Dissecting the predictive accuracy of polygenic indexes for behavioral phenotypes across genetic ancestries

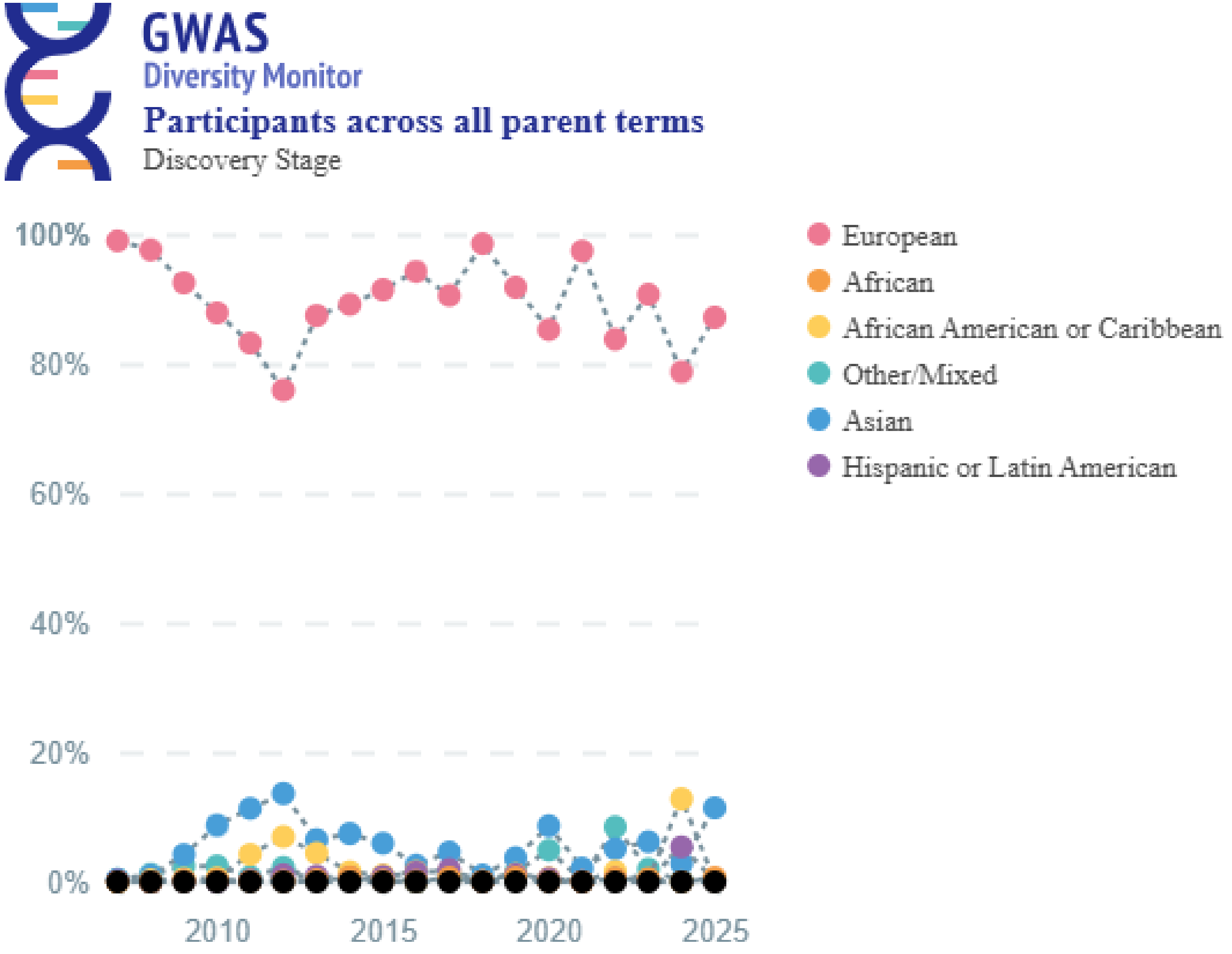
Aysu Okbay

Amsterdam UMC

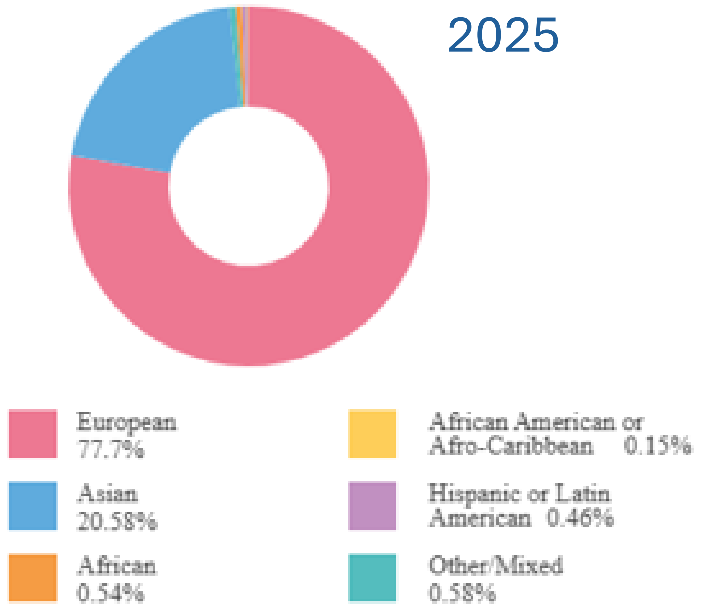
ESSGN Amsterdam Training

March 26, 2026

GWAS samples are dominated by individuals of European genetic ancestries

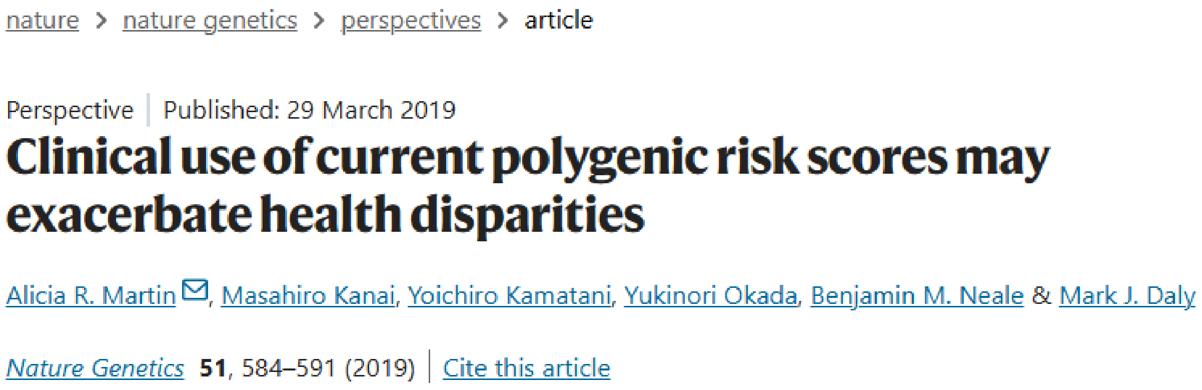


Please cite this as: Mills, M.C and Rahal, C., (2020). 'The GWAS Diversity Monitor Tracks diversity by disease in real time'. Nature Genetics, 52, 242-243. doi: 10.1038/s41588-020-0580-y



Please cite this as: Mills, M.C and Rahal, C., (2020). 'The GWAS Diversity Monitor Tracks diversity by disease in real time'. Nature Genetics, 52, 242-243. doi: 10.1038/s41588-020-0580-y

PGI predictive accuracy is lower in ancestrally divergent samples



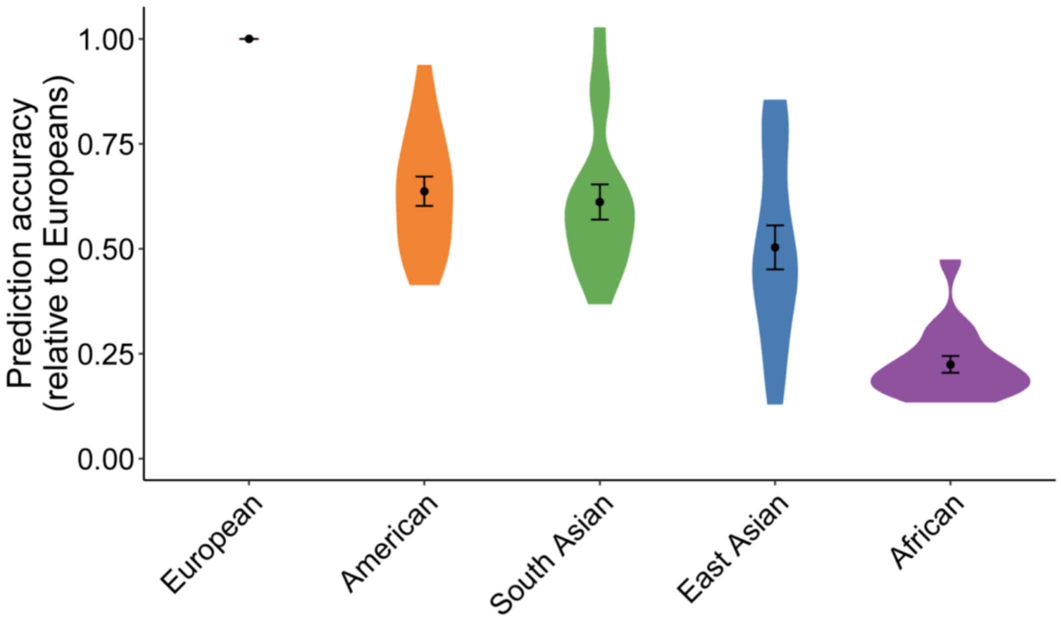
nature > nature genetics > perspectives > article

Perspective | Published: 29 March 2019

Clinical use of current polygenic risk scores may exacerbate health disparities

Alicia R. Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale & Mark J. Daly

Nature Genetics **51**, 584–591 (2019) | Cite this article

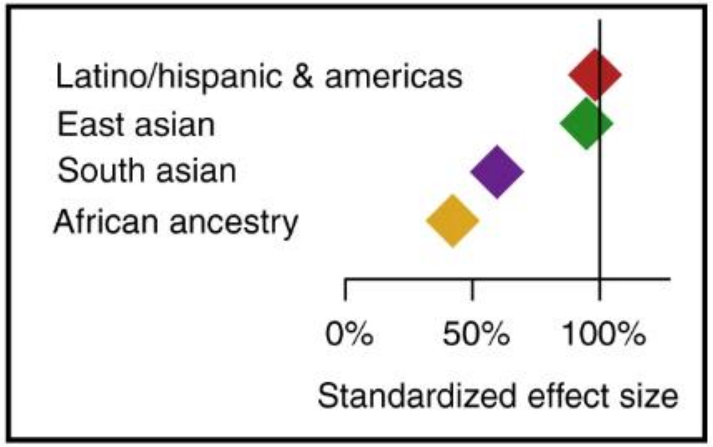


Article | Open access | Published: 25 July 2019

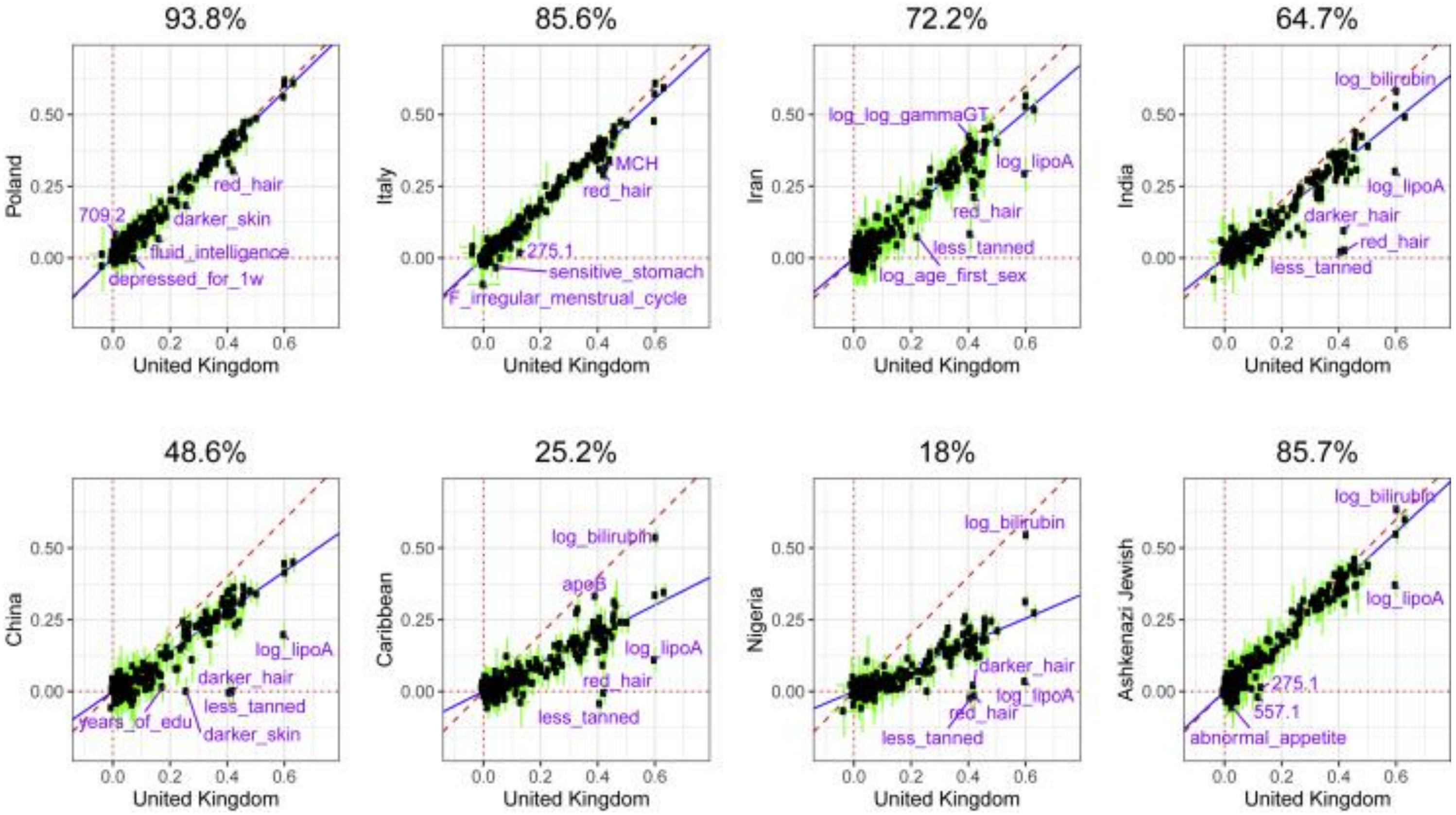
Analysis of polygenic risk score usage and performance in diverse human populations

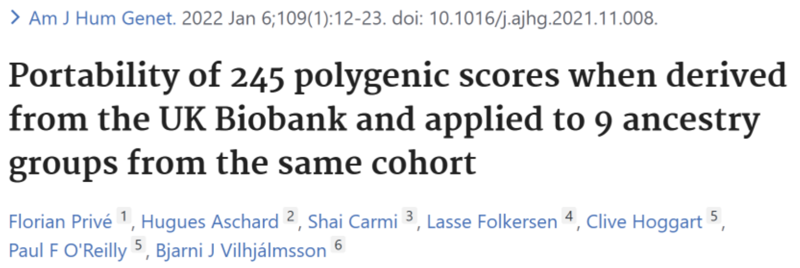
L. Duncan, H. Shen, B. Gelaye, J. Meijsen, K. Ressler, M. Feldman, R. Peterson & B. Domingue

Nature Communications **10**, Article number: 3328 (2019) | Cite this article



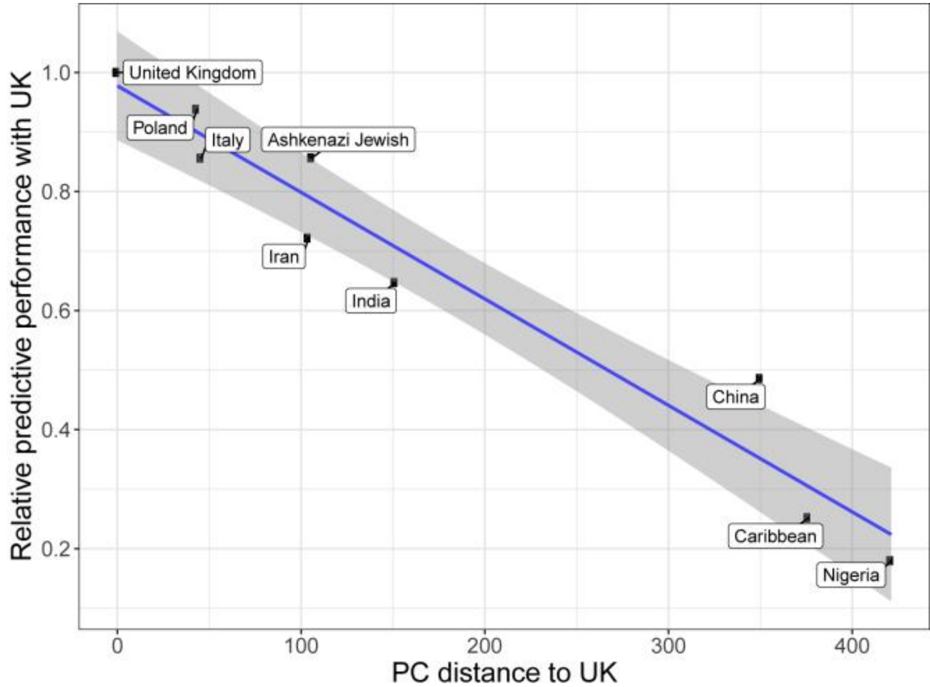
Predictive accuracy decays as a function of genetic distance from discovery sample




> Am J Hum Genet. 2022 Jan 6;109(1):12-23. doi: 10.1016/j.ajhg.2021.11.008.

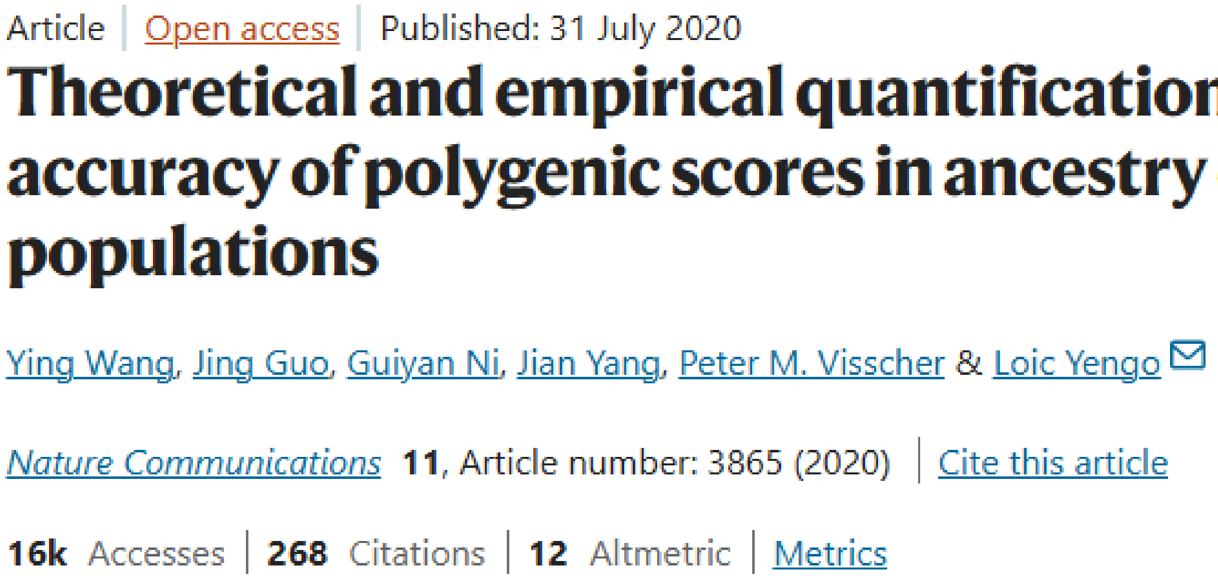
Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort

Florian Privé [1], Hugues Aschard [2], Shai Carmi [3], Lasse Folkersen [4], Clive Hoggart [5], Paul F O'Reilly [5], Bjarni J Vilhjálmsson [6]



WHY is PGI predictive accuracy is lower in genetically divergent samples?

- Cross-population differences in:
  - Allele frequencies (MAF)
  - Linkage disequilibrium (LD): causal SNPs – SNPs assayed in GWAS
  - Heritability
  - Causal variant effect sizes, population-specific causal SNPs
  - Negative selection
  - Gene-environment correlations
  - GxE interactions
  - G-E correlations
  - Assortative mating



Article | Open access | Published: 31 July 2020

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

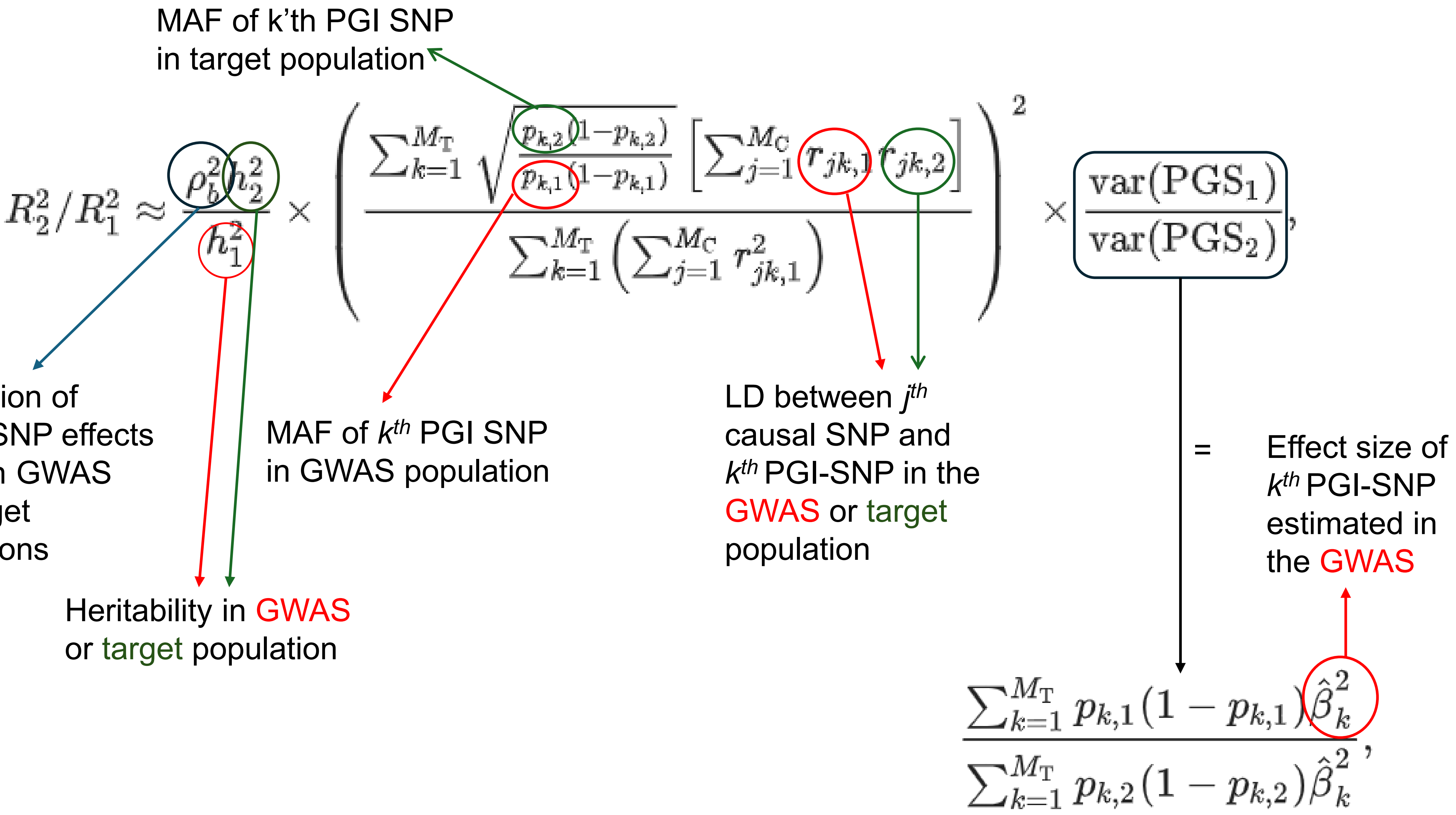
Ying Wang, Jing Guo, Guiyan Ni, Jian Yang, Peter M. Visscher & Loic Yengo

Nature Communications **11**, Article number: 3865 (2020) | Cite this article

16k Accesses | **268** Citations | **12** Altmetric | Metrics

- Calculate the theoretically expected relative accuracy of the PGI based only on cross-ancestry LD and MAF differences
- Compare to observed relative accuracy

→ Loss in predictive accuracy due to cross-ancestry LD and MAF differences



MAF of k’th PGI SNP
in target population

$$R_2^2/R_1^2 \approx \frac{\rho_b^2 h_2^2}{h_1^2} \times \left(\frac{\sum_{k=1}^{M_{\mathrm{T}}} \sqrt{\frac{p_{k,2}(1-p_{k,2})}{p_{k,1}(1-p_{k,1})}} \left[\sum_{j=1}^{M_{\mathrm{C}}} r_{jk,1} r_{jk,2}\right]}{\sum_{k=1}^{M_{\mathrm{T}}} \left(\sum_{j=1}^{M_{\mathrm{C}}} r_{jk,1}^2\right)} \right)^2 \times \boxed{\frac{\mathrm{var}(\mathrm{PGS}_1)}{\mathrm{var}(\mathrm{PGS}_2)}},$$

Correlation of
causal SNP effects
between GWAS
and target
populations

MAF of k^{th} PGI SNP
in GWAS population

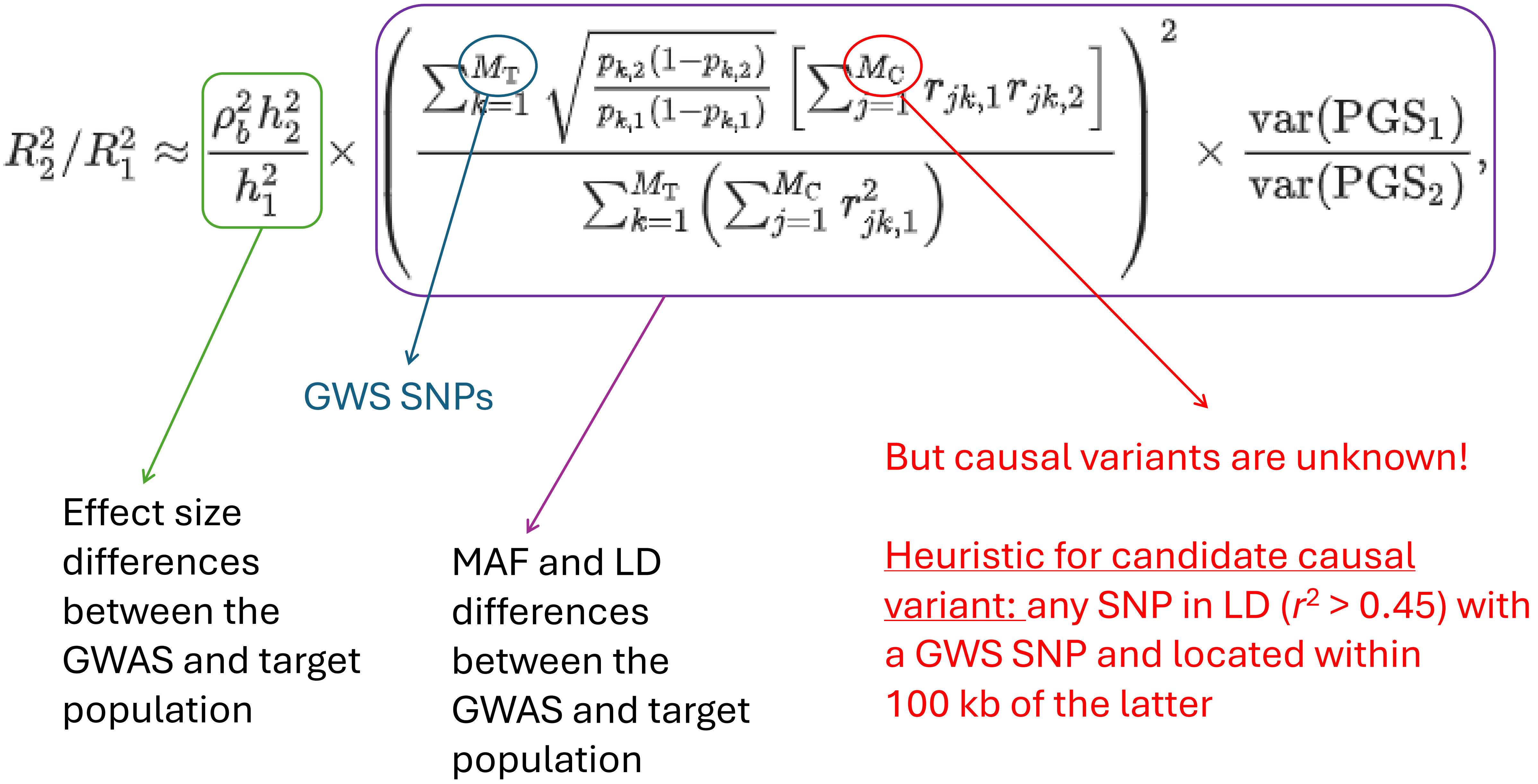
LD between j^{th}
causal SNP and
k^{th} PGI-SNP in the
GWAS or target
population

Heritability in GWAS
or target population

=

$$\frac{\sum_{k=1}^{M_{\mathrm{T}}} p_{k,1}(1-p_{k,1})\hat{\beta}_k^2}{\sum_{k=1}^{M_{\mathrm{T}}} p_{k,2}(1-p_{k,2})\hat{\beta}_k^2},$$

Effect size of
k^{th} PGI-SNP
estimated in
the GWAS



$$R_2^2/R_1^2 \approx \boxed{\frac{\rho_b^2 h_2^2}{h_1^2}} \times \boxed{\left(\frac{\sum_{k=1}^{M_{\mathrm{T}}}\sqrt{\frac{p_{k,2}(1-p_{k,2})}{p_{k,1}(1-p_{k,1})}}\left[\sum_{j=1}^{M_{\mathrm{C}}} r_{jk,1}r_{jk,2}\right]}{\sum_{k=1}^{M_{\mathrm{T}}}\left(\sum_{j=1}^{M_{\mathrm{C}}} r_{jk,1}^2\right)}\right)^2 \times \frac{\mathrm{var}(\mathrm{PGS}_1)}{\mathrm{var}(\mathrm{PGS}_2)}},$$

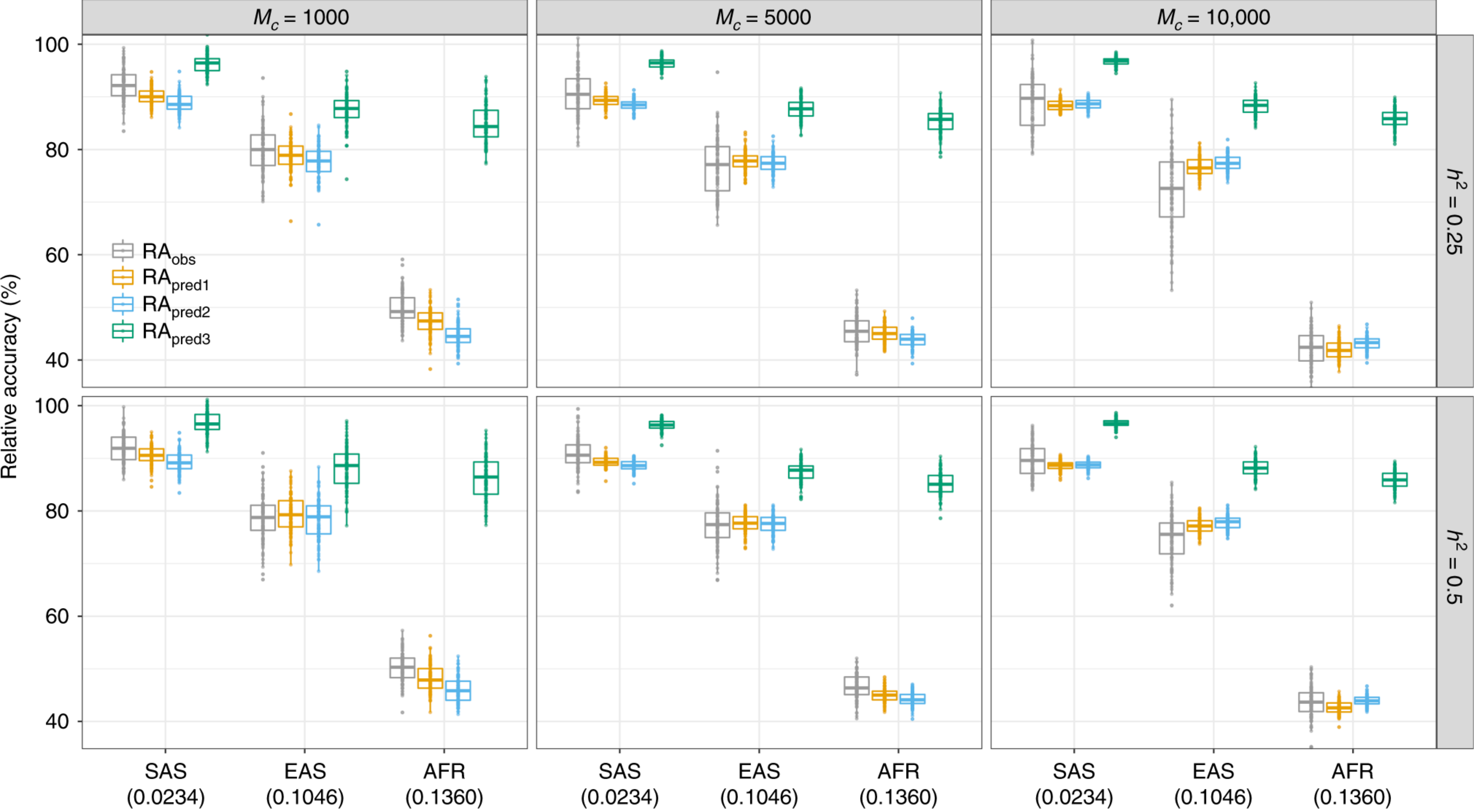
GWS SNPs

Effect size differences between the GWAS and target population

MAF and LD differences between the GWAS and target population

But causal variants are unknown!

Heuristic for candidate causal variant: any SNP in LD ($r^2 > 0.45$) with a GWS SNP and located within 100 kb of the latter


$M_c = 1000$
$M_c = 5000$
$M_c = 10,000$
$h^2 = 0.25$
$h^2 = 0.5$
Relative accuracy (%)
100
80
60
40
RA_{obs}
RA_{pred1}
RA_{pred2}
RA_{pred3}
SAS (0.0234)
EAS (0.1046)
AFR (0.1360)

$$RA_{Expected} = \frac{R^2_{2,exp}}{R^2_{1,exp}}$$

$$RA_{Observed} = \frac{R^2_{2,obs}}{R^2_{1,obs}}$$

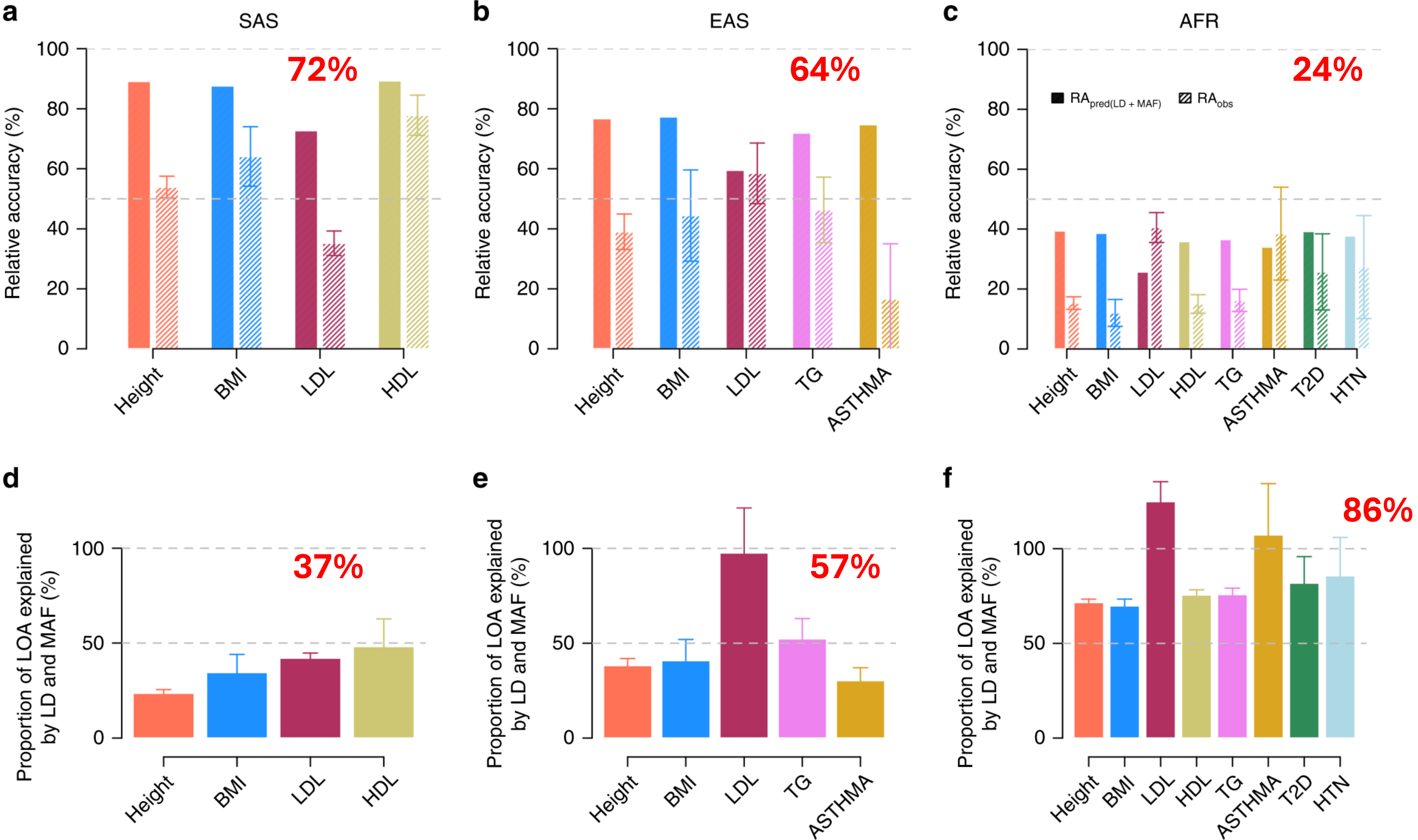
Loss of Accuracy (LoA)

$$\left(1 - RA_{Expected}\right) * 100$$

$$\left(1 - RA_{Observed}\right) * 100$$

LoA due to MAF+LD

$$LoA_{(MAF+LD))} = \left(\frac{1 - RA_{Expected}}{1 - RA_{Observed}}\right) * 100$$


a
SAS
Relative accuracy (%)
0
20
40
60
80
100
72%
Height
BMI
LDL
HDL
b
EAS
Relative accuracy (%)
0
20
40
60
80
100
64%
Height
BMI
LDL
TG
ASTHMA
c
AFR
Relative accuracy (%)
0
20
40
60
80
100
24%
■ $RA_{pred(LD + MAF)}$
▨ RA_{obs}
Height
BMI
LDL
HDL
TG
ASTHMA
T2D
HTN
d
Proportion of LOA explained by LD and MAF (%)
0
50
100
37%
Height
BMI
LDL
HDL
e
Proportion of LOA explained by LD and MAF (%)
0
50
100
57%
Height
BMI
LDL
TG
ASTHMA
f
Proportion of LOA explained by LD and MAF (%)
0
50
100
86%
Height
BMI
LDL
HDL
TG
ASTHMA
T2D
HTN

Dissecting the Predictive Accuracy of Polygenic Indexes for Behavioral Phenotypes Across Genetic Ancestries

Robel Alemu, Alexander S. Young, Daniel J. Benjamin, Patrick Turley, Aysu Okbay

doi: https://doi.org/10.1101/2025.09.11.675704

Contribution:

- Cohorts: UKB & HRS, and (in the revision) All of Us
- Phenotypes: 52 phenotypes from PGI Repository [46 in UKB, 38 in HRS]
  - Do social and behavioral phenotypes have weaker portability?
- PGI Repo SNP weights → obtained from largest GWAS samples
- Examine
  - the share of cross-ancestry differences in MAF, LD and heritability contributing to the loss of PGI RA
  - how using genome-wide PGIs changes things
  - how using within-family PGI weights changes things

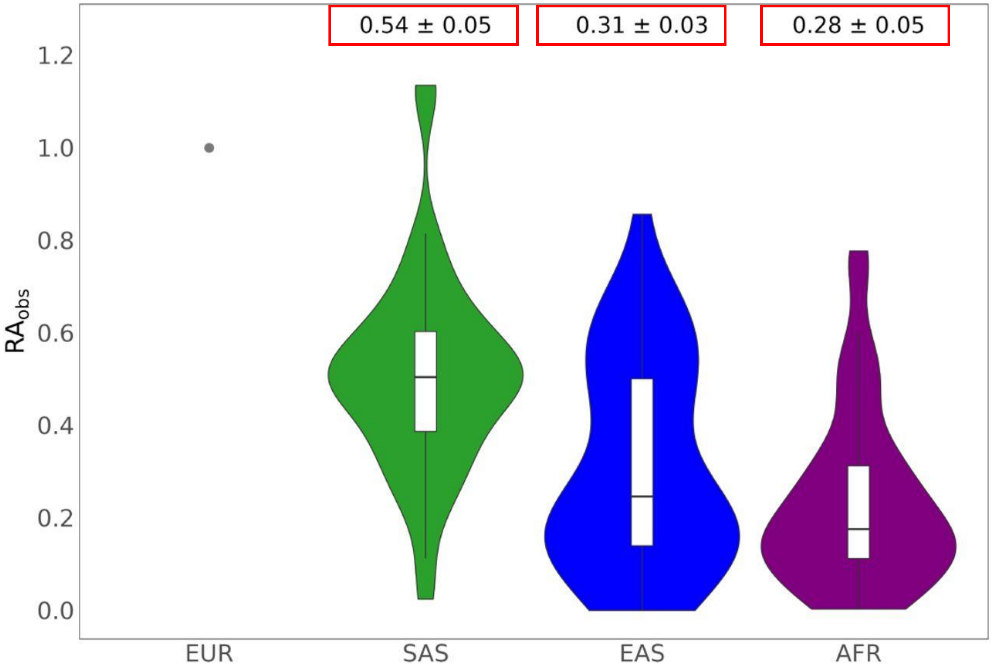
Methods overview

- Identify ancestries based on first 10 PCs
- SNP weight estimation:
  - *Standard PGIs: SBayesR: 2.8M SNPs,*
  - *fGWAS PGIs: PRS-CS: Good quality HapMap3 SNPs (filtered for INFO>0.99, etc.)*
- RA estimation: ratio of incremental R^2:
  - $RA_{Observed}$
  - $RA_{Expected}$
- Implement a modification of Wang et al. (2020) to estimate contribution of different factors to the loss of PGS predictive accuracy:
  - LD
  - MAF
  - h^2

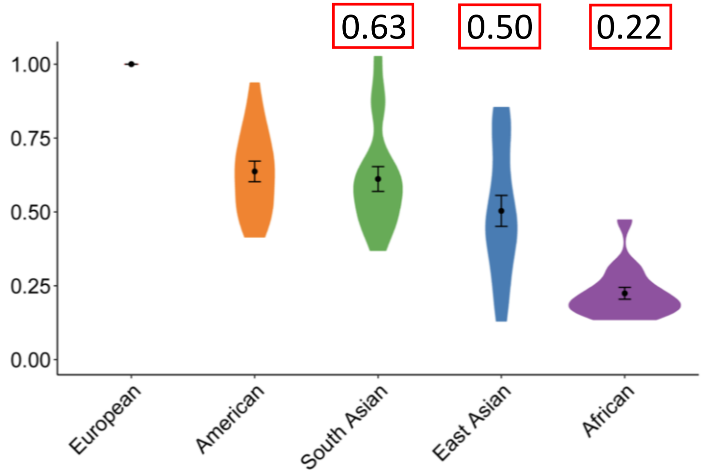
Ancestry identification

- Estimate PC weights for first 10 PCs using HapMap 3 SNPs in 1000 Genomes Ph 3
- Project UKB onto the PC space
- For each ancestry in 1000 Genomes, identify UKB/HRS individuals falling within mean ± 4 SD in all 10 PCs.
- UKB:
  - 162,963 EUR
  - 11,413 SAS
  - 2,216 EAS
  - 9,494 AFR
- HRS:
  - 12,774 EUR
  - 3,593 AFR

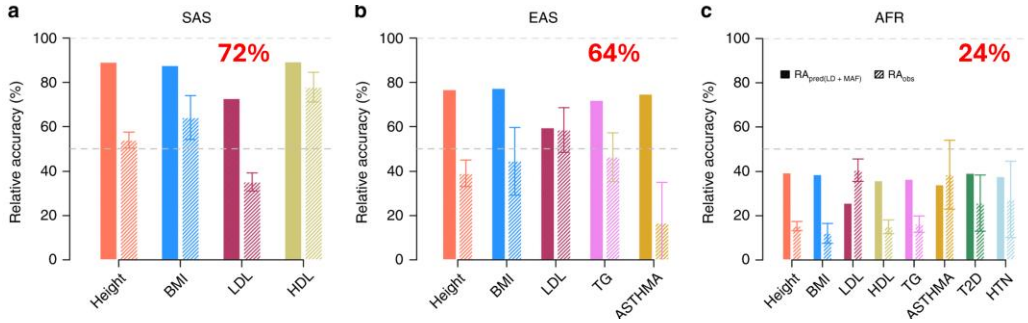
Observed relative accuracy averaged across 46 phenotypes; loss of accuracy highest in AFR, followed by EAS & SAS



Current Study
46 phenotypes in UKB [anthropometric,reproductive, cognitive, health, psychiatric, substance use, personality & well-being]

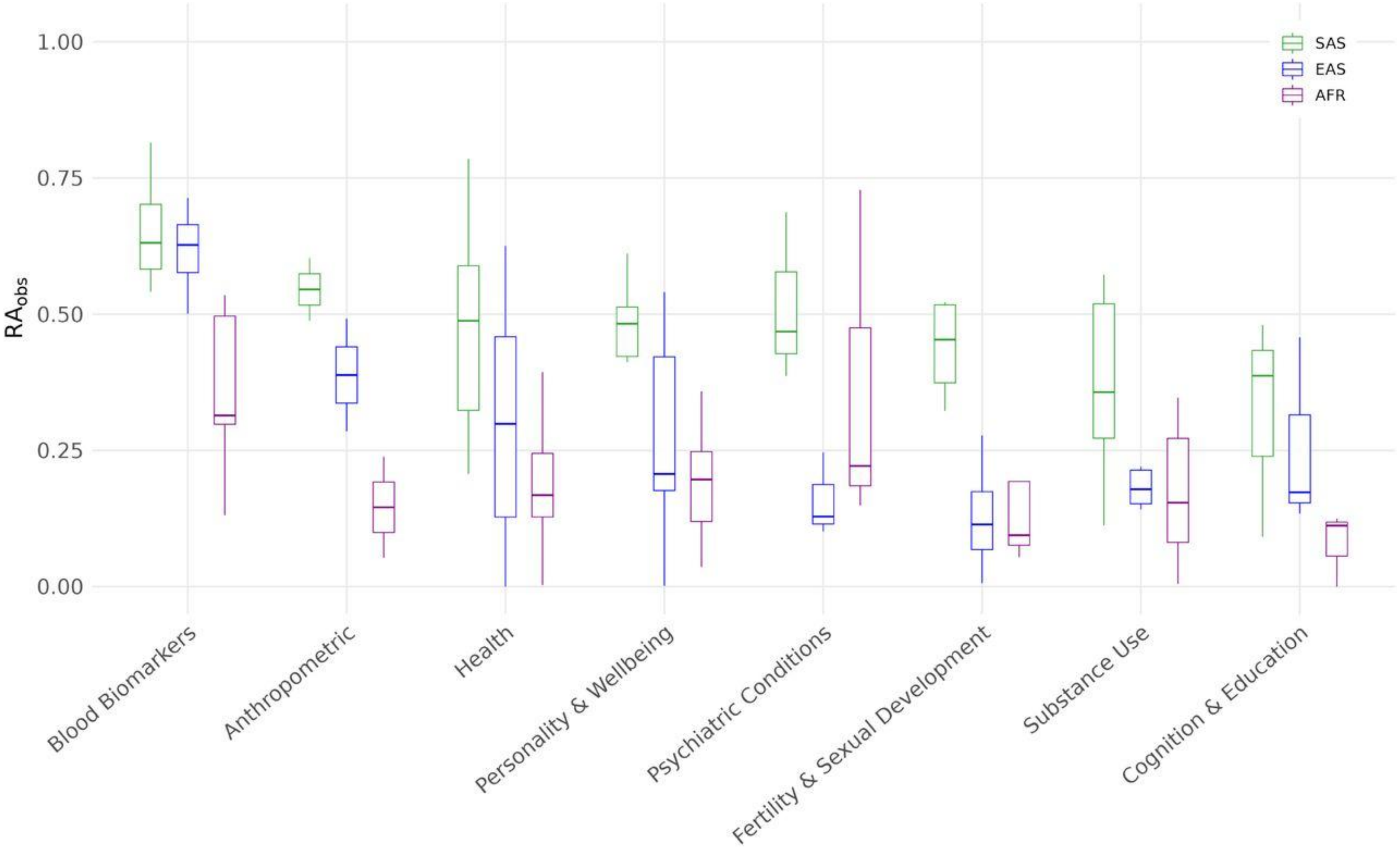


Martin et al. (2019)
17 phenotypes [anthropometric + biomarkers]

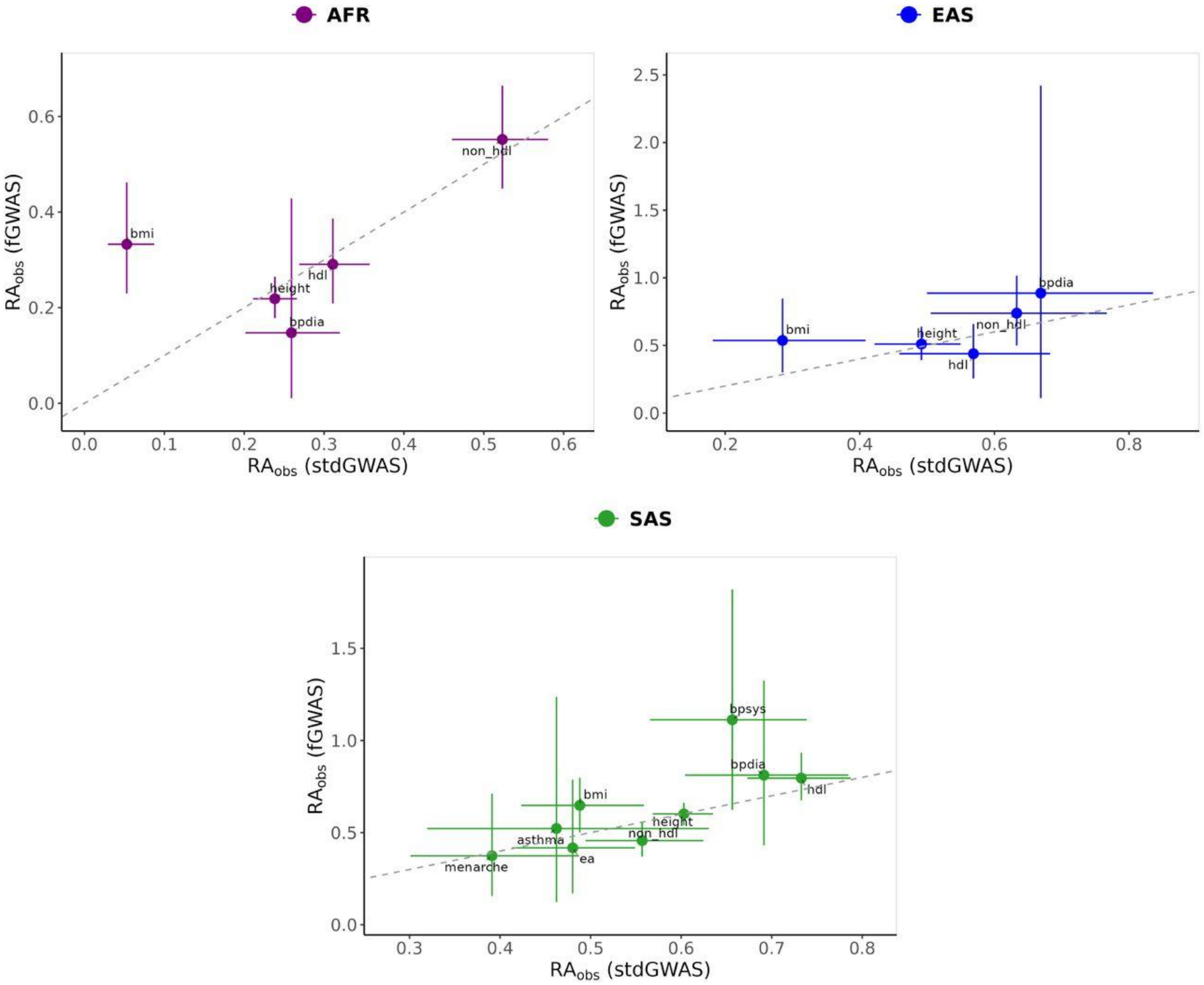


Wang et al. (2022)
8 phenotypes [anthropometric + biomarkers + metabolic]

RA_{Obs} varies by phenotype category; highest for anthropometric traits & biomarkers; higher in SAS across all categories



RA_{Obs} using standard vs fGWAS weights



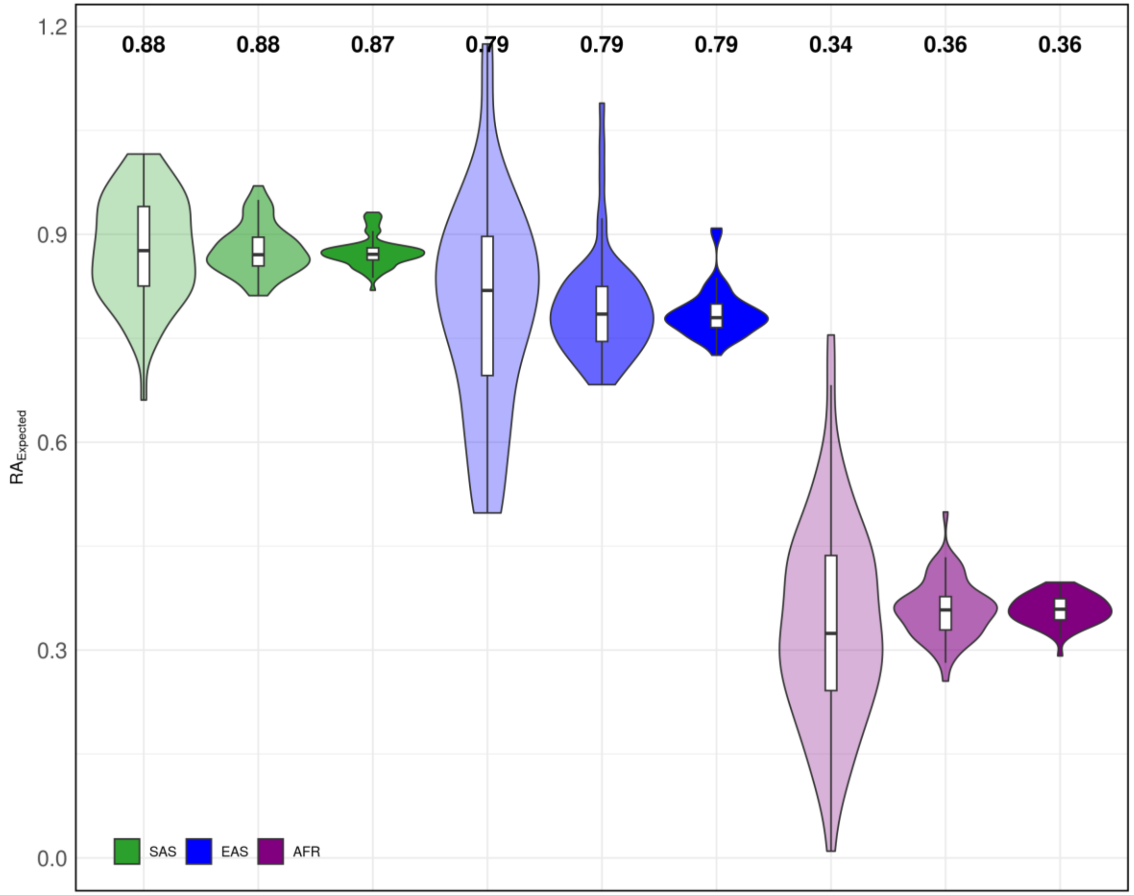
How to select candidate causal SNPs?

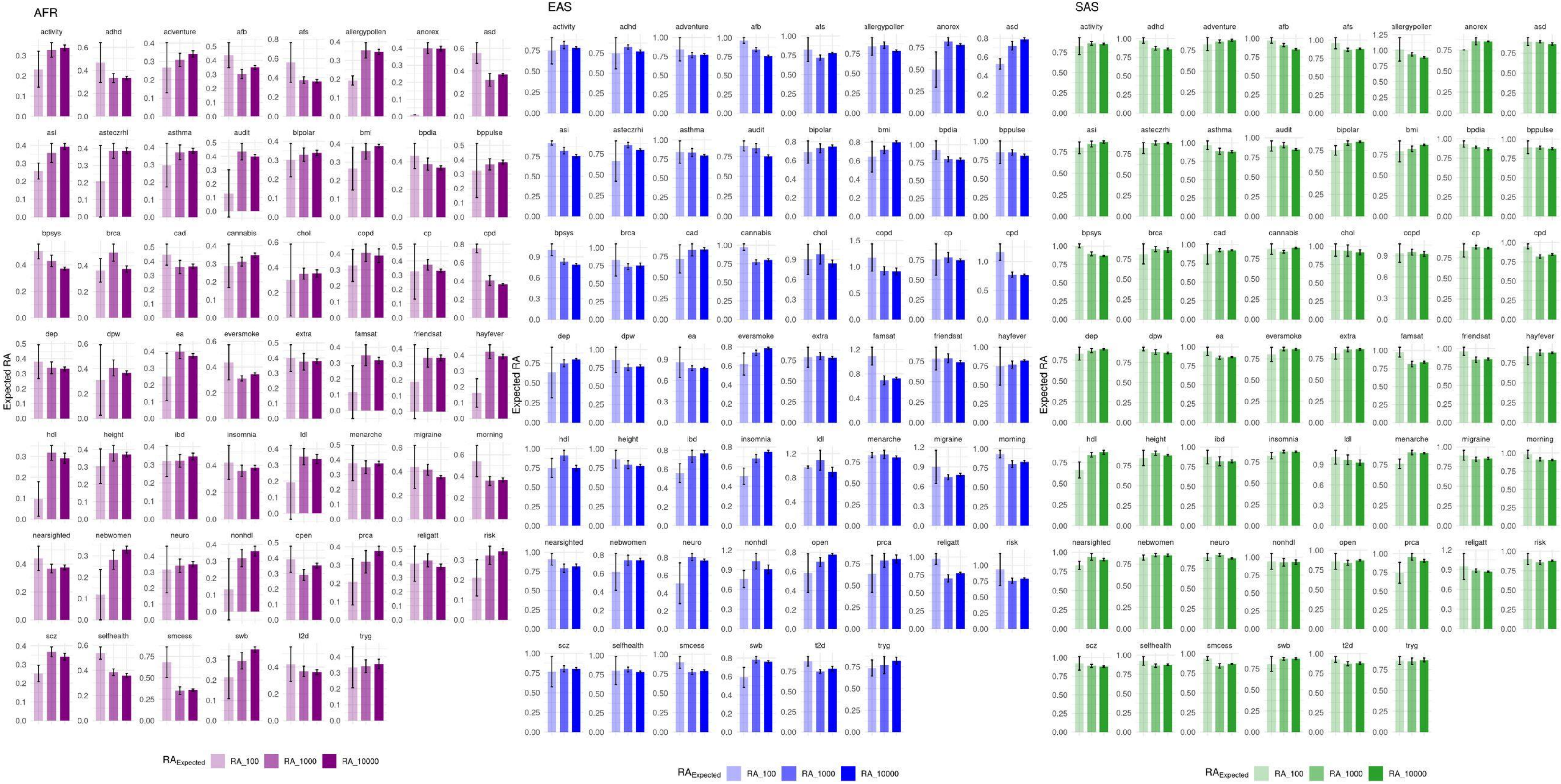
- We include ∼2.9 million SNPs in the PGIs instead of only GWS SNPs → Why?
  1. Wider set of behavioral and health-related phenotypes analyzed which are highly polygenic
     - Only a handful of GWS SNPs identified
     - Predictive accuracy of GWS SNPs is low, relative accuracy becomes imprecise
  2. Most PGI studies construct PGIs with a focus on maximizing predictive power so would be good to know if Wang *et al.*'s results hold under practically more relevant conditions.

- But then how to select candidate causal SNPs?
  - Need a set of SNPs that would approximate the genome-wide level expected predictive accuracy of PGIs
  - Because the PGI-SNPs are not selected based on their p-values and are not pairwise independent, we cannot assume that causal SNPs are located within a 100kb window and are in LD with the PGI-SNPs.
  - Use Wang *et al.*'s approach of defining candidate causal SNPs, but relax the p-value cutoff to consider more than only GWS SNPs.
    - **Challenge:** include enough SNPs to correspond to the predictive power of genome-wide PGIs while keeping the computational burden manageable

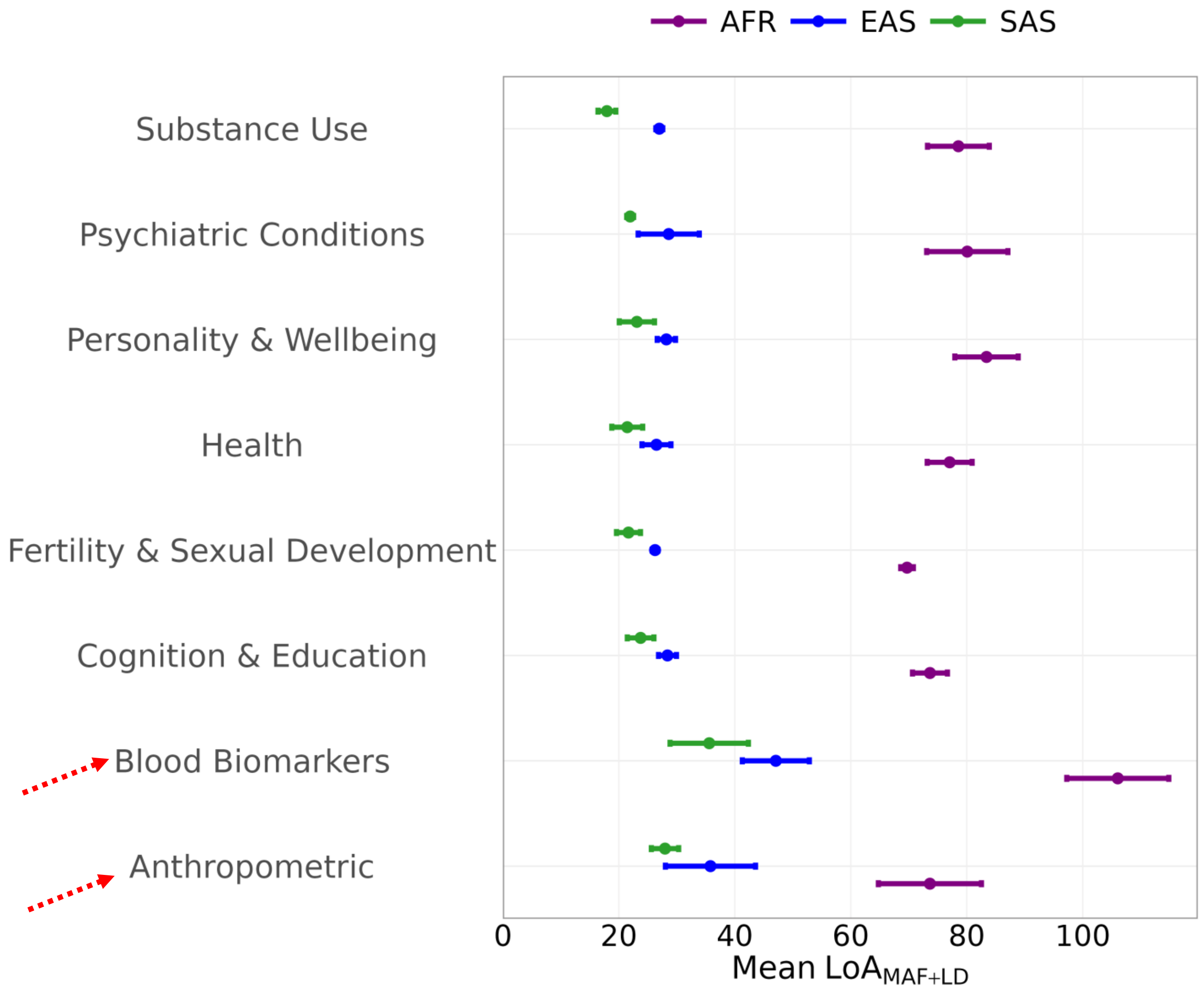
RA_{exp} stabilizes after 1000 SNPs, most precise with 10k SNPs

- Generate three different candidate causal SNP sets for each phenotype based on the top 100, 1,000, and 10,000 independent SNPs by p-value
- Expand each set to include any SNPs in LD ($R^2 > 0.45$) within 100 kb of these top SNPs
- Compute the expected RA of the PGIs based on each candidate causal SNP set
- For the fGWAS PGIs, repeat the procedure selecting top SNPs based on fGWAS p-values




AFR
EAS
SAS
Expected RA
activity
adhd
adventure
afb
afs
allergypollen
anorex
asd
asi
asteczrhi
asthma
audit
bipolar
bmi
bpdia
bppulse
bpsys
brca
cad
cannabis
chol
copd
cp
cpd
dep
dpw
ea
eversmoke
extra
famsat
friendsat
hayfever
hdl
height
ibd
insomnia
ldl
menarche
migraine
morning
nearsighted
nebwomen
neuro
nonhdl
open
prca
religatt
risk
scz
selfhealth
smcess
swb
t2d
tryg
$RA_{Expected}$
RA_100
RA_1000
RA_10000

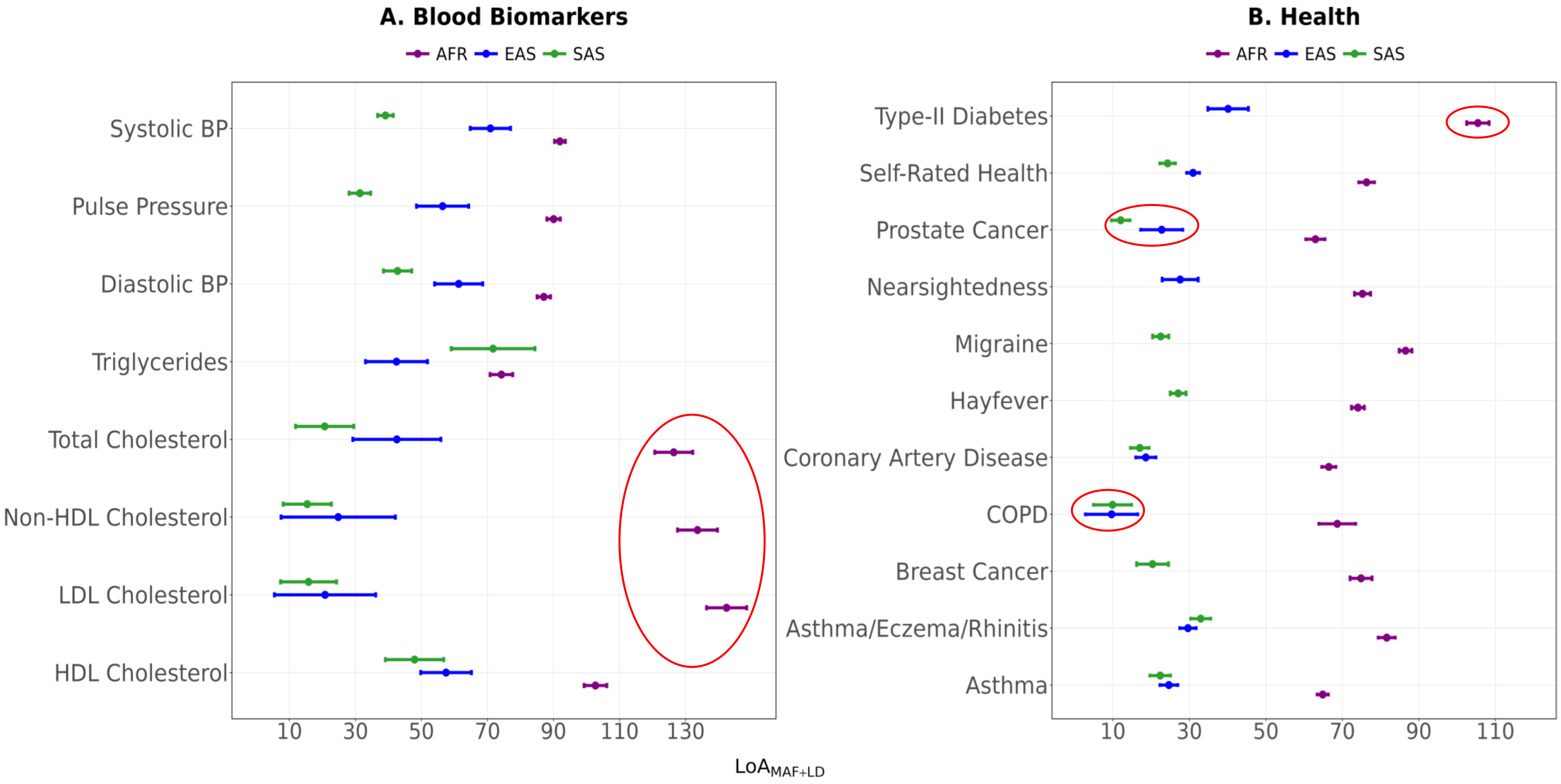
Loss of predictive accuracy attributable to MAF and LD by phenotype category in UKB



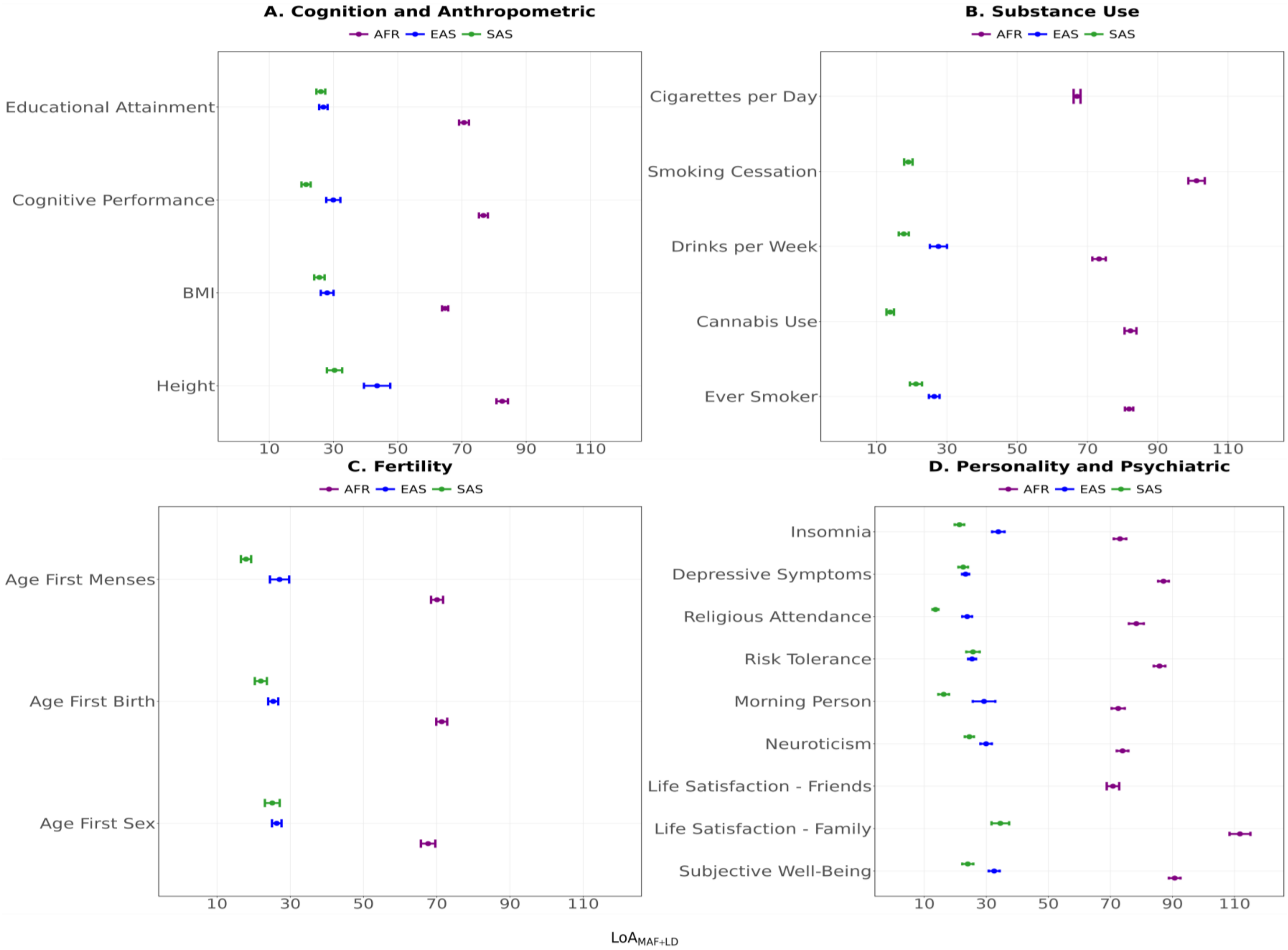
- Mean LoA_{LD+MAF}
  - 83% in AFR
  - 33% in EAS
  - 25% in SAS
- Highest in blood biomarkers in all ancestries
- Lowest in fertility and substance use

$$LoA_{(MAF+LD))} = \left(\frac{1-RA_{Expected}}{1-RA_{Observed}}\right) \times 100$$

Loss of predictive accuracy attributable to MAF and LD by phenotype category in UKB

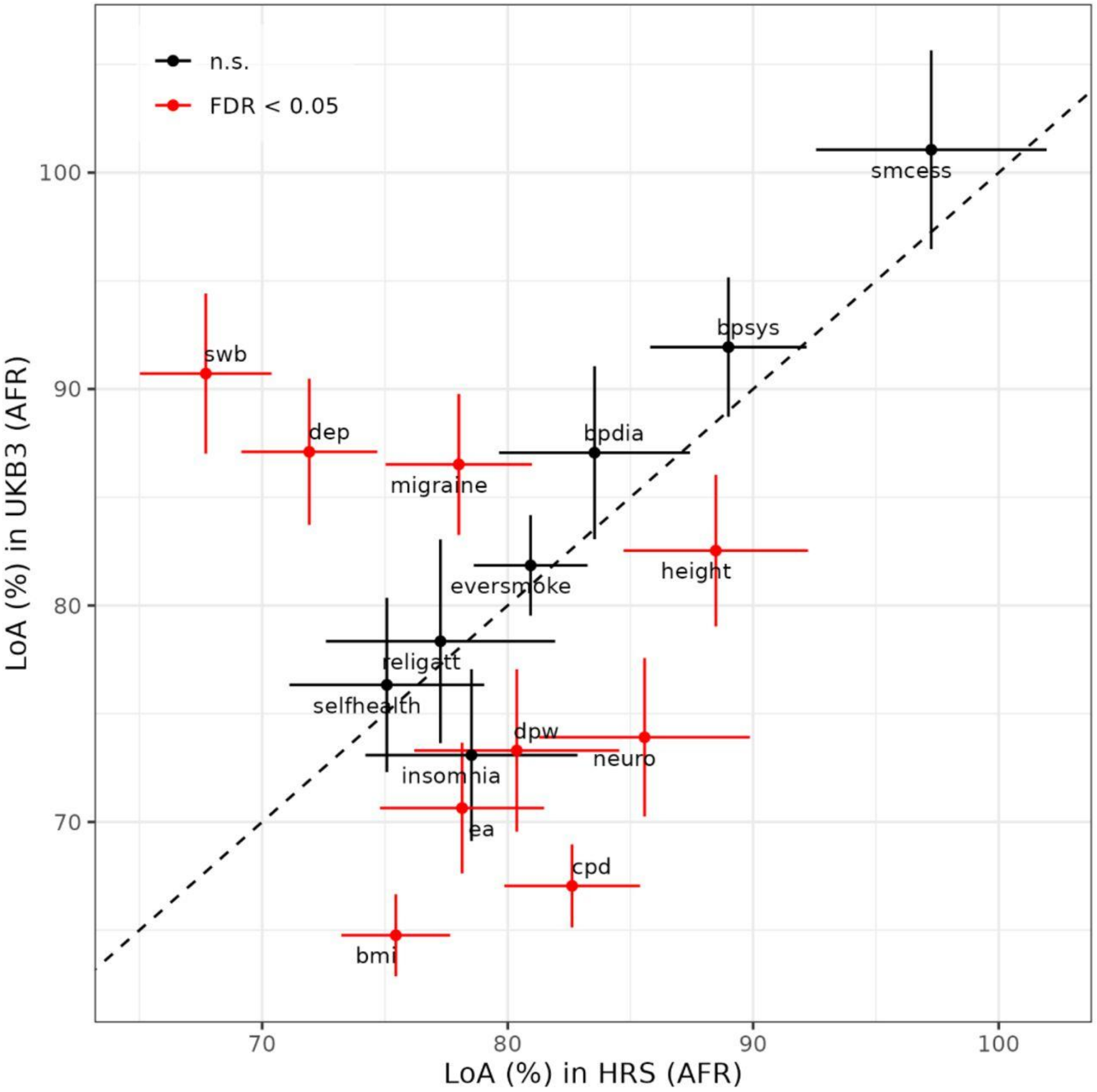


Loss of predictive accuracy attributable to MAF and LD by phenotype category in UKB

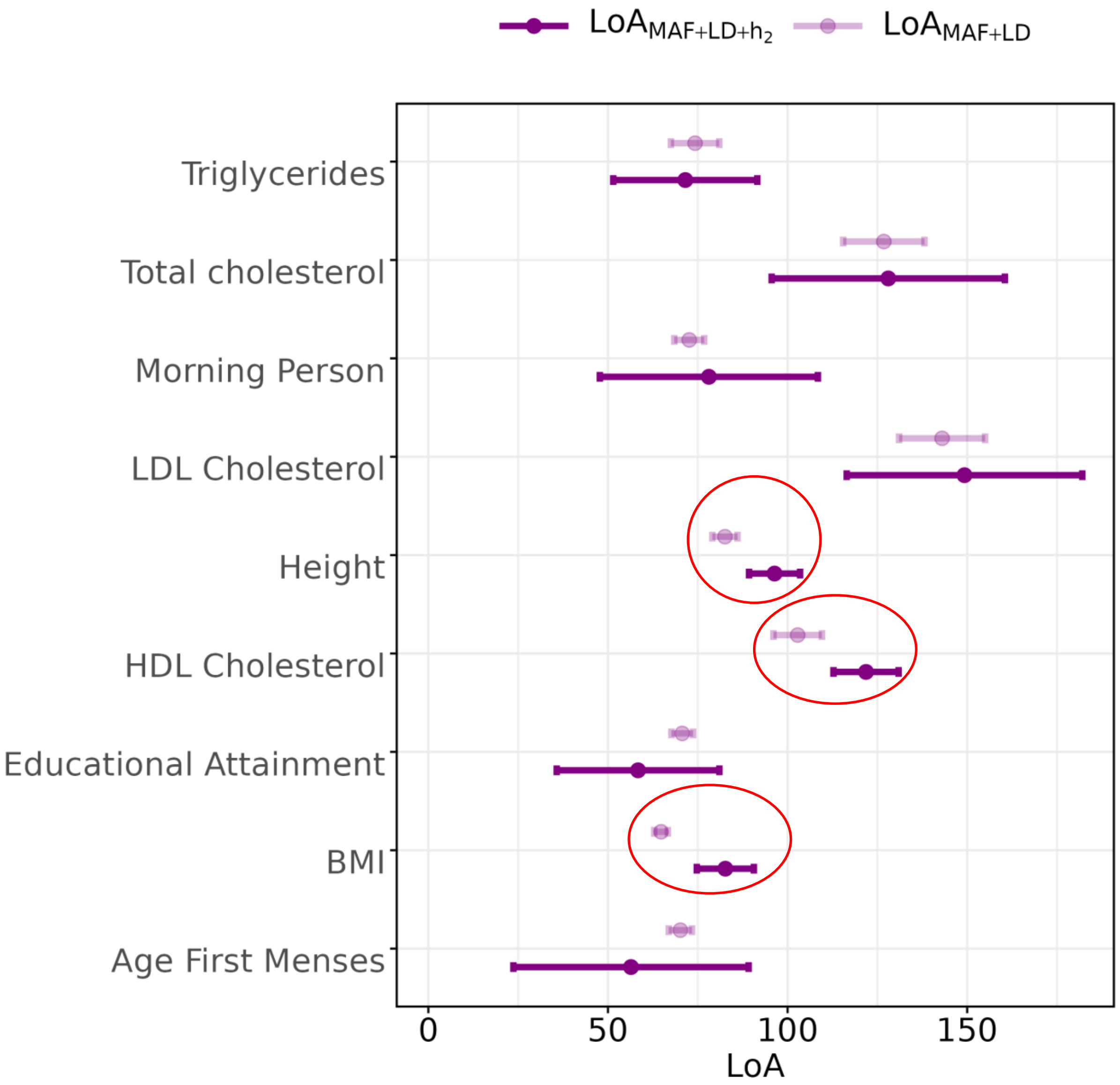


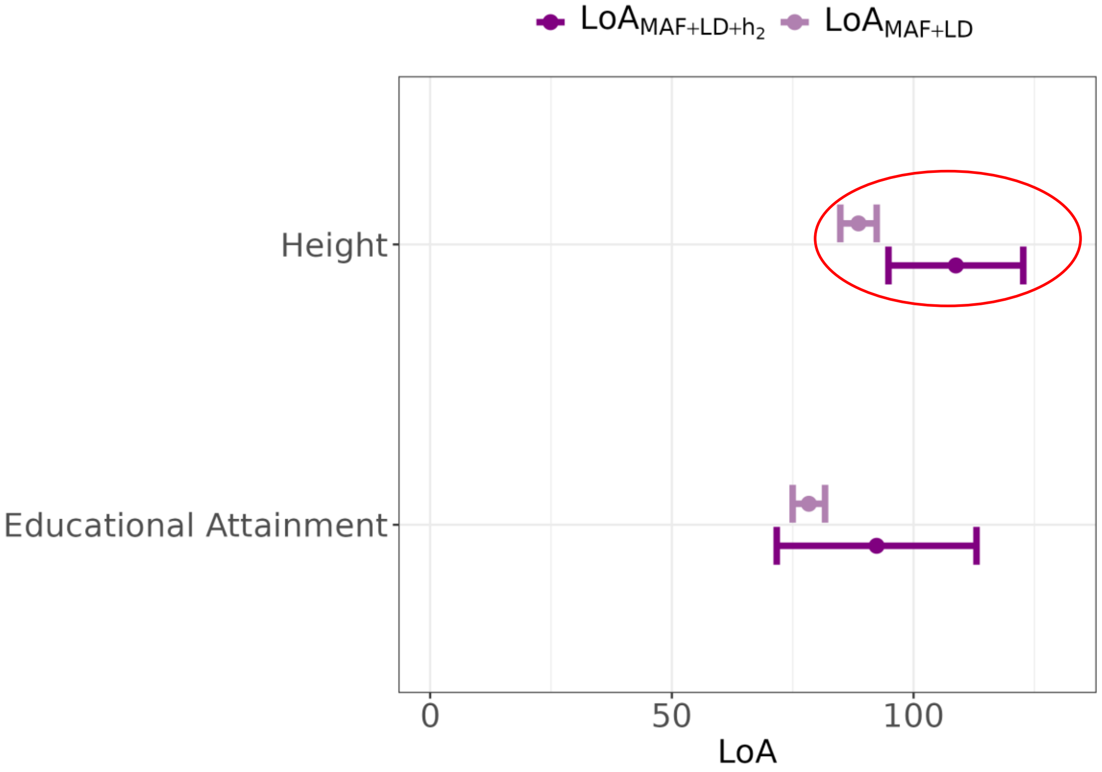
UKB vs HRS, AFR ancestry

- Contribution of MAF and LD to PGI predictive accuracy varies across cohorts
- May reflect
  - differences in environmental exposures
  - differences in sample characteristics
  - differences in phenotype definitions

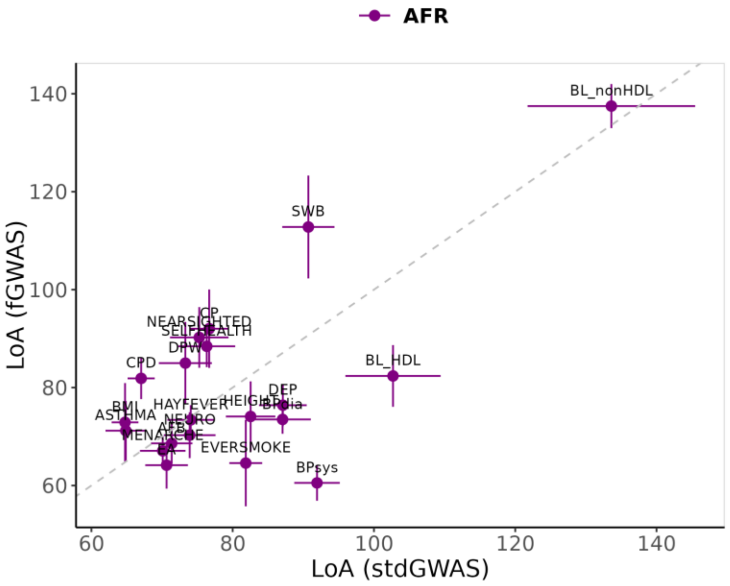


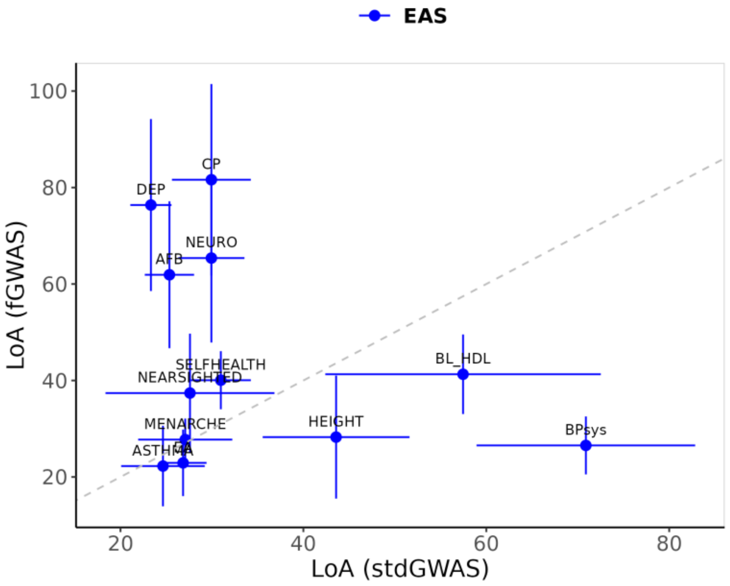
Introducing heritability differences, UKB vs HRS, AFR ancestry

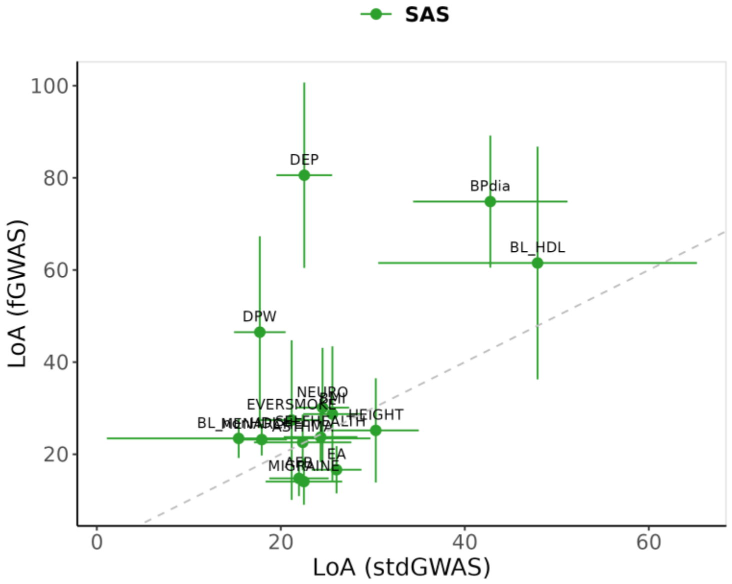




LoA$_{MAF+LD}$ based on fGWAS Vs stdGWAS SNP weights







Meta-analyzed fGWAS SNP weights: Tan et al. (2024). Family-GWAS reveals effects of environment and mating on genetic associations. medRxiv.

Summary

- We confirm substantial reductions in the predictive accuracy of PGIs for non-European ancestries:
  - $RA_{EUR \rightarrow AFR} < RA_{EUR \rightarrow EAS} < RA_{EUR \rightarrow SAS}$
- The loss is not uniform across trait categories
  - Portability is substantially **lower for behavioral and social traits** compared to more biologically proximal phenotypes
- Core findings are broadly consistent with Wang *et al.*, showing that differences in LD and MAF account for a substantial portion of accuracy loss:
  - $LoA^{AFR}_{LD+MAF} = 83\% > LoA^{EAS}_{LD+MAF} = 34\% > LoA^{SAS}_{LD+MAF} = 25\%$
- fGWAS-based PGIs can improve portability for some traits (e.g. BMI in AFR), suggesting that some of the portability gap may be attributable to population-specific biases present in standard PGIs
- fGWAS findings except for BMI in AFR align with Zhang and Conley, who reported no improvement in cross-ancestry prediction accuracy in HRS and Add Health
- We find differences in LoA^{AFR}_{LD+MAF} residing in different contexts, i.e. UK versus US

Source: Wang et al., (2020)

Limitations & Future research

- Only common SNPs (MAF>0.01) used for PGI computation – may not generalize to rare SNPs which are often population specific
- Incorporating rare variants and improving imputation strategies for diverse populations could alter the cross-ancestry portability of PGIs.
- The use of genotype reference panels like the 1000 Genomes which are biased toward EUR ancestry may introduce inaccuracies in estimating population parameters such as MAF and LD.
- While fGWAS-based PGIs showed promise for improving portability for certain traits, their smaller effective discovery sample sizes and stringent QC filters limit power and genome-wide SNP coverage.
- Expanding family-based GWAS datasets in scale and coverage could enhance the robustness of fGWAS-based approaches

This study is supported by funding from:

- NIA/NIH through grants R24-AG065184, R01-AG042568, R01-AG083379, R00-AG062787, R01-AG081518
- The Open Philanthropy Project (010623-00001)
- Amsterdam UMC fellowship
- Swedish Research Council (2023-01343 and 2024-06499)

Robel Alemu

UCLA

Broad Institute of
MIT and Harvard

Patrick Turley

USC

Daniel Benjamin

UCLA

NBER

Alexander Young

UCLA

Thank you!

ESSGN

GENETICS
NETWORK
AMSTERDAM

Social
Science
Genetic
Association
Consortium

Amsterdam UMC