Polygenic risk prediction in diverse populations and contexts: scientific considerations

Alicia Martin, PhD

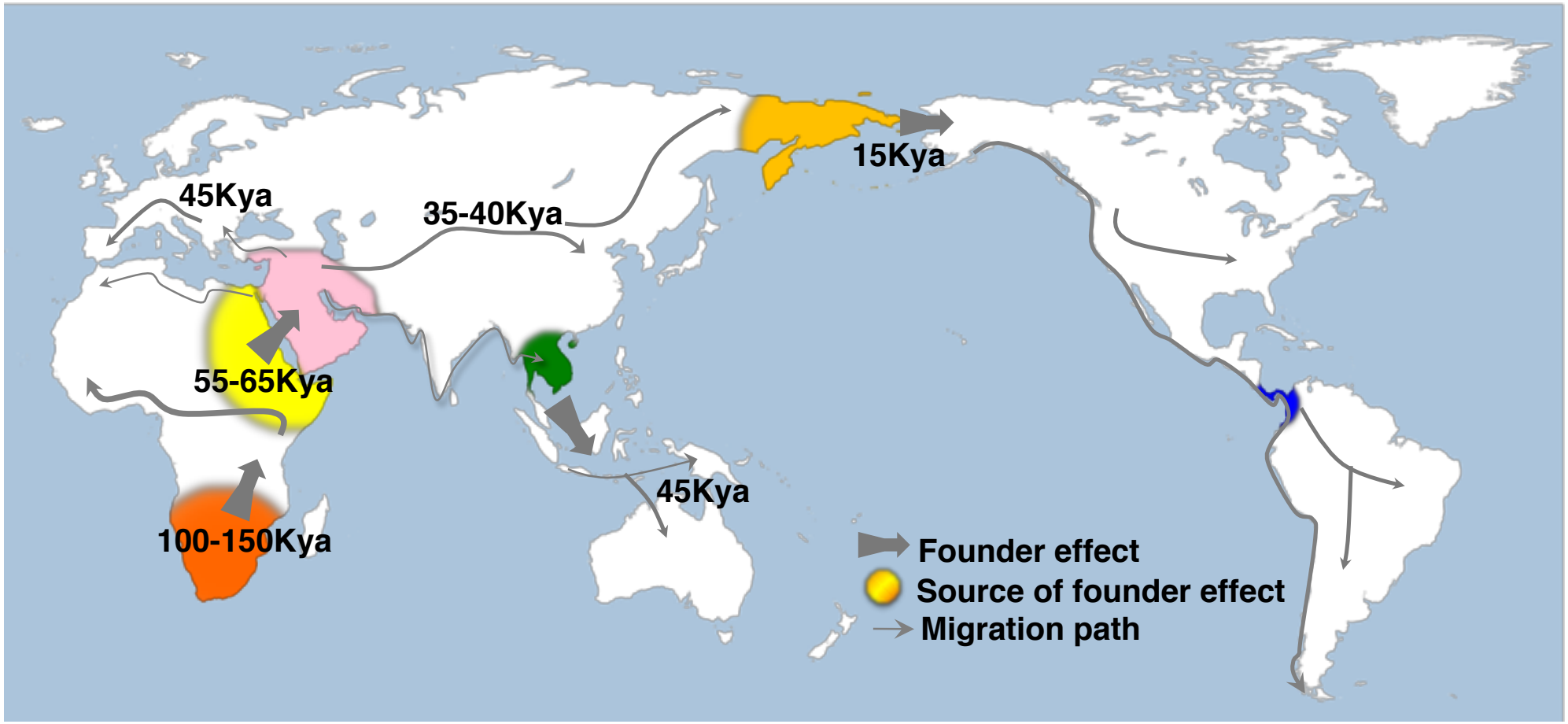
CERA ELSI-hub

January 12, 2024

armartin@broadinstitute.org

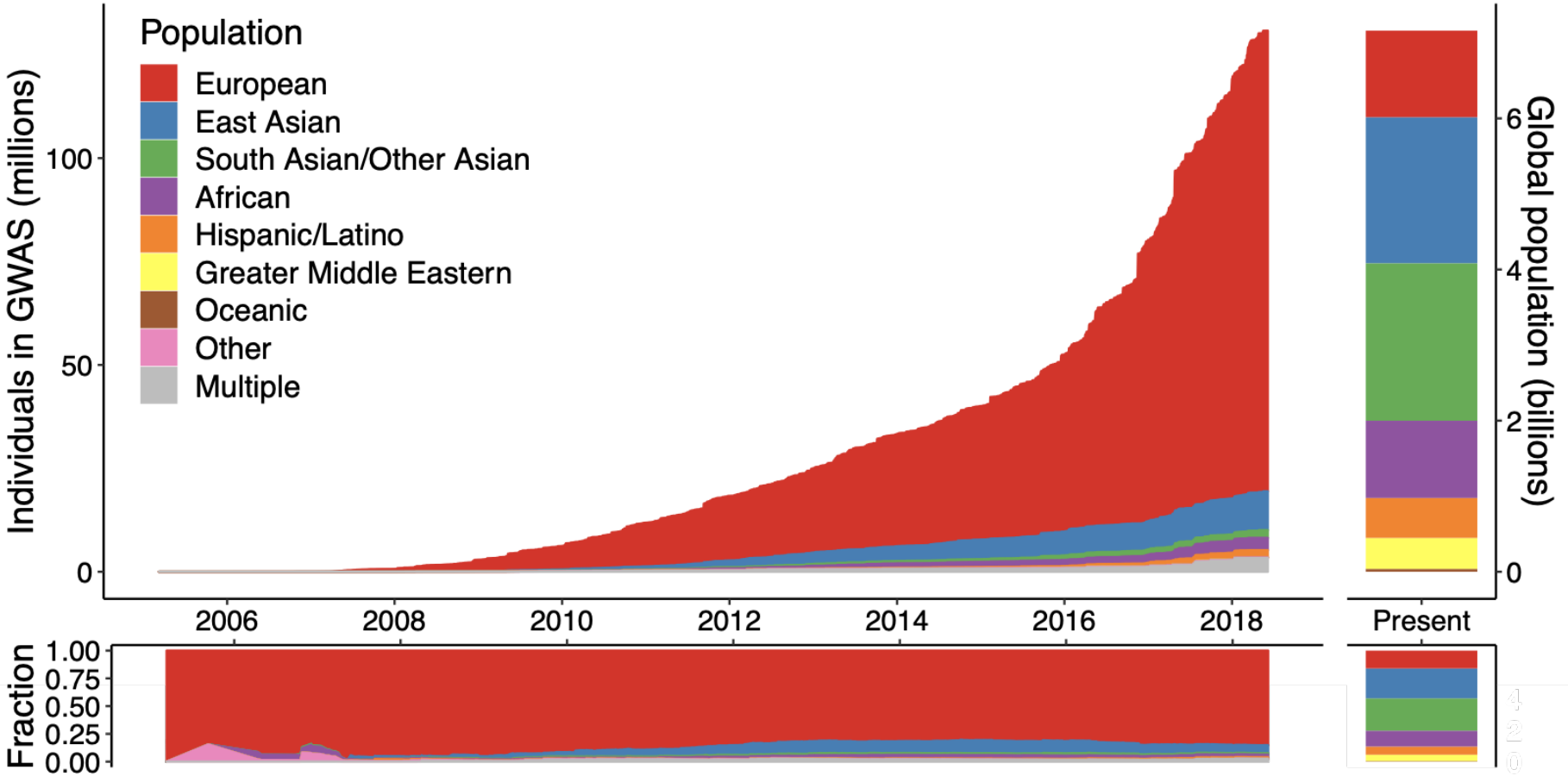
@genetisaur

Genetic variation reflects human history



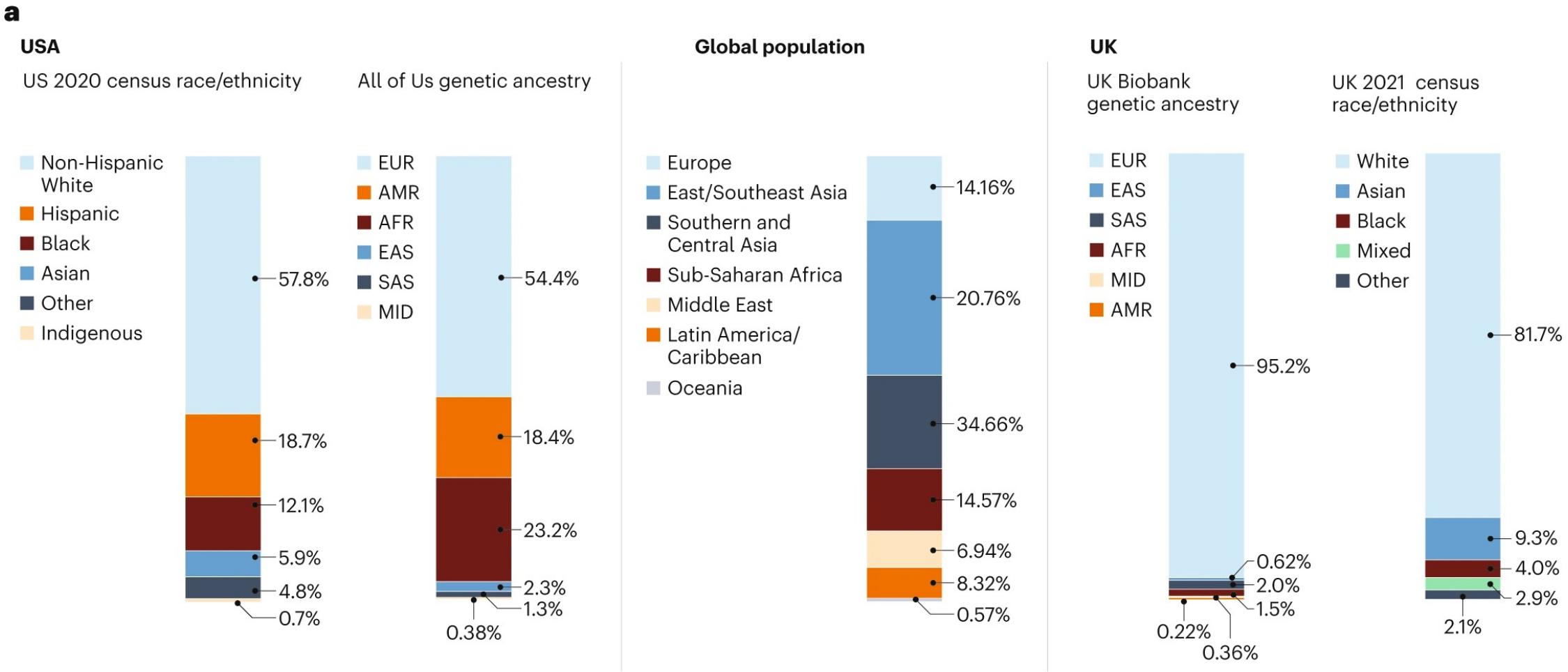
Henn, Cavalli-Sforza, and Feldman (2012). PNAS

Equity in genomics: a long shot?



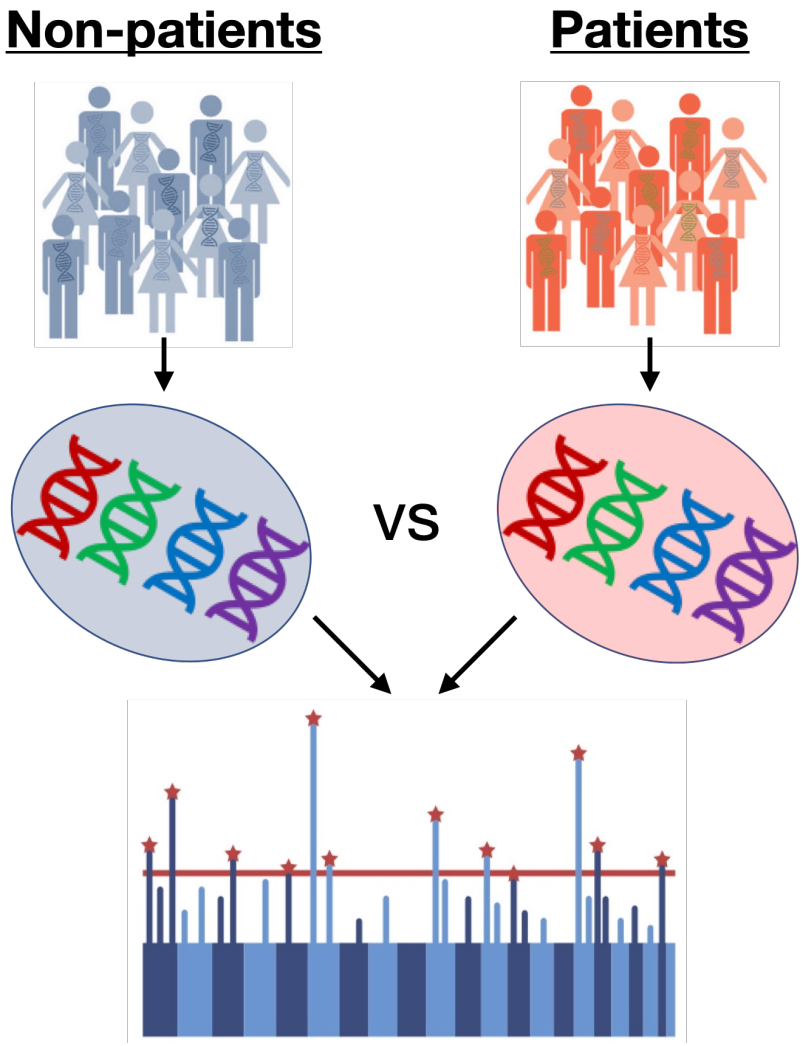
Martin, A. R., *et al.*, (2019) Nature Genetics

National biobanks do not match global or national diversity

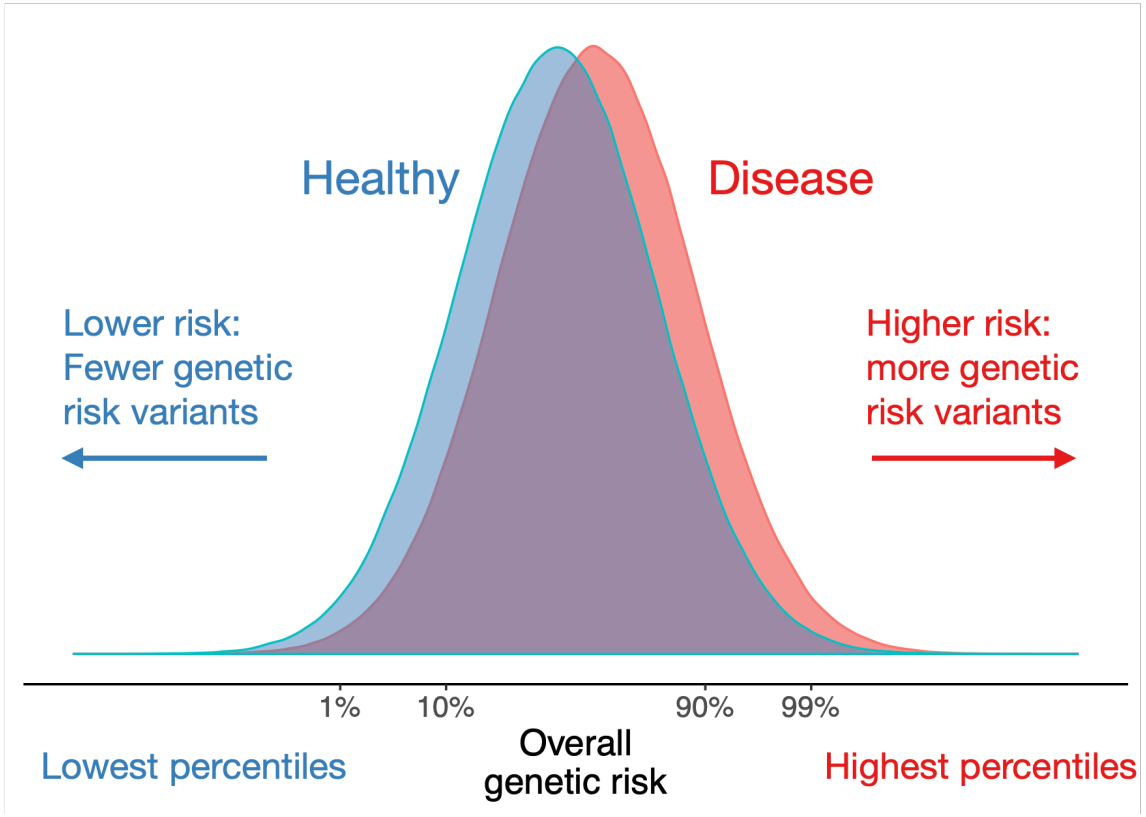


He, Y., and Martin, A.R. (2023). We need more-diverse biobanks to improve behavioural genetics. Nat Hum Behav.

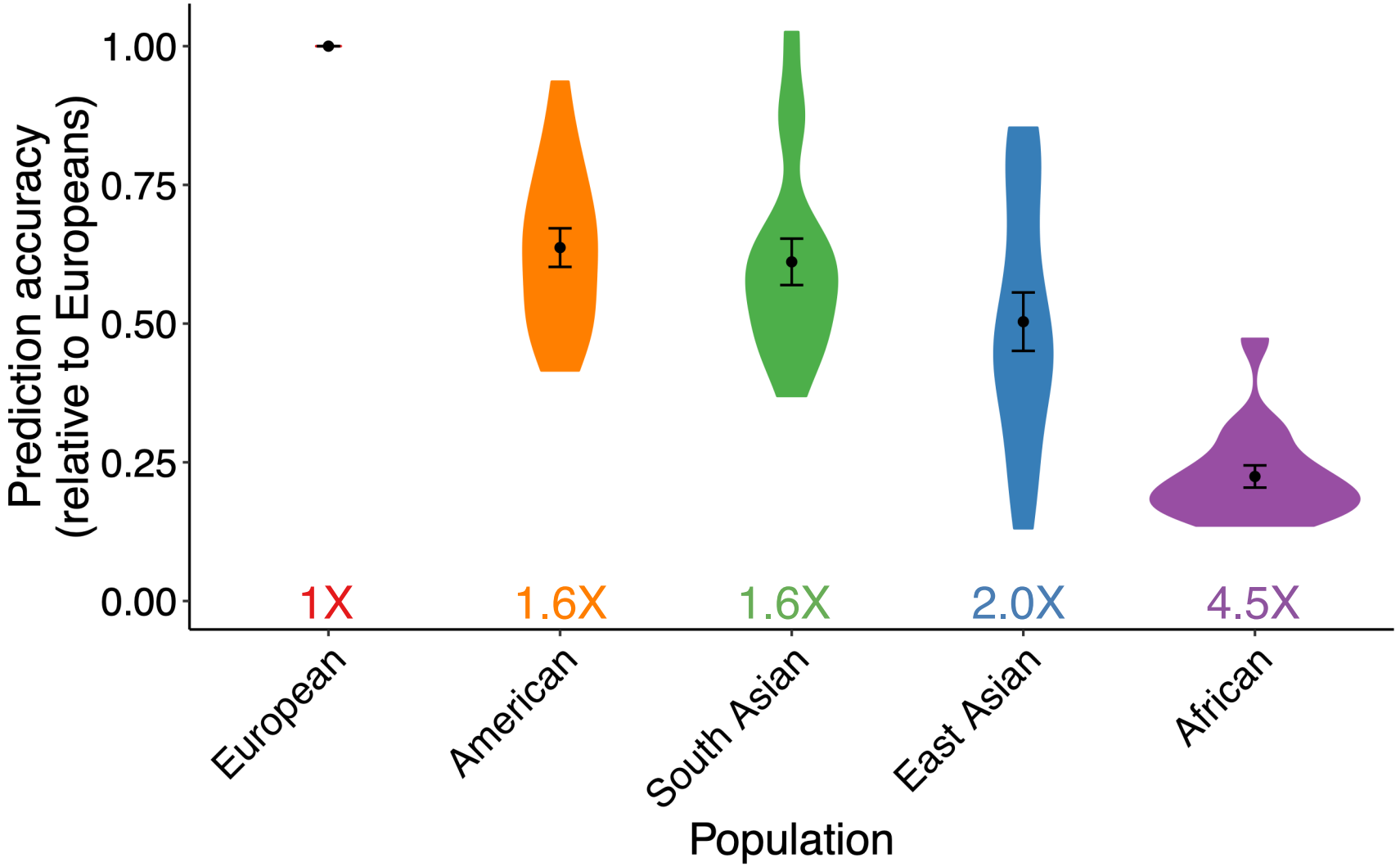
The rise of the polygenic score



Polygenic risk score: risk prediction of an individual's phenotype from DNA



Staggering disparities in accuracy



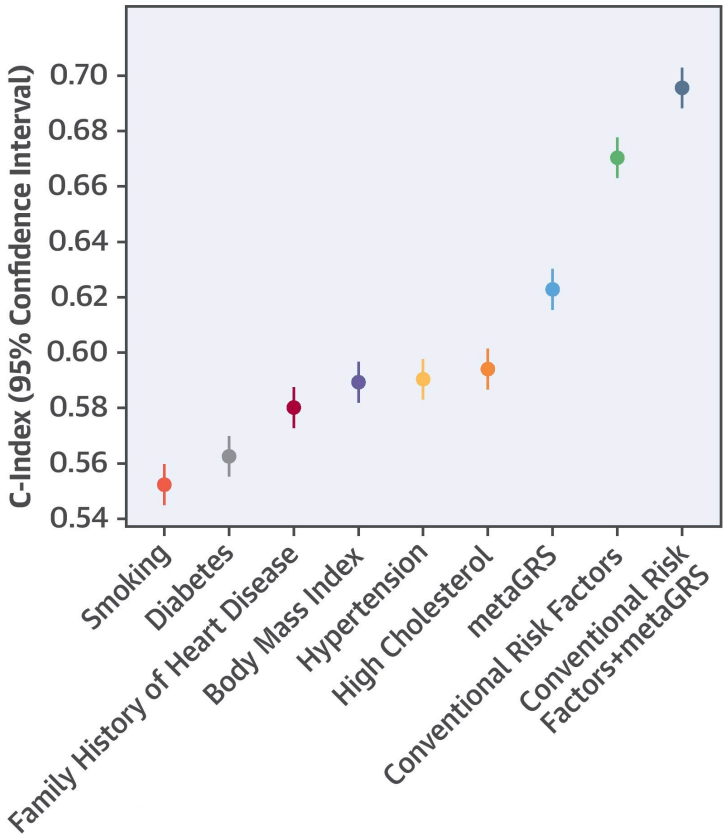
biobank uk
Improving the health of future generations

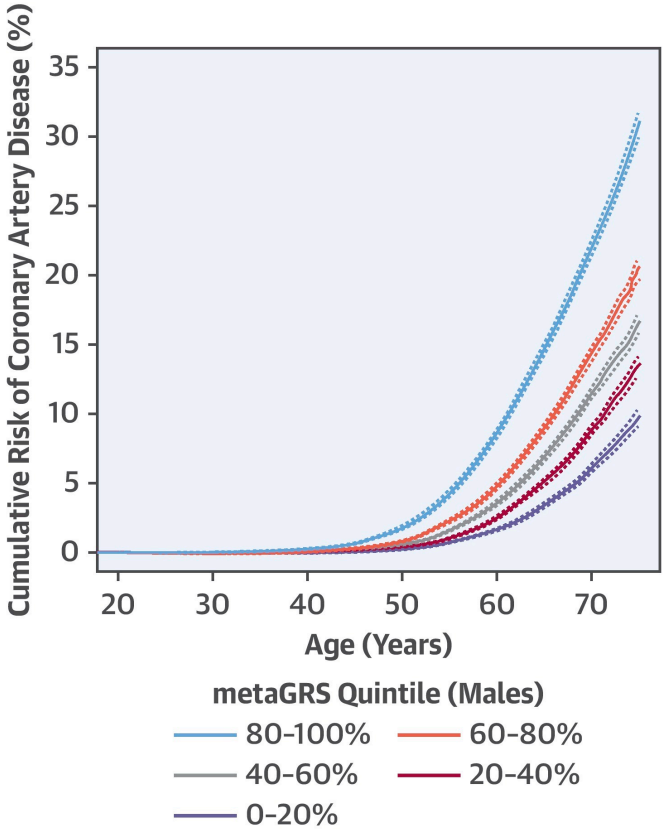
Why?

- Causal variants are shared
- *Estimated effects* (i.e. from GWAS) vary with allele frequency, LD

Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).

PRS may help prevent heart disease when considered alongside known risk factors





Inouye, M., et al. (2018). Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J. Am. Coll. Cardiol. *72*, 1883–1893.

PRS (in European ancestry) already perform better than well-known *individual* heart disease risk factors, improves existing models

Variants along allele frequency spectrum inform risk additively (e.g. breast cancer)

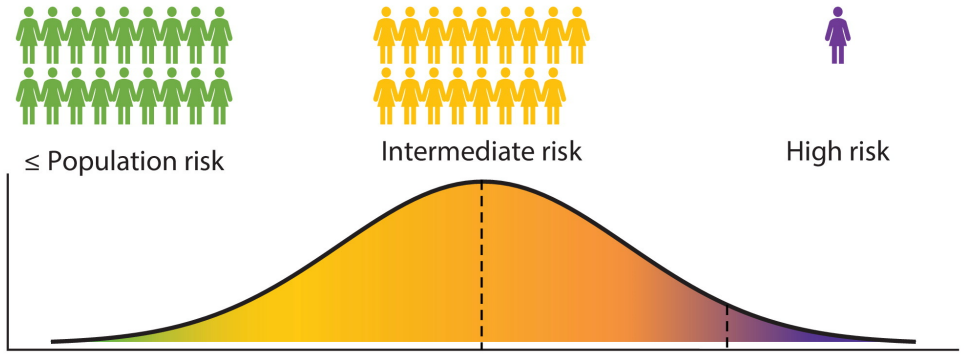
- BOADICEA =Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm
- Note use of **absolute** risk, varies by population (e.g. prostate cancer)

BOADICEA risk factors considered:

- **Family history:** cancers (breast, ovarian, prostate, and pancreatic)
- **Age** (at cancer diagnosis or current/death for all unaffected family members)
- **Genetic risk** (rare variants in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, and *ATM* and common variants from PRS)
- **Lifestyle/hormonal/reproductive:** height, BMI, parity, age at first birth, age at menarche, age at menopause, use of oral contraceptive, use of hormone replacement therapy, alcohol intake
- **Mammographic density**
- **Breast tumor morphology:** ER, progesterone receptor, HER2 receptor, CK14, CK5/6 status
- **Demographic:** country of origin, Ashkenazi Jewish origin

Combined model
(allows for missing information)

$$Y \sim X\hat{\beta}_{\text{Fam hx}} + X\hat{\beta}_{\text{Genetics}} + \ldots + X\hat{\beta}_{\text{Lifestyle}}$$



Lifetime risk < 17%

Screening
- Mammogram every 2 years (50–69)
- Clinical breast exam every 1–2 years

17% ≤ Lifetime risk < 30%

Screening
- Mammogram every 1–2 years (40+)
- If breast density >75%: annual mammogram, consider ultrasound
- Clinical breast exam (annual)

Lifetime risk ≥ 30%

Screening
- Annual mammogram (35+ if MRI, 30–35 otherwise)
- Consider annual MRI (30+)
- Clinical breast exam (annual)

Prevention
- Chemoprevention?
- Mastectomy?

Wang Y, et al. 2022
Annu. Rev. Biomed. Data Sci. 5:293–320

Key takeaways

- Eurocentric genetic study biases influence the accuracy of polygenic scores – highest in training population, decays with genetic divergence based on human history
- PRS will never be diagnostic. Use is in risk stratification context and should be integrated alongside other known clinical risk factors
- PRS can also be modeled with known rare variants
- Differences in prevalence motivate use of absolute risk and evaluation metrics