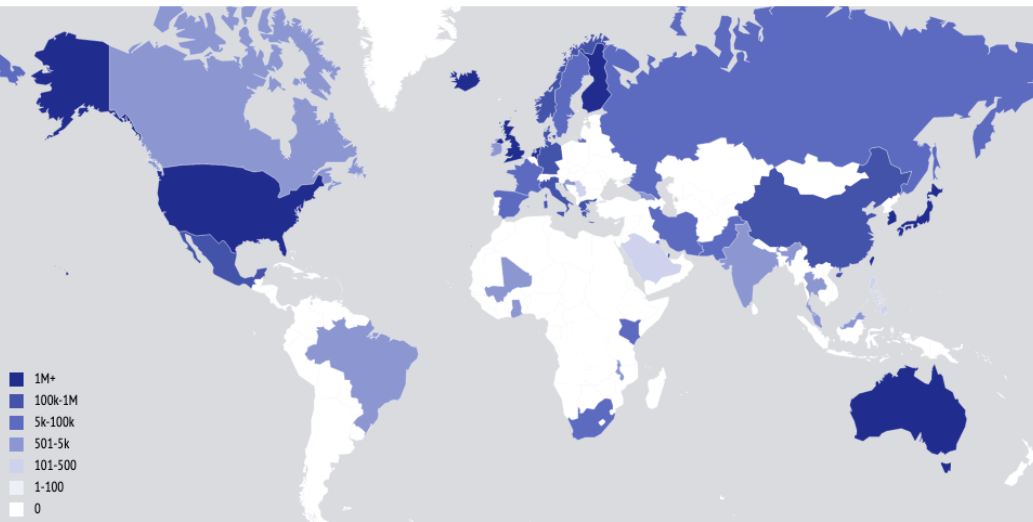
Parent term by 'broader' ancestry

By Participants - Discovery Stage



Participants by country (all parent terms)

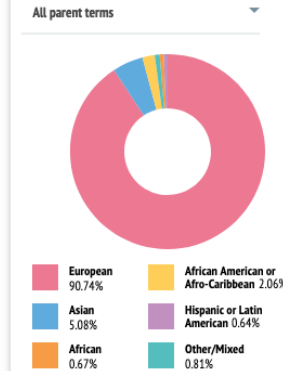
Both Stages



Participants by ancestry

Discovery Stage

Click to show associations discovered



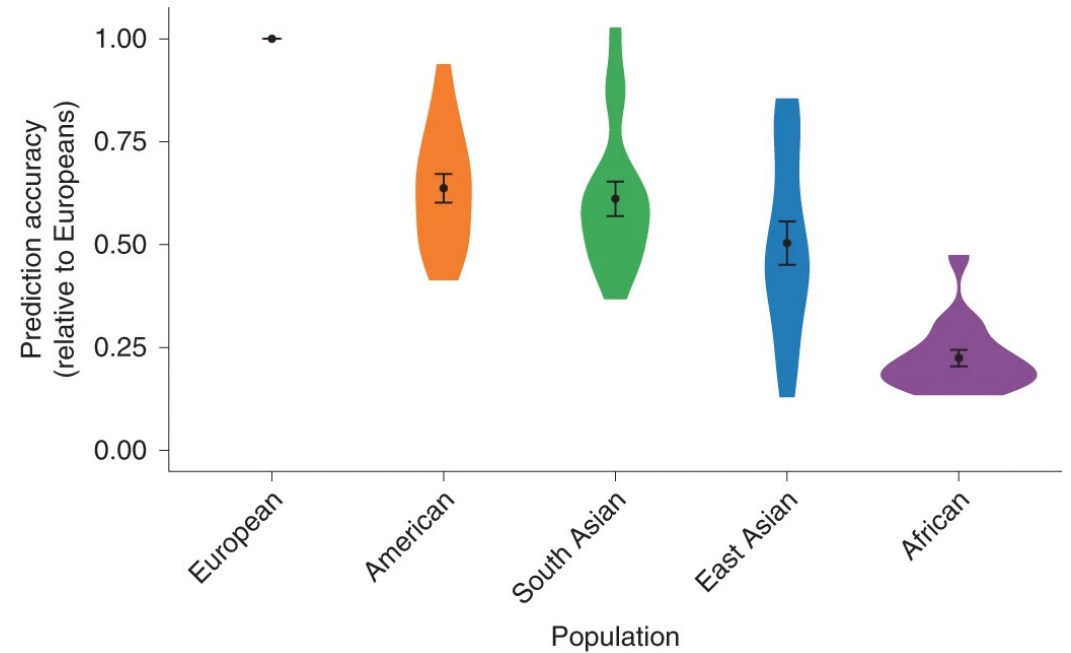
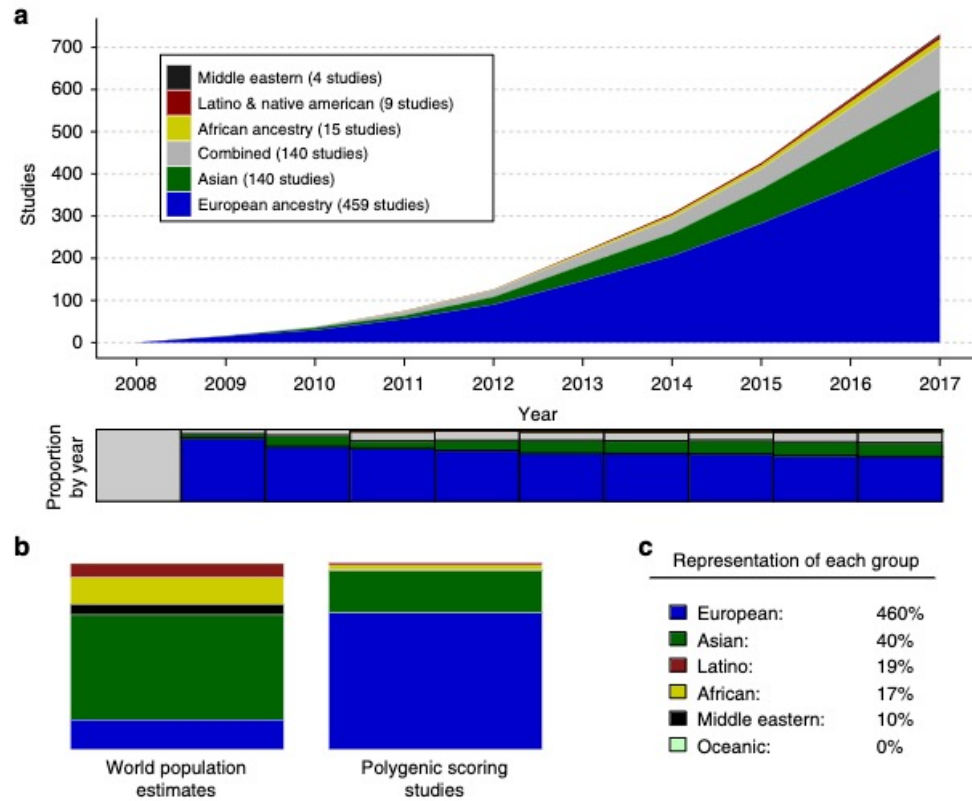
In 2023...

>95% European

<4% Asian

<0.5% African

Euro-centric GWASs bias Loss of prediction accuracy



Martin et. al., *Nature Genetics*, 2019

Duncan et. al., *Nature Communication*, 2019

Why?

- Ancestry-specific causal variants (worse case scenario)
- Same causal variants but different effect sizes
- Same causal variants, same effect sizes, same heritability, but different haplotype frequency (demographic and evolutionary history of populations)
- ...

Why?

- ~~Ancestry-specific causal variants (worse case scenario)~~
- ~~Same causal variants but different effect sizes~~

ARTICLE

Multi-ancestry fine-mapping improves precision to identify causal genes in transcriptome-wide association studies

Zeyun Lu,^{1,12,*} Shyamalika Gopalan,^{2,3,12} Dong Yuan,¹ David V. Conti,^{1,2} Bogdan Pasaniuc,^{4,5,6,7} Alexander Gusev,^{8,9,10} and Nicholas Mancuso^{1,2,11,*}

nature genetics

Article

<https://doi.org/10.1038/s41588-023-01338-6>

Causal effects on complex traits are similar for common variants across segments of different continental ancestries within admixed individuals

Received: 10 August 2022

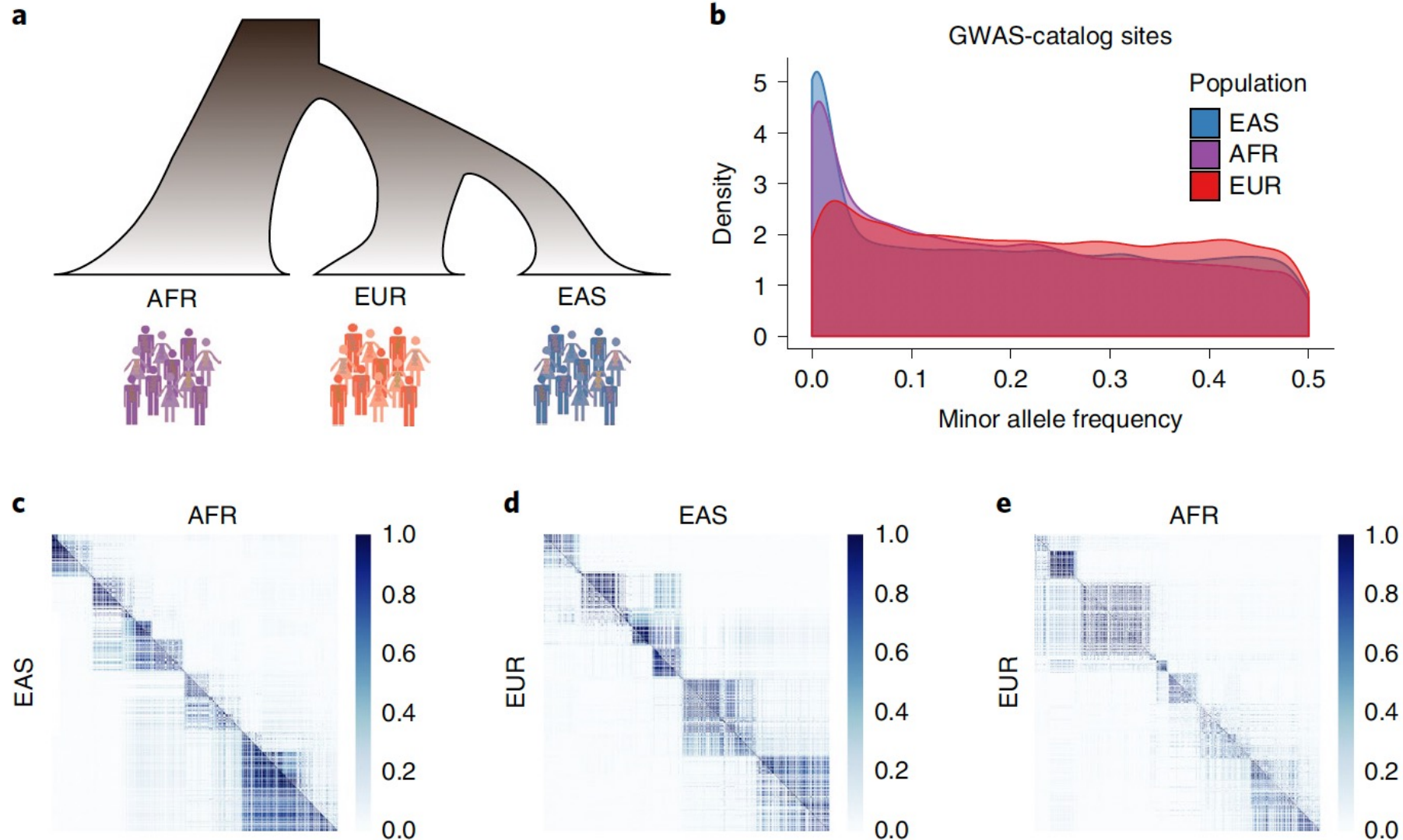
Accepted: 16 February 2023

Published online: 20 March 2023

 Check for updates

Kangcheng Hou¹✉, Yi Ding¹, Ziqi Xu², Yue Wu², Arjun Bhattacharya³, Rachel Mester⁴, Gillian M. Belbin^{5,6,24}, Steve Buyske^{7,24}, David V. Conti^{8,24}, Burcu F. Darst^{9,24}, Myriam Fornage^{10,24}, Chris Gignoux^{11,24}, Xiuqing Guo^{12,24}, Christopher Haiman^{8,24}, Eimear E. Kenny^{5,13,14,24}, Michelle Kim^{9,24}, Charles Kooperberg^{9,24}, Leslie Lange^{15,24}, Ani Manichaikul^{16,24}, Kari E. North^{7,17,24}, Ulrike Peters^{9,24}, Laura J. Rasmussen-Torvik^{18,24}, Stephen S. Rich^{16,24}, Jerome I. Rotter^{12,24}, Heather E. Wheeler^{19,20,24}, Genevieve L. Wojcik^{21,24}, Ying Zhou^{9,24}, Sriram Sankararaman^{1,2,22,23} & Bogdan Pasaniuc^{1,3,22,23}✉

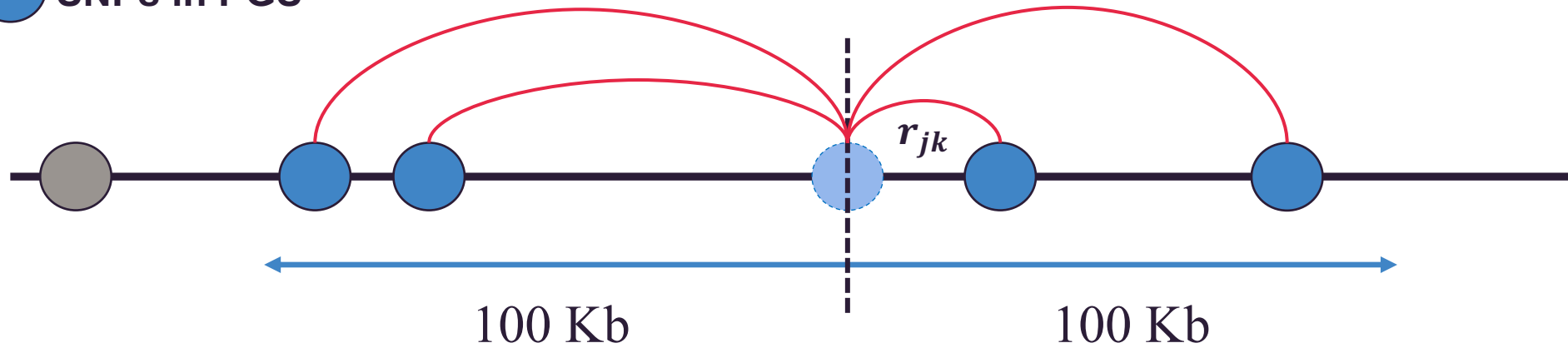
Factors affecting PGS accuracy disparity



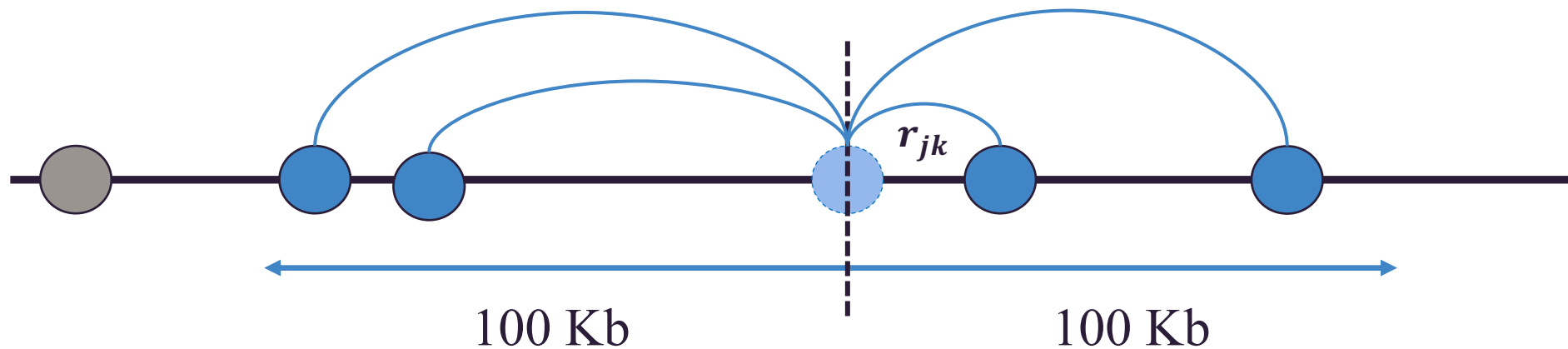
LD differences

True causal variant

SNPs in PGS



Population 1



Population 2

Factors affecting PGS accuracy disparity



How to **quantify** the **loss of accuracy** attributable to MAF and LD?



ARTICLE

<https://doi.org/10.1038/s41467-020-17719-y>

OPEN

Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations

Ying Wang ¹, Jing Guo¹, Guiyan Ni¹, Jian Yang ^{1,2}, Peter M. Visscher ¹ & Loic Yengo ^{1✉}

Theory

Relative Accuracy (RA)



$$\frac{R_2^2}{R_1^2} = \text{Function of } \frac{r_b^2 h_2^2}{h_1^2}, p_{k1}^{(t)}, p_{k2}^{(t)}, \frac{\text{var}(\hat{y}_1)}{\text{var}(\hat{y}_2)}, \text{ and } r_{jk,1}, r_{jk,2}$$

r_b^2 : Squared genetic correlation

h_l^2 : Trait heritability in Population l

p_{kl} : MAF of k-th tag SNP in Population l

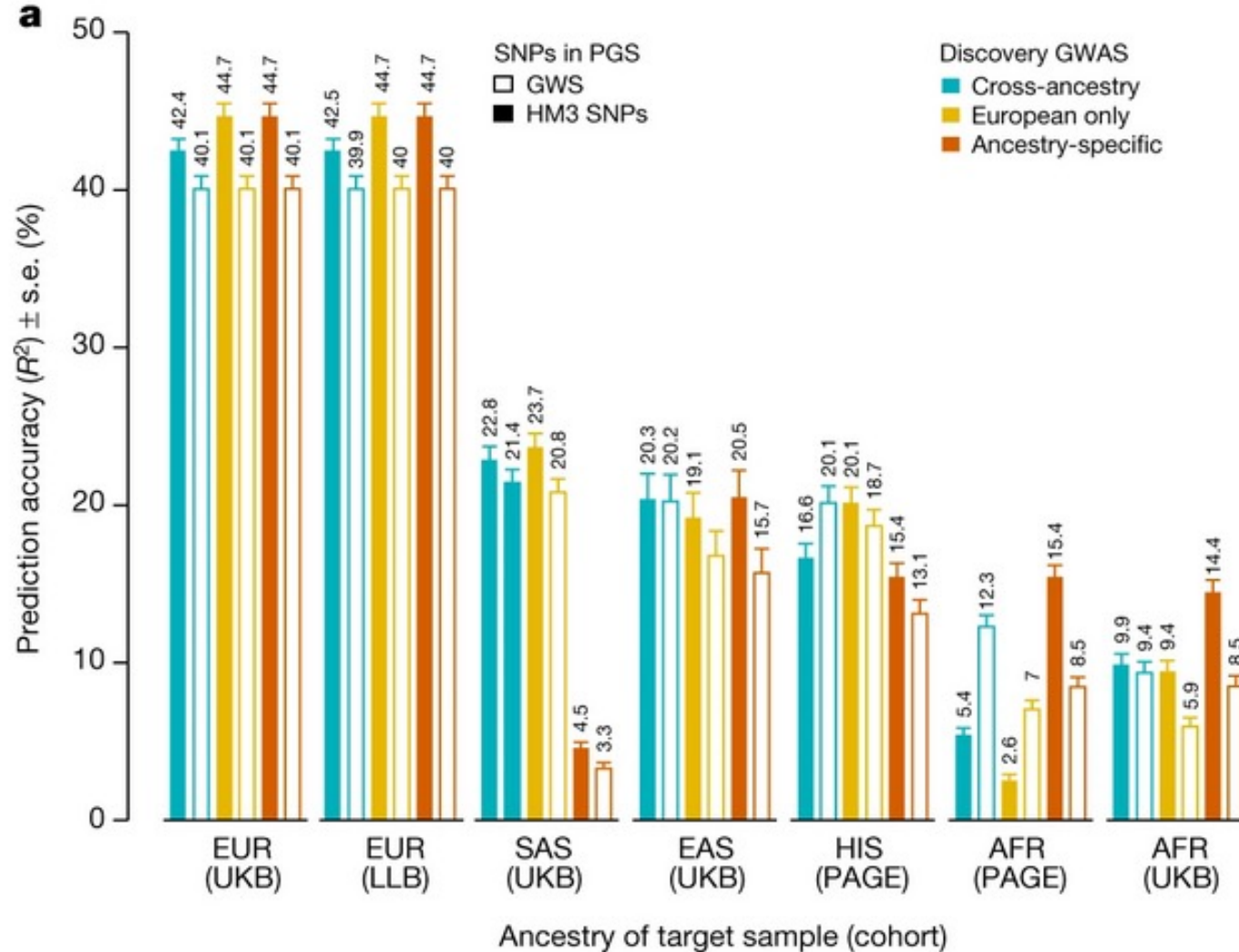
$r_{jk,l}$: LD correlation between j-th causal variant and k-th tag SNP in Population l

$\text{var}(\hat{y}_l)$: variance of PGS in Population l

Conclusions

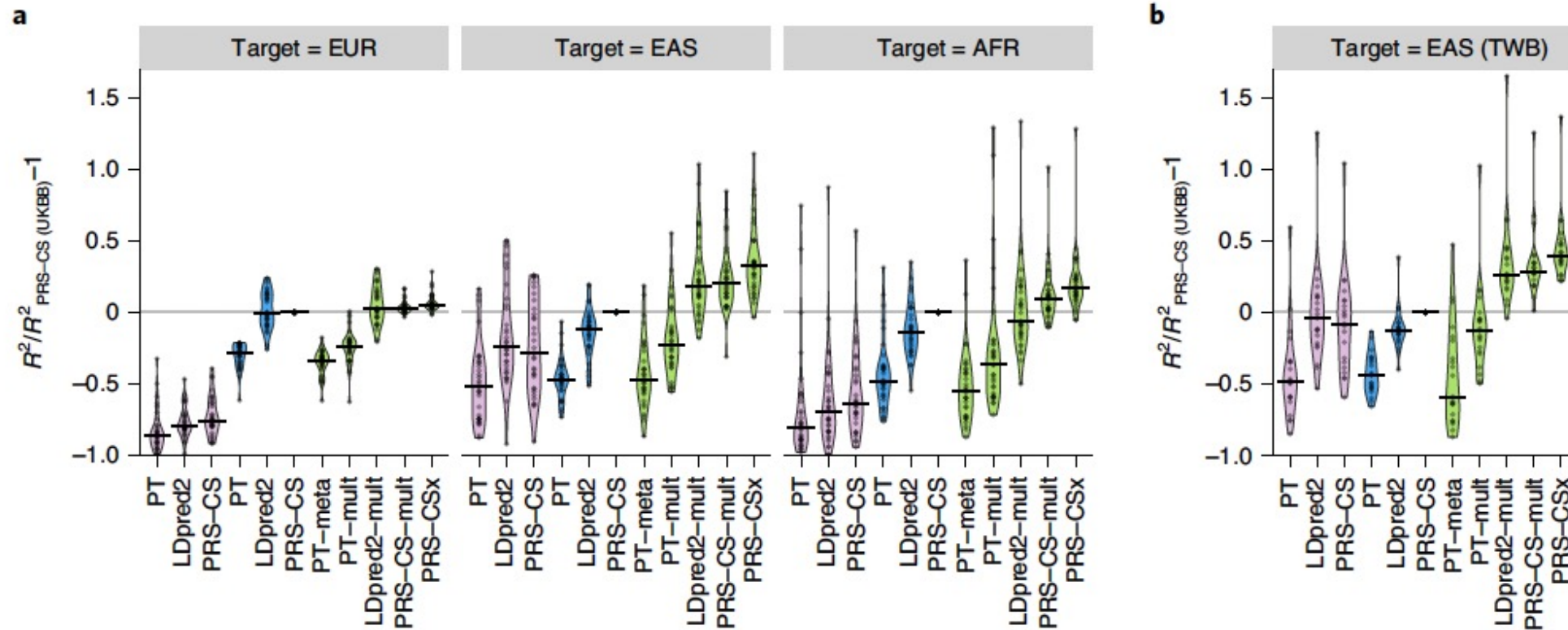
1. Empirical data suggest that causal variants (and effects) are largely shared across populations
2. Relative accuracy is lower than expected if assuming GWAS SNPs are causal variants (→ wrong conclusion)
3. LD and MAF differences between populations account for the majority of the loss of prediction accuracy

Implication: meta-analysis can improve accuracy



Yengo et al., *Nature*, 2022

Implication: methods accounting for MAF/LD differences can improve accuracy



Ruan et al., *Nature Genetics*, 2022

Summary

- Polygenic predictors lose a significant amount of their accuracy across populations of different genetic ancestries
- MAF and LD differences between populations can largely account for this phenomenon
- Evidence suggests that causal variants (and effects are largely shared across populations)
 1. Implication 1: meta-analysis can improve cross-ancestry prediction
 2. Implication 2: methods accounting for MAF and LD can improve further