POLYGENIC SCORES FOR HEIGHT IN ADMIXED POPULATIONS

Bárbara D. Bitarello[1] & Iain Mathieson[1]

[1,] University of Pennsylvania, Perelman School of Medicine, Genetics Department

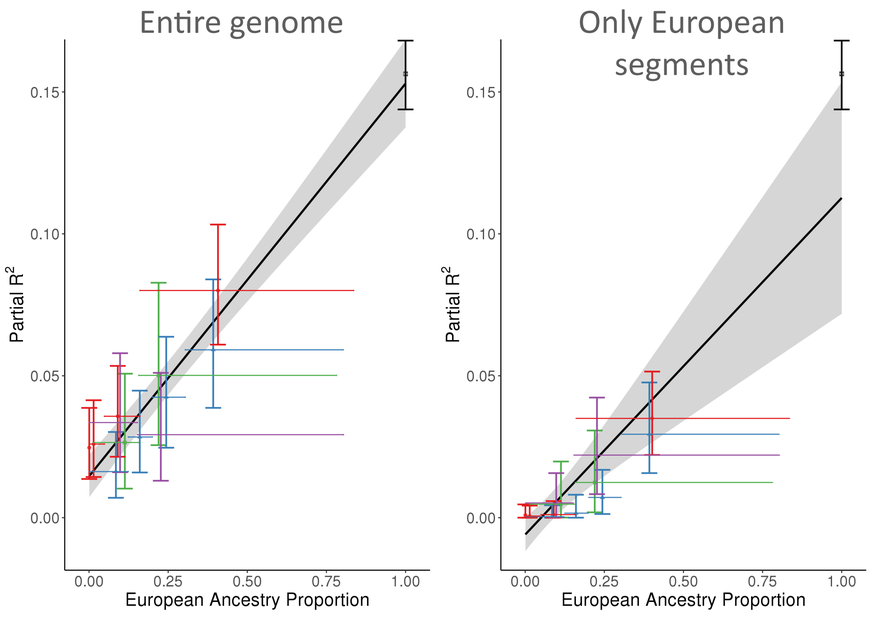
Scan the code or click here for preprint

Dataset	Group	*N*
UK Biobank (UKB_eur)	European	9,998
UK Biobank (UKB_afr)	African + European	8,700
Women's Health Initiative (WHI_afr)	African American	6,863
Jackson Heart Study (JHS_afr)	African American	1,773
Health & Retirement Study (HRS_afr)	African American	2,251
Health & Retirement Study (HRS_eur)	European American	10,159

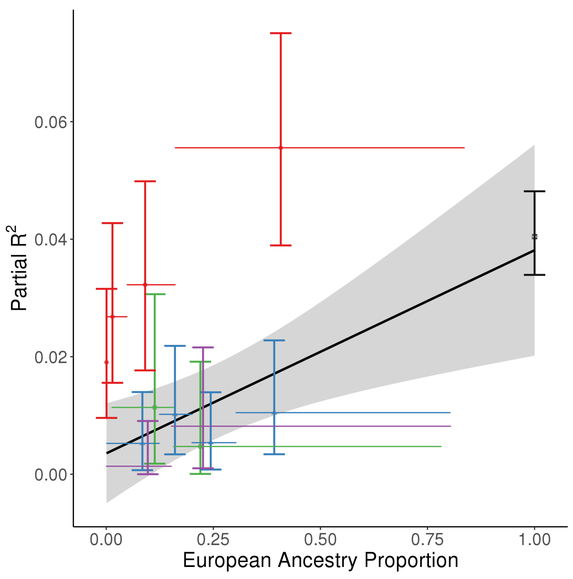
BACKGROUND

PRS-based phenotypic predictions based on European GWAS transfer poorly to other ancestry groups. A major barrier to the use of polygenic risk scores (PRS) is that the majority of GWAS (80%) are carried out in cohorts of European ancestry. We investigate the genetic and non-genetic reasons driving this pattern in height PRS derived from European GWAS when used to predict height in admixed individuals with both African and European ancestry.

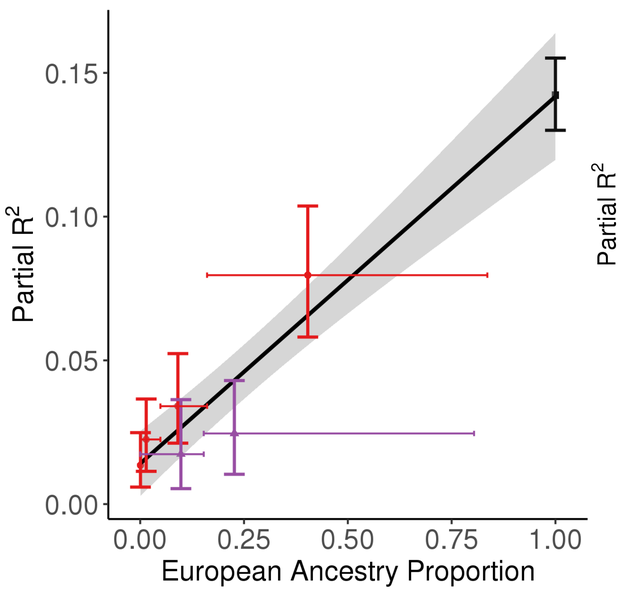
PREDICTIVE POWER INCREASES LINEARLY WITH PROPORTION OF EUROPEAN ANCESTRY



POPULATION STRUCTURE?

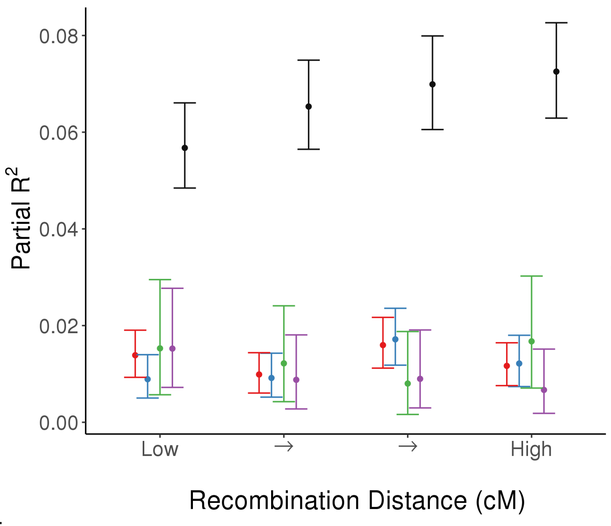


PRS from sibling-pair effect sizes display same pattern as PRS from GWAS effect sizes.



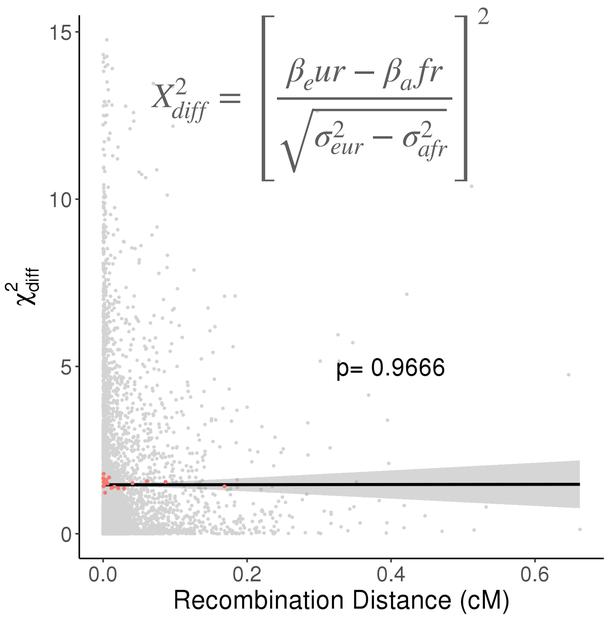
Imputed data: similar pattern as that observed for genotyped data.

DIFFERENCES IN LINKAGE DISEQUILIBRIUM?



Partial-R^2 84% and 76% lower for admixed individuals for 4th and 1st quantiles of recombination rate, respectively.

$$X^2_{diff} = \left[\frac{\beta_e ur - \beta_a fr}{\sqrt{\sigma^2_{eur} - \sigma^2_{afr}}}\right]^2$$



Differences in effect size do not increase with recombination rate.

ALLELIC FREQUENCY DIFFERENCES?

Dataset	Genetic variance ratio
UKB	0.78
WHI	1.07
JHS	1.04
HRS	0.92

$$= \frac{\sum_{i=1}^{n} 2f_1(1-f_1)\beta_i^2}{\sum_{i=1}^{n} 2f_2(1-f_2)\beta_i^2}$$

f_1=allele frequency in EUR

Frequency spectrum differences in PRS SNPs not large enough to explain reduction in R^2.

ANCESTRY-SPECIFIC EFFECT SIZES IN PRS?

PRS_C^1 weights the African PRS by α:

$$PRS_C^1 = \alpha PRS_{AFR} + (1-\alpha)PRS_{EUR}$$

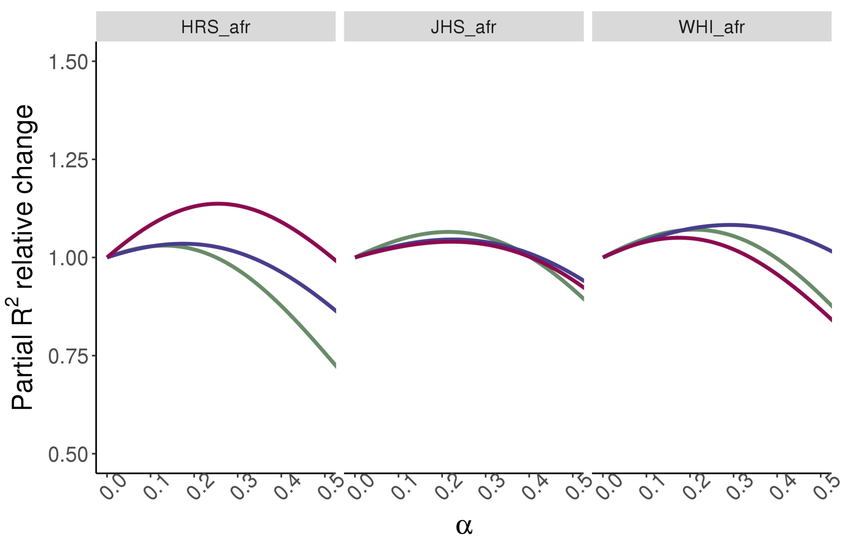
PRS_C^2 weights the African PRS by α:

$$PRS_C^2 = \alpha(1-p_{eur,j})PRS_{AFR,j} + (1-\alpha p_{eur,j})PRS_{EUR,j}$$

PRS_C^3 is defined for each haplotype in each individual, and weights SNPs in African segments by α:

$$PRS_C^3 = \alpha\left[\sum_{i\in AFR}\beta_{i,afr}G_i\right] + (1-\alpha)\left[\sum_{i\in AFR}\beta_{i,eur}G_i\right] + \left[\sum_{i\in EUR}\beta_{i,eur}G_i\right]$$

j, individuals; i, SNPs, *G*, genotype



CONCLUSIONS

- differences in LD structure and SFS do affect the transferability of PRS but do not, by themselves, explain the magnitude of the decrease in predictive power.
- Marginal effect sizes differ across ancestries and therefore:
  -prediction for admixed individuals can be improved by using a linear combination of PRS that includes ancestry-specific effect sizes — at present limited by the small size of non-European ancestry discovery cohorts.
  -large cohorts of diverse ancestries are needed in order to make PRS applicable to diverse ancestry groups and admixed populations.

FUNDING

THE PITTSBURGH FOUNDATION | Charles E. Kaufman Foundation

Alfred P. Sloan FOUNDATION

NIH NIGMS