



wellcome
connecting
science

connecting people with science



MAKERERE UNIVERSITY



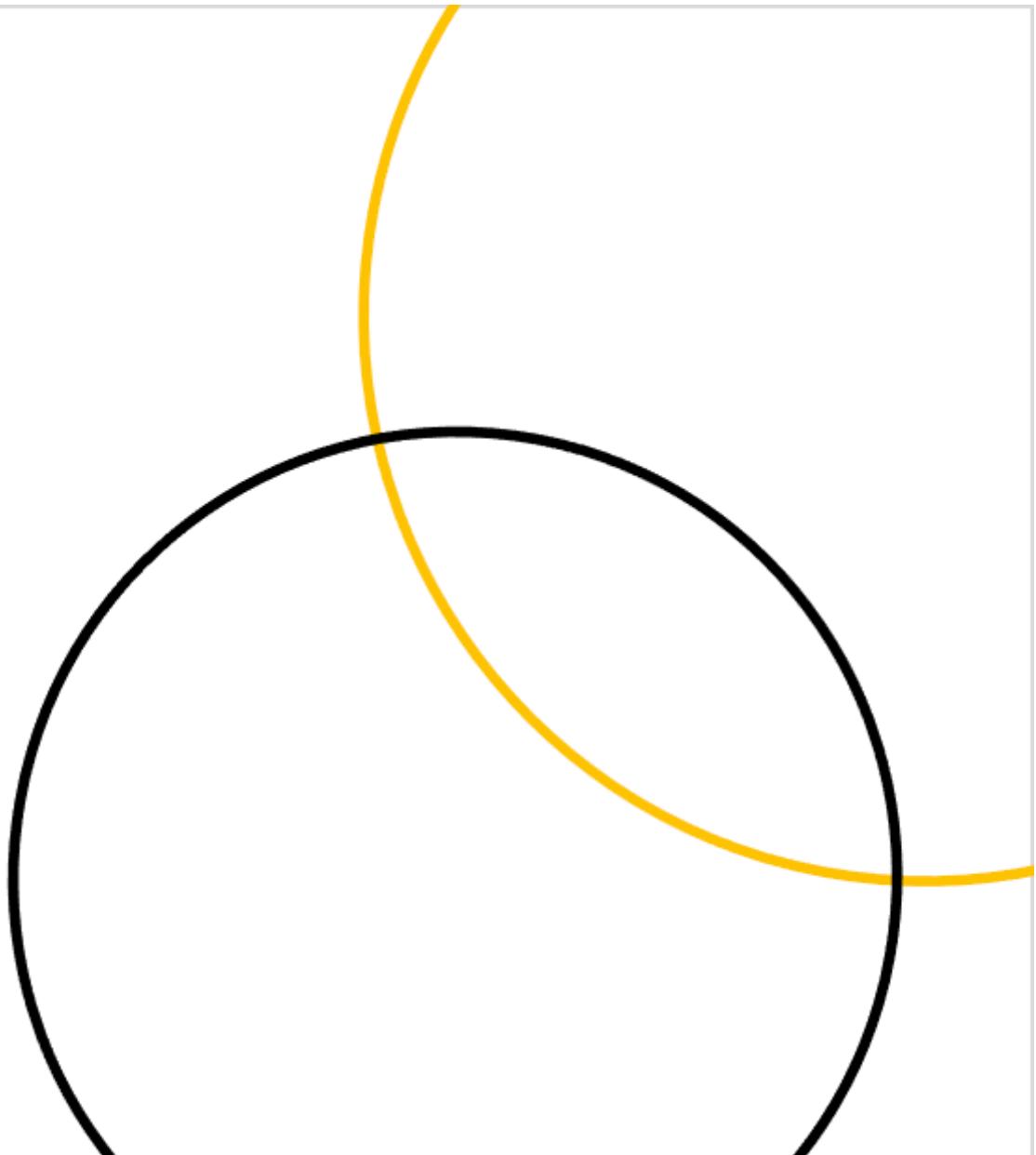
IDI
INFECTIOUS
DISEASES
INSTITUTE



AFRICAN
CENTER
OF EXCELLENCE
IN BIOINFORMATICS
& DATA INTENSIVE SCIENCES

Dr Conrad Iyegbe
Instructor
Introduction to BridgePRS

June 2025



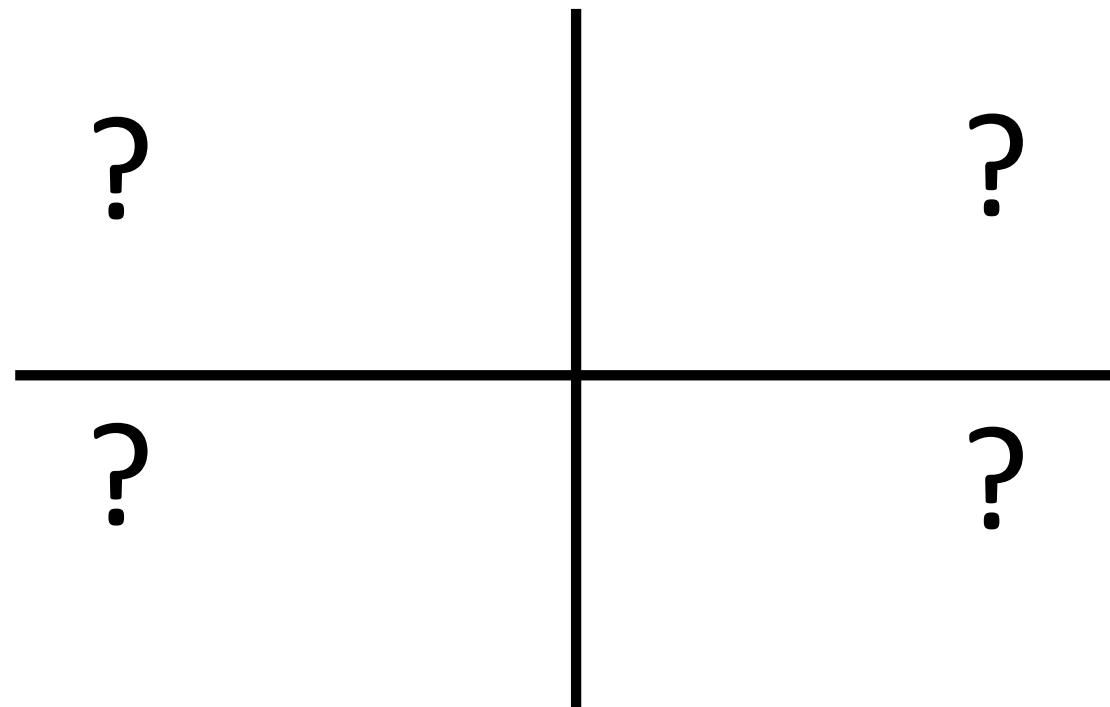
Talk plan

Introduction to BridgePRS

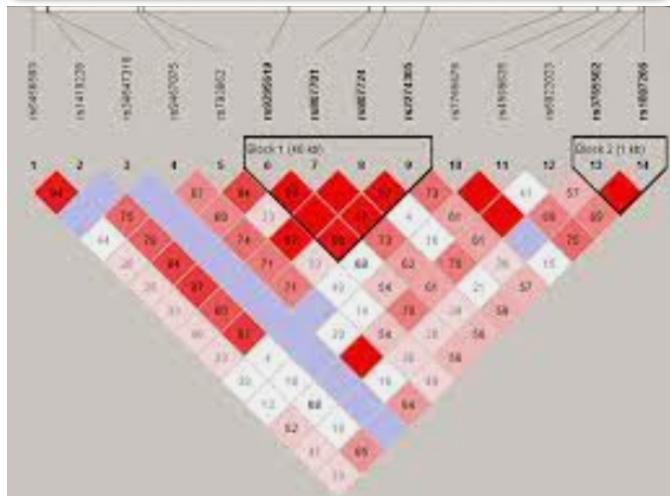
- Quick recap of the portability problem and its causes
- Explanation of F_{ST}
- Methodological developments in cross-ancestry genetic research
- BridgePRS: Rationale and differences in approach to CSX

What are the determinants of poor PRS portability?

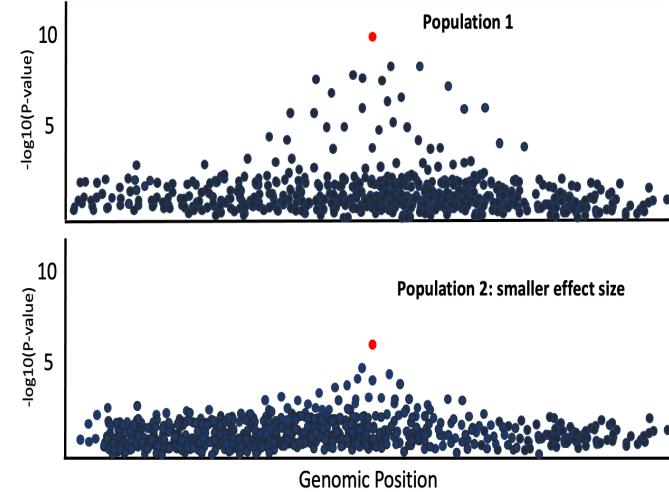
What are the determinants of poor PRS portability?



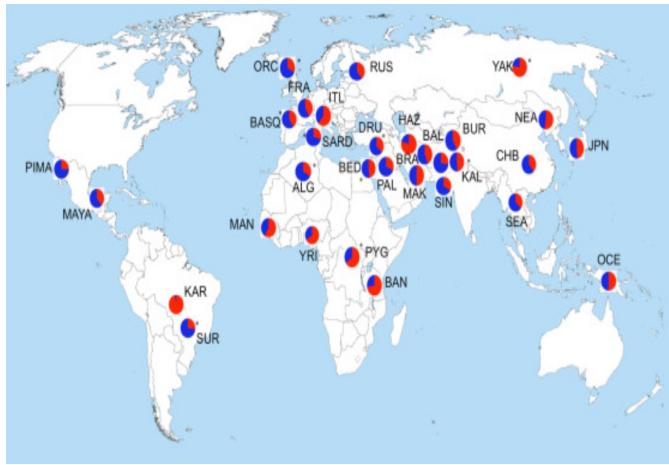
1. Linkage Disequilibrium



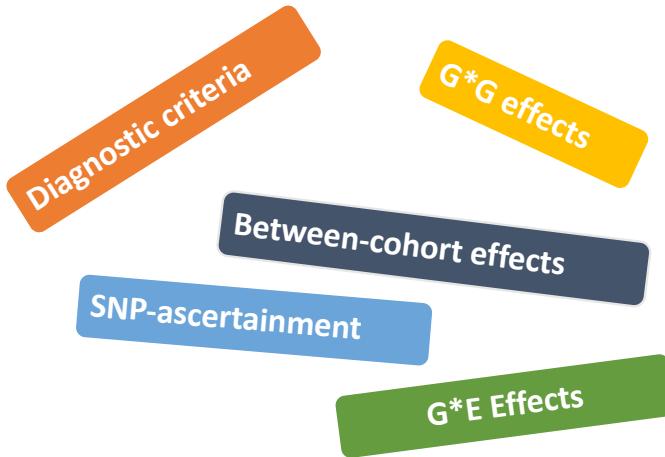
2. Effect Sizes



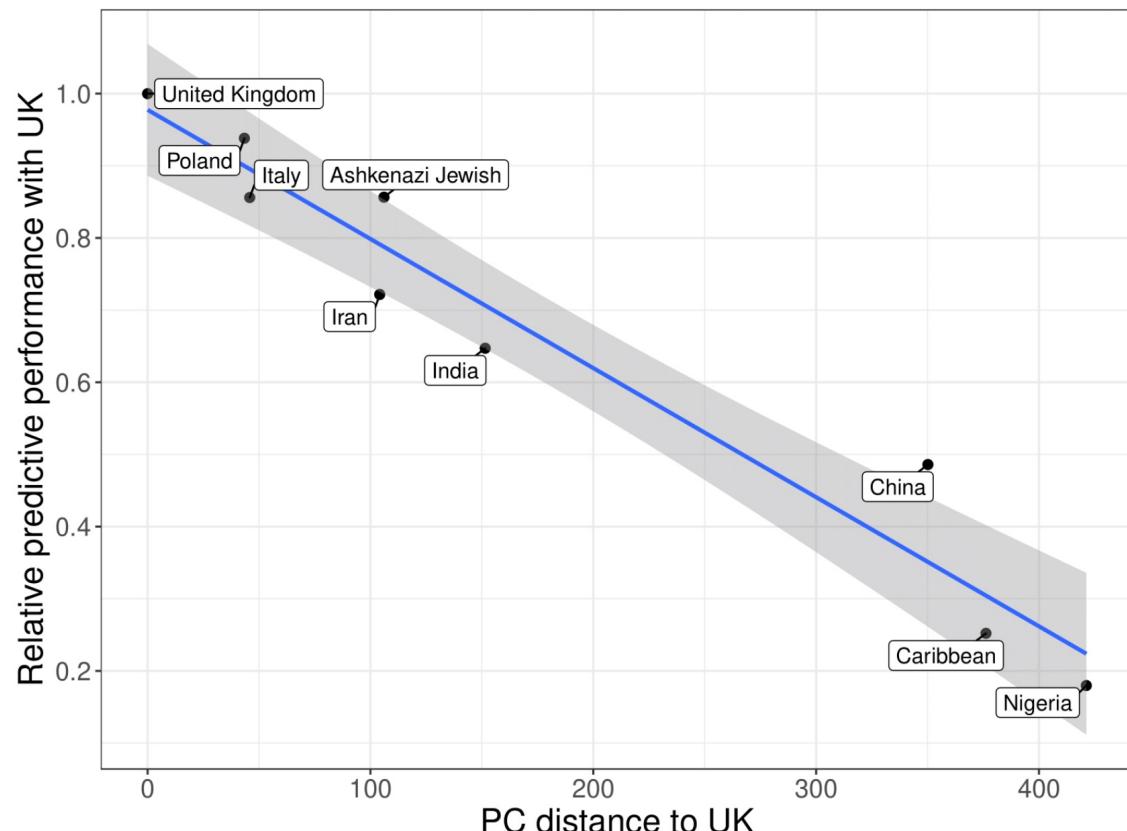
3. Allele Frequencies



4. Other Causes

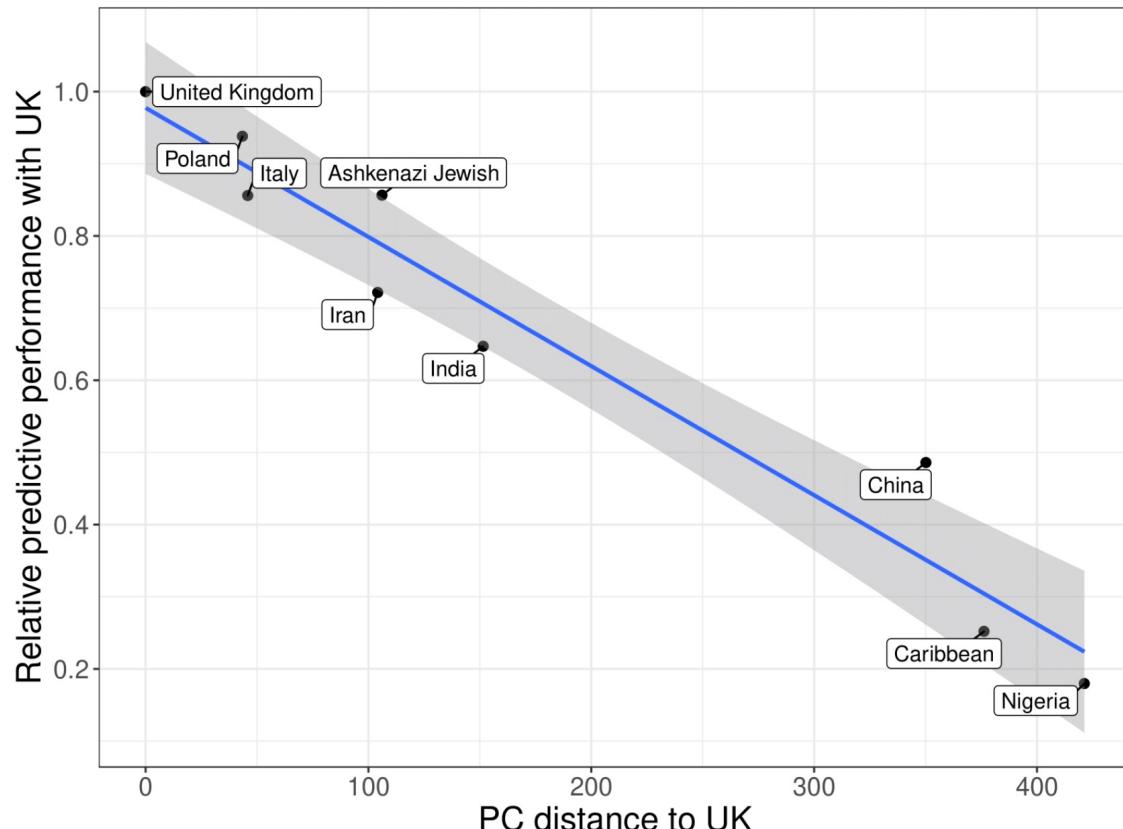


What is this plot telling us?



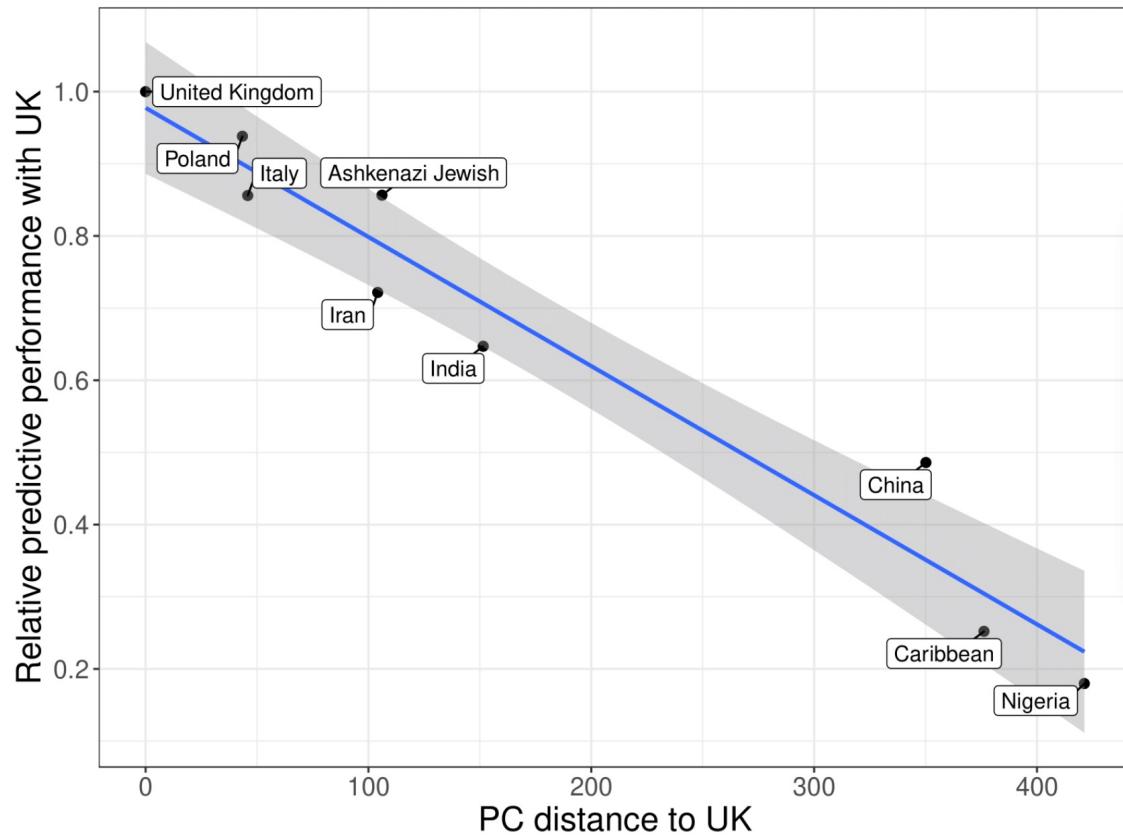
Privé et al, 2022

The PRS portability gap begins in Europe (and worsens with further genetic distance)

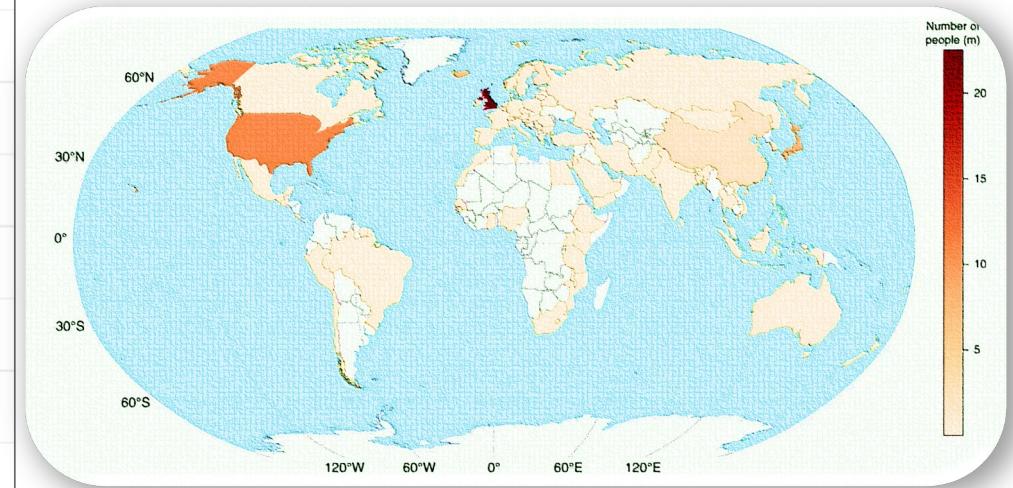


Privé et al, 2022

The PRS portability gap starts in Europe (and ends in Africa)

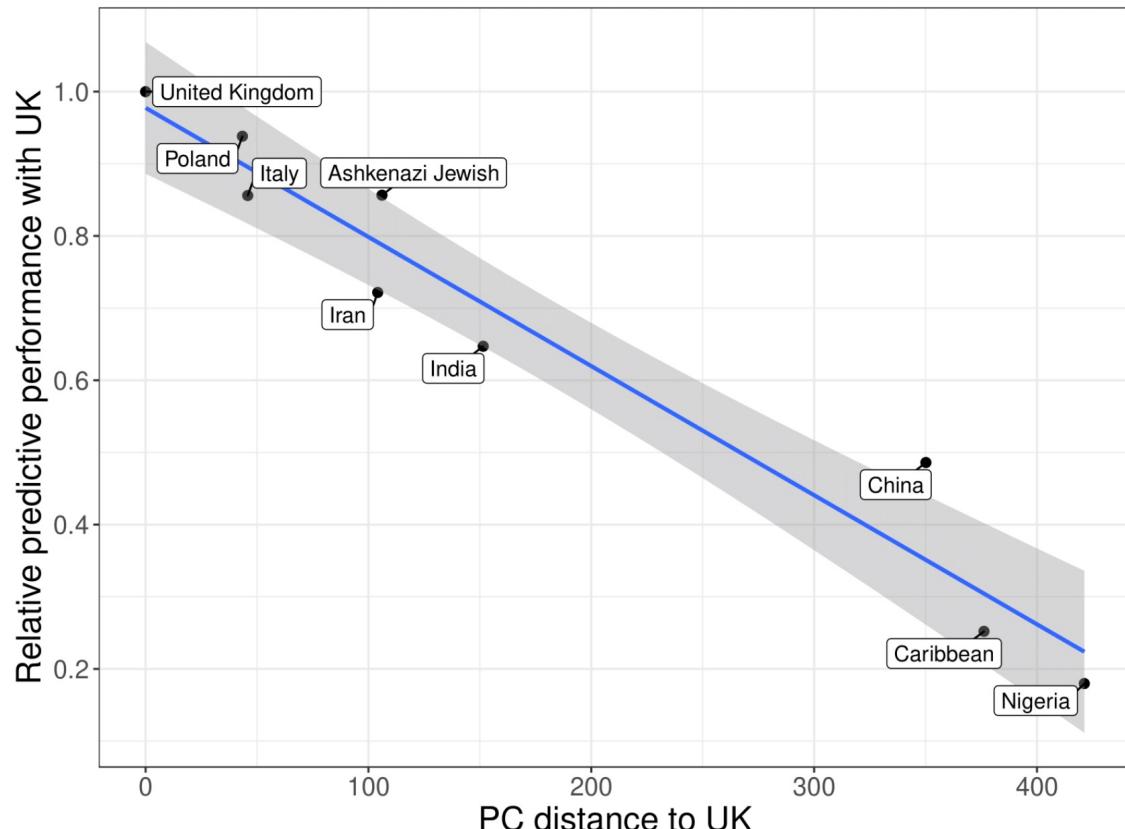


Privé et al, 2022

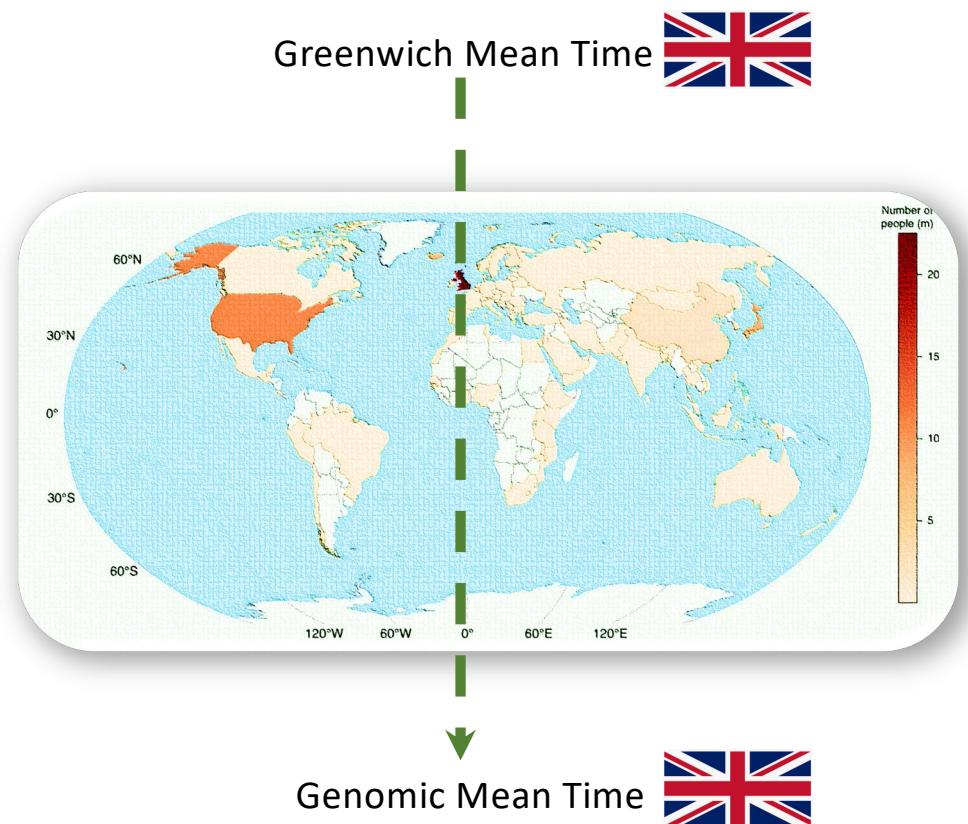


Mills & Rahal, 2019

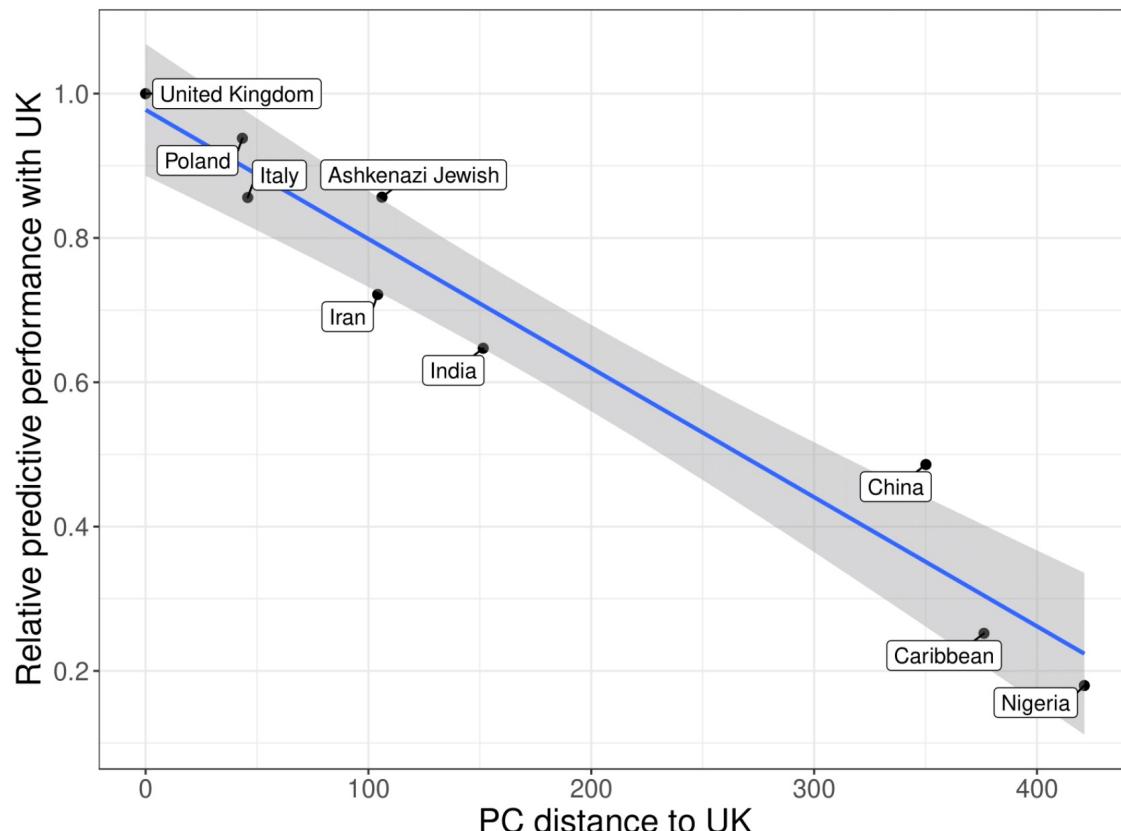
The PRS portability gap starts in Europe (and ends in Africa)



Privé et al, 2022



The Genetic Fixation Index F_{ST}



Privé et al, 2022

Fixation Index: A formal measure of genetic differentiation between populations

$Fst = 0$: No genetic differentiation between populations. Populations are genetically identical.

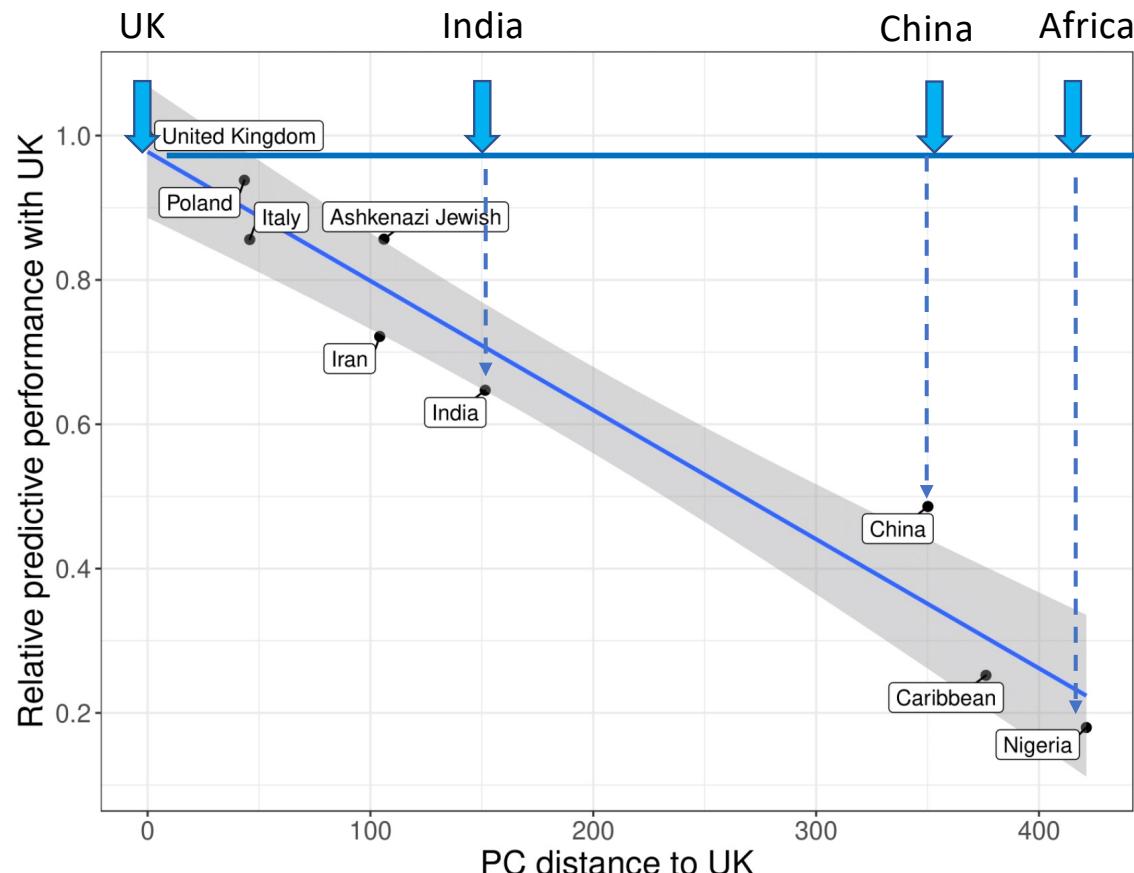
$Fst = 1$: Complete genetic differentiation. The populations are entirely distinct. No shared genetic variation

Fixation Index in 1000 Genomes sample

POP1	POP2	Fst
AFR	AMR	0.10
AFR	EAS	0.15
AFR	EUR	0.12
AFR	SAS	0.11
AMR	EAS	0.06
AMR	EUR	0.02
AMR	SAS	0.03
EAS	EUR	0.09
EAS	SAS	0.06
EUR	SAS	0.03

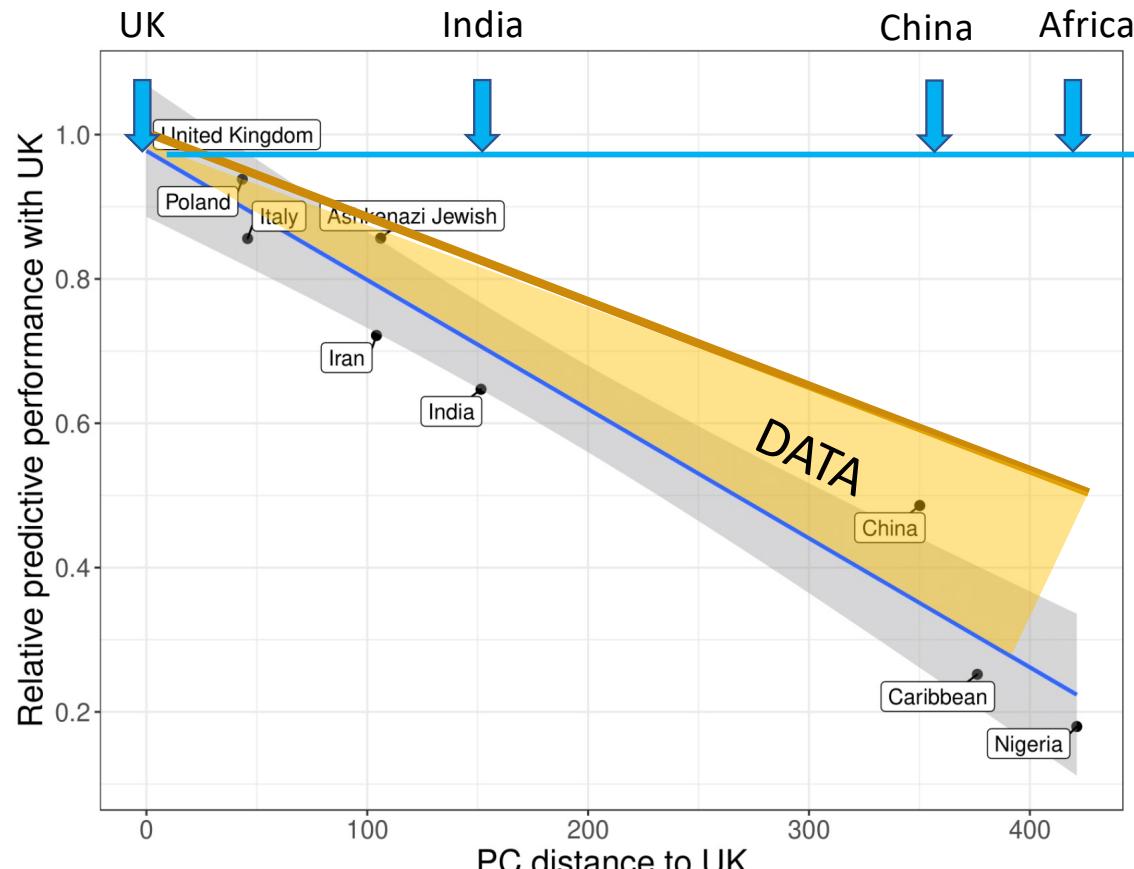
AFR: African
EAS: East Asian
AMR: Native American
SAS: South Asian
EUR: European

Polygenic Risk Scores and the Portability Problem



Privé et al, 2022

Closing the Gap with Diverse Data

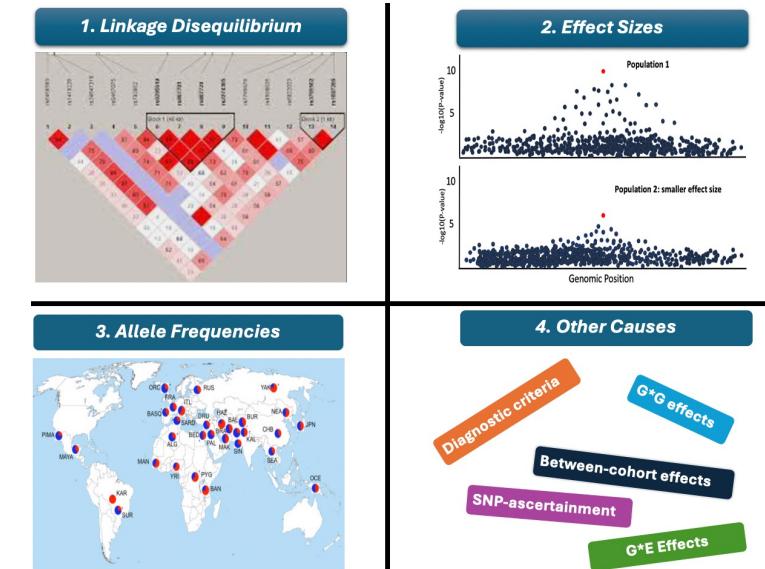
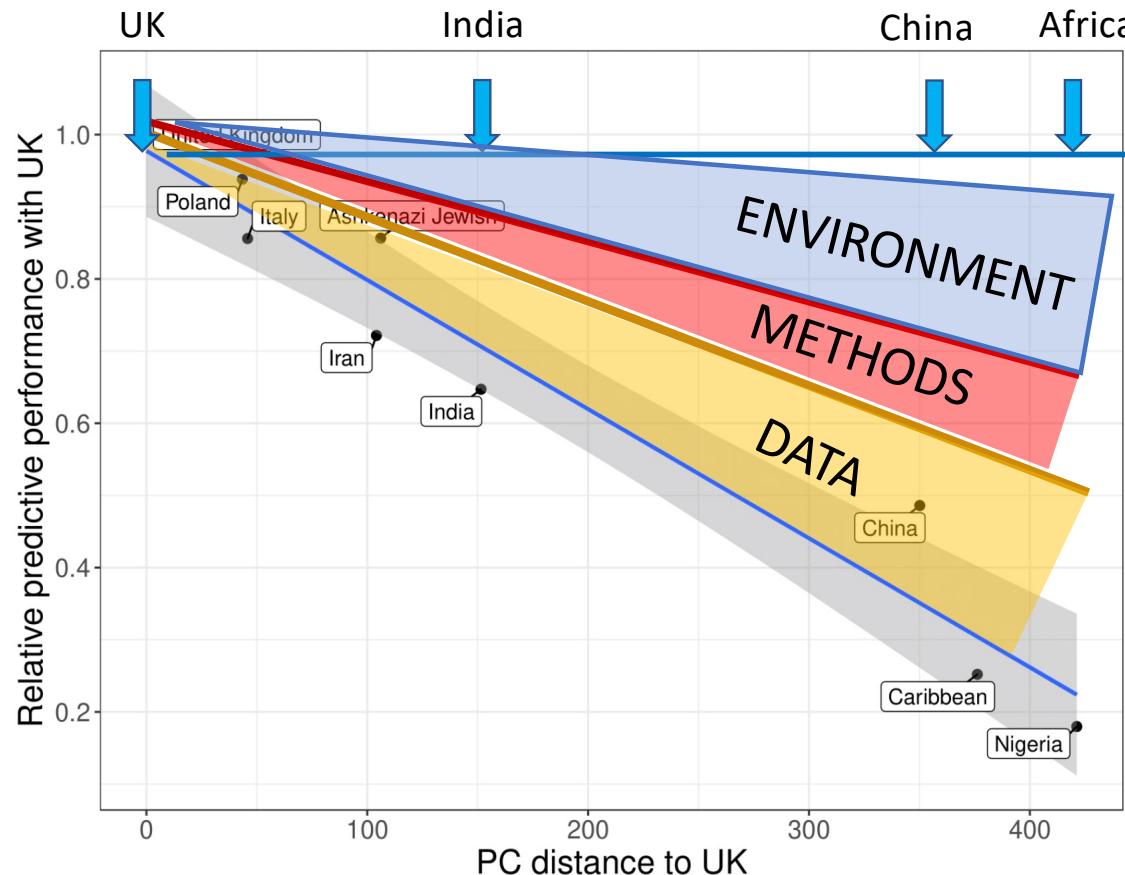


The inclusion of more globally diverse data will:

- Capture more population-specific causal variants.
- Refinement of LD structure across ancestries
- Recalibration of effect size weights in global summary statistics

Privé et al, 2022

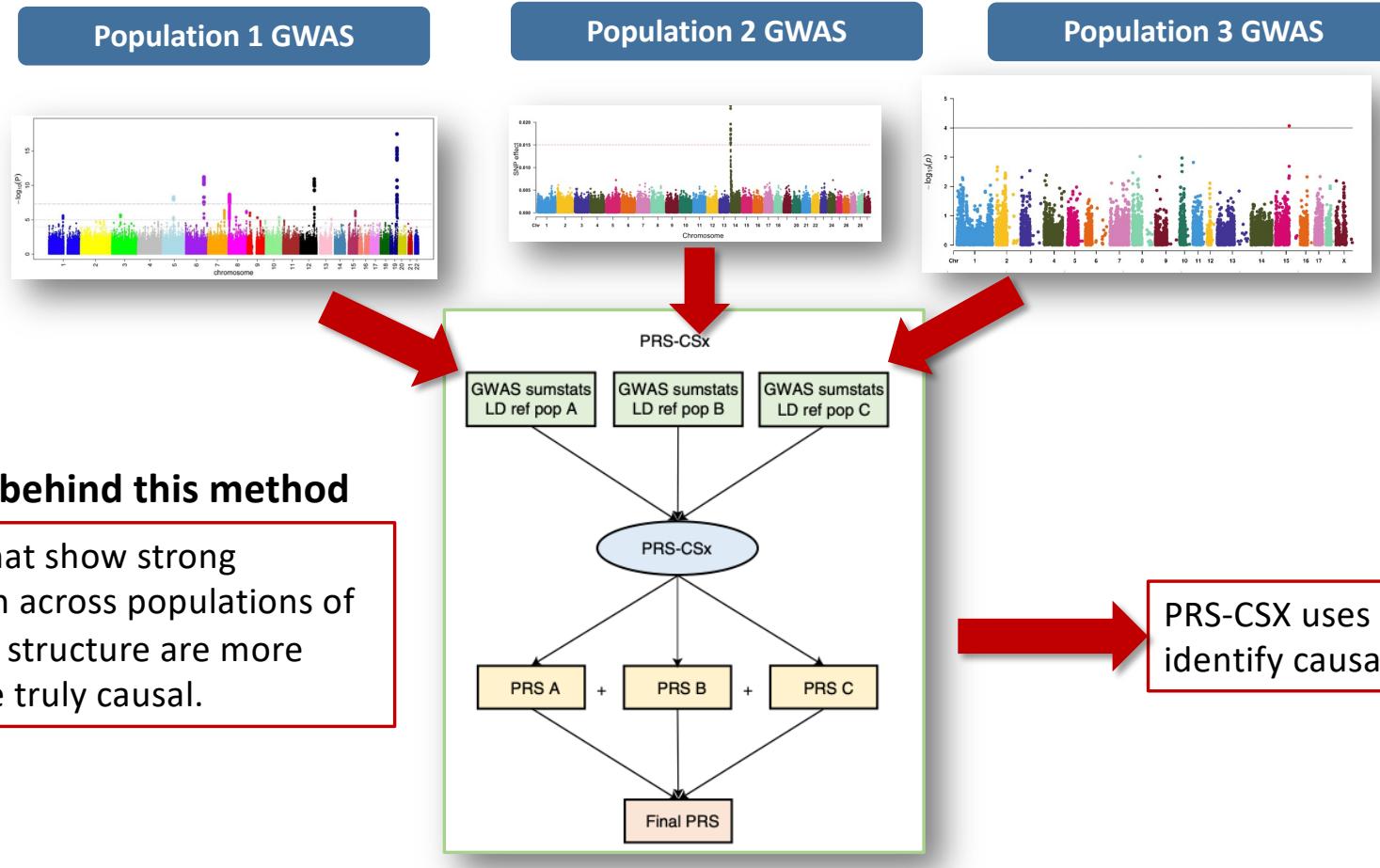
Further gains possible using dedicated methods

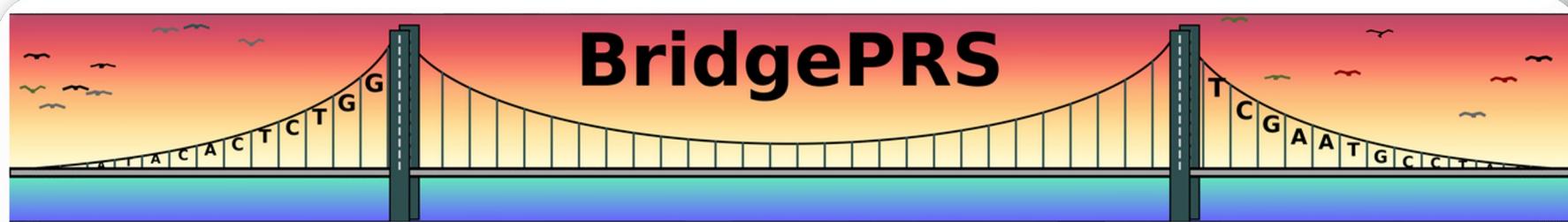


Statistical modelling can overcome practical challenges of bringing together globally diverse genetic datasets.

Privé et al, 2022

PRS-CSx Bayesian polygenic modeling and prediction framework





nature genetics



Technical Report

<https://doi.org/10.1038/s41588-023-01583-9>

BridgePRS leverages shared genetic effects across ancestries to increase polygenic risk score portability

Received: 18 January 2022

Clive J. Hoggart , Shing Wan Choi , Judit García-González ,

Accepted: 20 October 2023

Tade Souaiaia

Published online: 20 December 2023

Michael Preuss & Paul F. O'Reilly

Check for updates

Here we present BridgePRS, a novel Bayesian polygenic risk score (PRS) method that leverages shared genetic effects across ancestries to increase

https://www.bridgeprs.net/guide_args/

uncertainty increases: with lower trait heritability, higher polygenicity and greater between-population genetic diversity; and when causal variants are not present in the data. In real data, BridgePRS has a 61% larger average R^2 than PRS-CSx in out-of-cohort prediction of African ancestry samples in BioMe ($P = 6 \times 10^{-5}$). BridgePRS is a computationally efficient, user-friendly and powerful approach for PRS analyses in non-European ancestries.

BridgePRS Method Overview

Two-stage modeling: (i) Builds an initial PRS from a large discovery GWAS (ii) Updates variant weights using the target-population GWAS.

Inclusive variant weighting: Retains all SNPs in each genomic region, applying shrinkage rather than hard pruning.

Cross-population “bridge”: Leverages shared genetic effects across ancestries to inform weight adjustments in the target group.

Capturing unique signals: Optionally incorporates target-only associations for loci missing in the discovery study.

Key Features

Improved portability: Boosts PRS accuracy in underrepresented ancestries by borrowing strength from larger GWAS.

Robust to LD differences: Averages evidence across correlated SNPs, reducing bias from divergent linkage patterns.

Adaptable to complexity: Outperforms single-population and hard-selection methods when traits are highly polygenic.

Designed based on African contexts

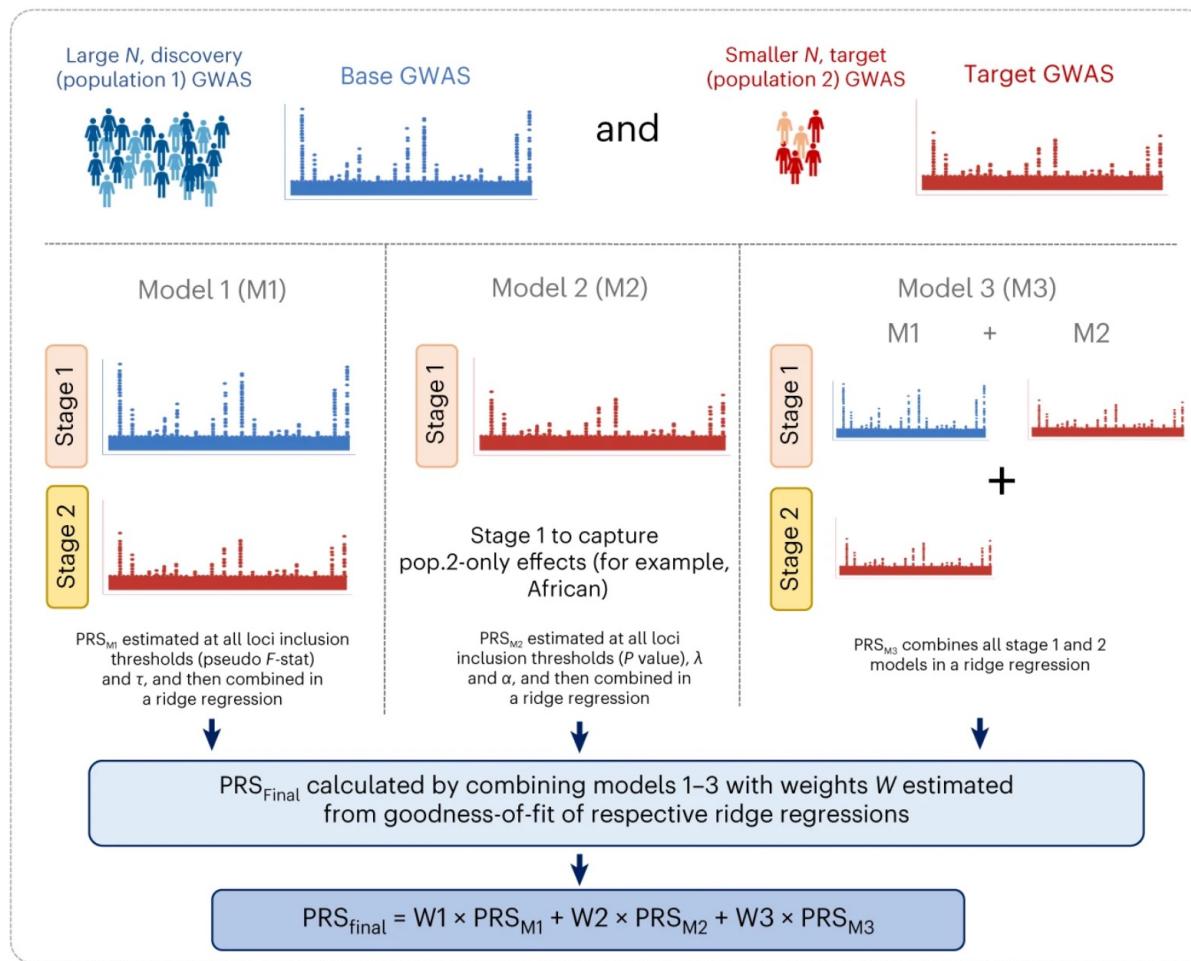
BridgePRS workflow

Stage 1
Shrinkage of GWAS betas
(Bayesian Ridge Regression)

Stage 2
PRS SNP effects treated as priors, updated in a Bayesian framework using effect sizes from smaller population.



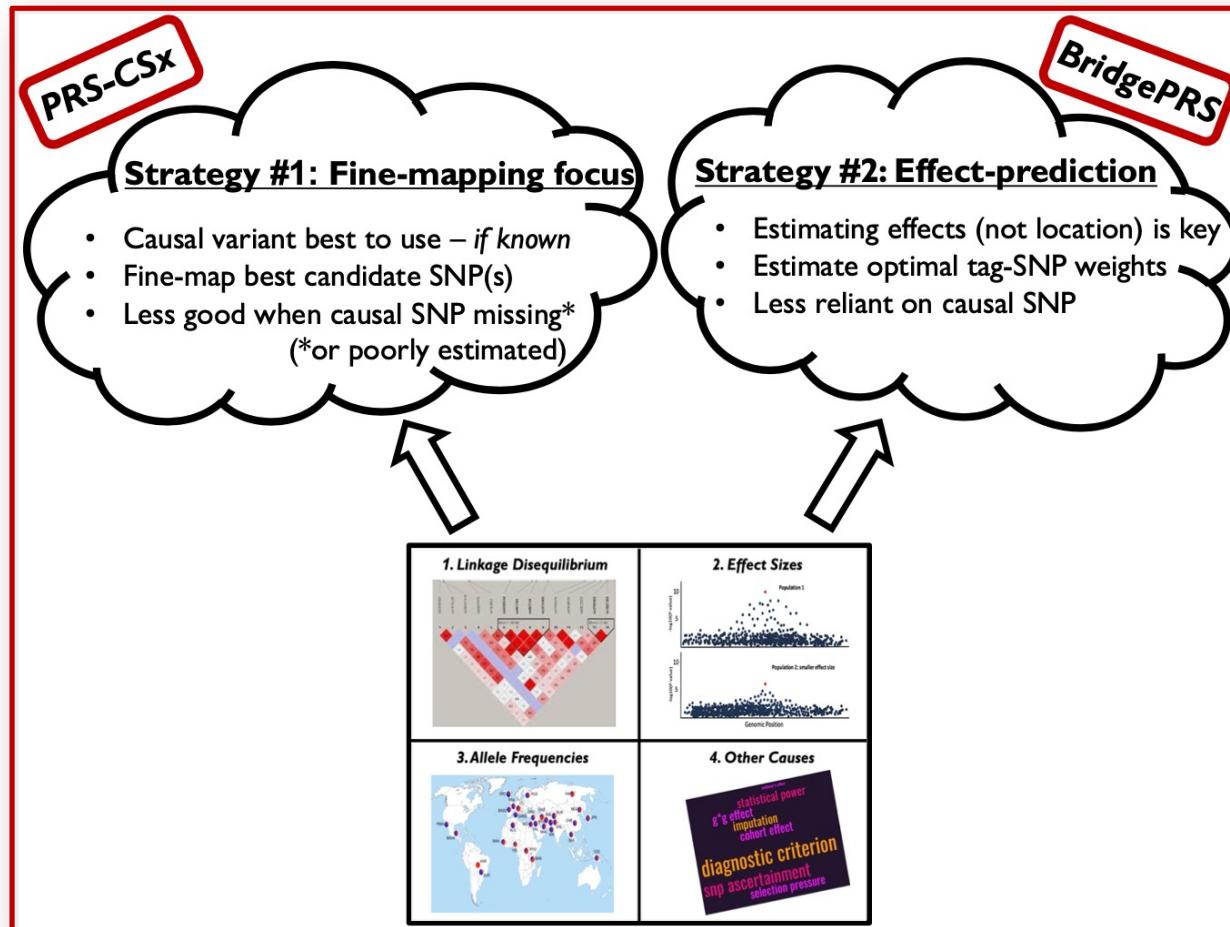
Bridging of PRS between populations



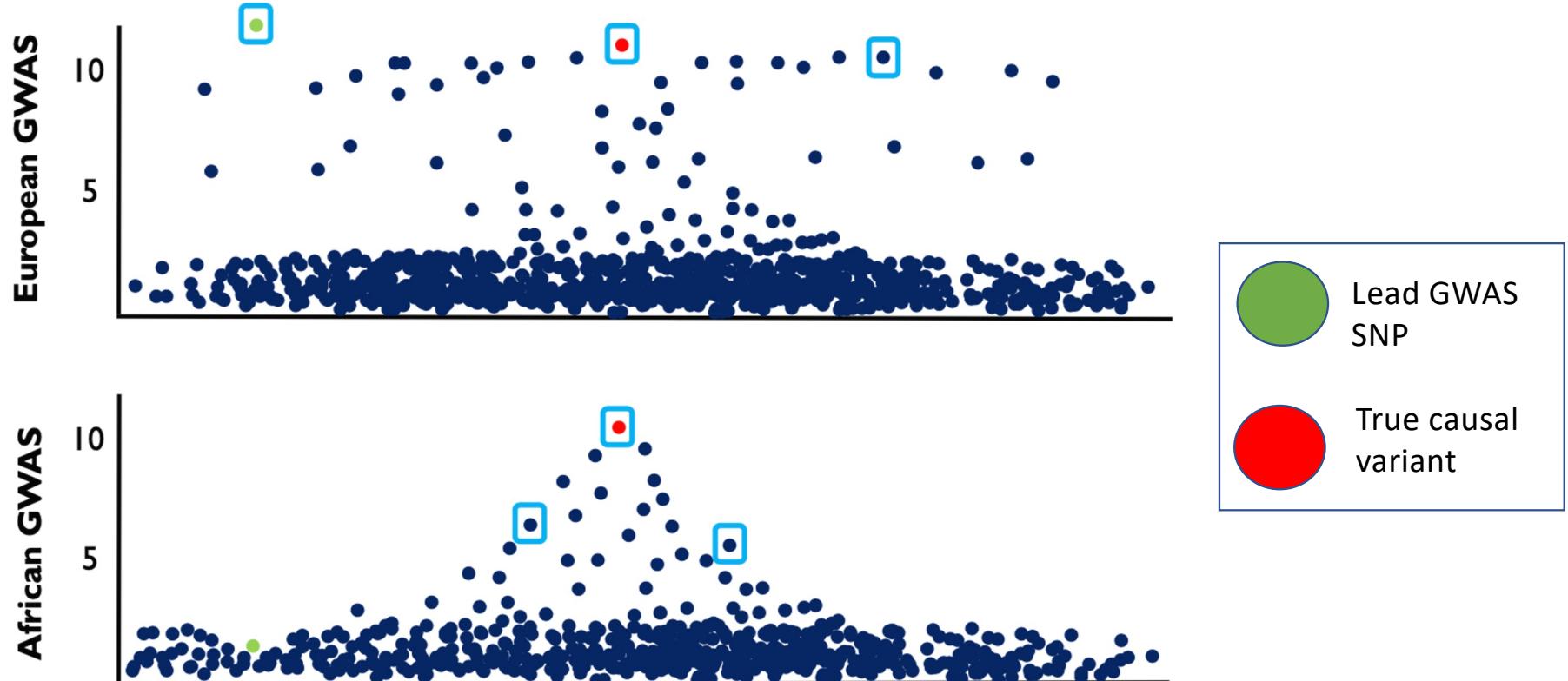
How BridgePRS and PRS-CSx differ

Feature / Aspect	BridgePRS	PRS-CSx
Modeling Approach	Two-stage Bayesian update: build from discovery GWAS, then adjust with target GWAS.	Joint Bayesian model: couples effect sizes across all ancestries simultaneously.
Variant Inclusion & Shrinkage	Inclusive weighting: *retains all SNPs per locus, applies moderate shrinkage.	Harder shrinkage: zeroes out many SNPs, focusing on a minimal set of top signals.
Handling LD Differences	**Averages evidence across correlated SNPs to accommodate different LD patterns.	Uses ancestry-specific LD priors to fine-map and pinpoint causal variants.
Effect-size Estimation	Sequentially updates effect sizes with target data, blending large- and small-study signals.	Estimates all effect sizes jointly, borrowing strength but enforcing sparsity.
Best-case Performance	***Excels in highly polygenic traits, low heritability, or when causal SNPs are missing.	Excels when few large-effect variants are known and present in the data.
Computational & Practical Notes	Streamlined pipeline, minimal tuning, efficient even with many variants.	More complex fine-mapping step; may require careful parameter settings.

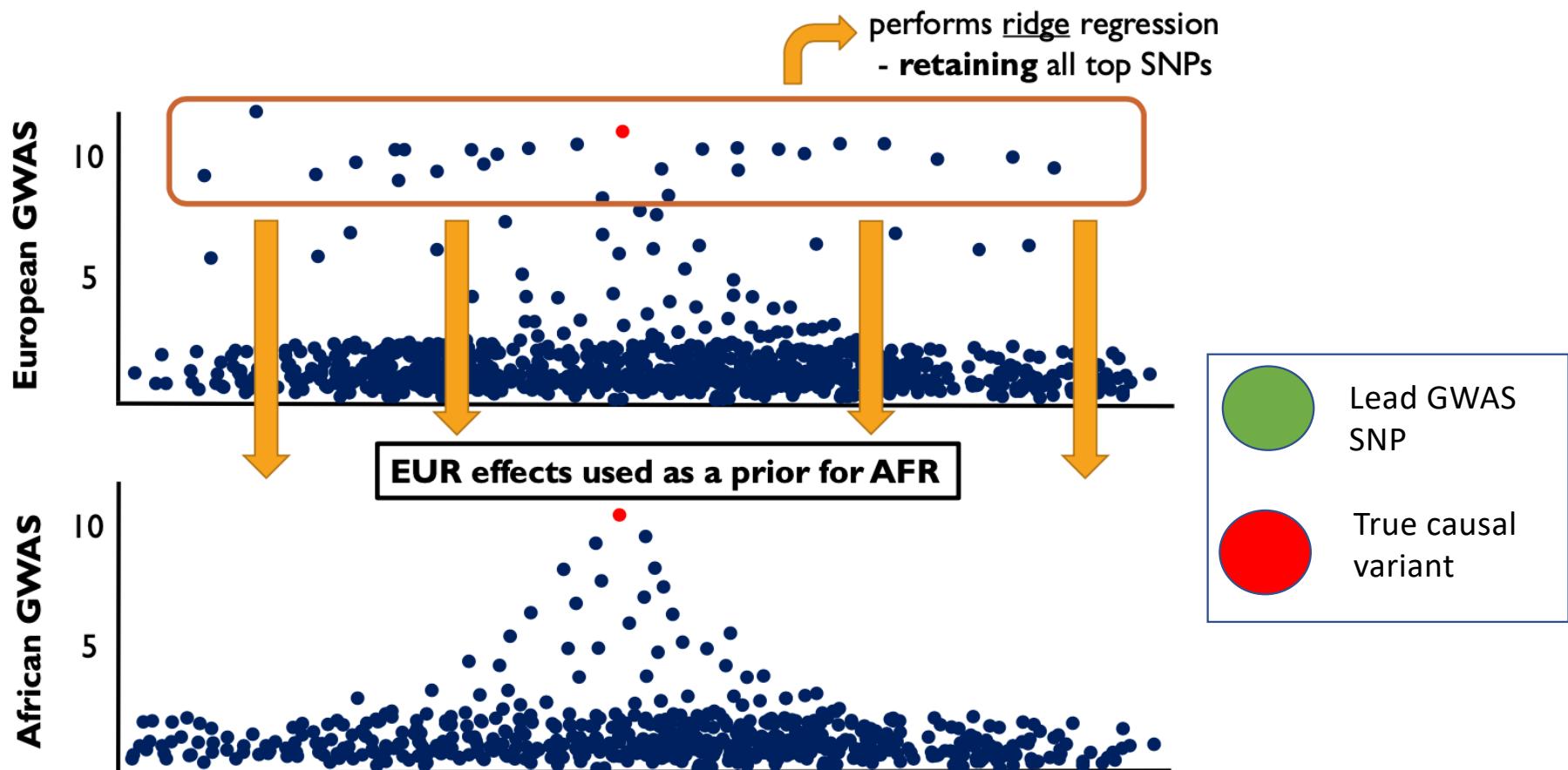
Comparison of PRS-CSx and BridgePRS



PRS-CSx tries to ‘pick a winner’



BridgePRS: Stage 2



BridgePRS: Software Features

Sub-routines used by Bridge

Subprogram	Input	Subcommands	Output
prs-single	Target Pop Data	run,clump,beta,predict,quantify	prs-result
build-model	Model Pop Data	**run**clump,beta,predict,prior	model-params
prs-port	Target Pop + Model Result	run,predict,quantify	prs-port-result
prs-prior	Target Pop Data + Model Result	run,clump,beta,test,predict	prs-prior-result
analyze	PRS Result Files	**run**result,combine	single-plot,weighted prs-result,weighted prs plot

Input data for configuration file command

Input Type	Command Line Flag	Base Population	Target Population
Population Information	--pop	Required	Required
LD-Reference Pop.	--ld_pop	Preloaded with 1000 genomes data (but need to specify which)	Preloaded with 1000 genomes data (but need to specify which)
Summary statistics	--sumstats_prefix	Required	Required
Genotypes	--genotype_prefix	-	Required
Phenotype File	--phenotype_file	-	Required
Validation File	--validation_file	-	Bridge derives from phenotype file
QC SNP list	--snp_file	Uses all SNPs unless specified otherwise by user	Uses all SNPs unless specified otherwise by user

Summary Stats: Arguments

Default Headers	CHR	ID	REF	A1	A1_FREQ	OBS_CT	BETA	SE	T_STAT	P
Argument		--ssf-snpid	--ssf-ref	--ssf-alt	--ssf-maf	--sdf-n	--ssf-beta	--ssf-se		--ssf-p
Data	1	rs12184325	T	G	0.0257573	4853	0.820864	0.413692	1.98424	0.0472871
Data	1	rs4970382	C	A	0.483495	4847	0.0011142	0.128347	0.00868116	0.993074
Data	1	rs2710890	G	G	0.424387	4814	0.108094	0.132225	0.817497	0.413687

Input data for configuration file command

Input Type	Command Line Flag	Base Population	Target Population
Population Information	--pop	Required	Required
LD-Reference Pop.	--ld_pop	1000 genomes data comes preloaded (the user just needs to specify which)	1000 genomes data comes preloaded (the user just needs to specify which)
Summary statistics	--sumstats_prefix	Required	Required
Genotypes	--genotype_prefix	-	Required
Phenotype File	--phenotype_file	-	Required
Validation File	--validation_file	-	Bridge derives from phenotype file
QC SNP list	--snp_file	Uses all SNPs unless specified otherwise by user	Uses all SNPs unless specified otherwise by user

Phenotype file

FID	IID	y	y.binary	PC1	PC2
afr1_1	afr2_1	24.4	1	0.53	0.950
afr1_2	afr2_2	4.10	0	0.59	0.450
afr1_3	afr2_3	37.2	1	0.73	-0.13
afr1_4	afr2_4	5.40	0	0.44	-0.55

Phenotype can be provided to BridgePRS using **--phenotype_files**

- The names of columns containing covariates (age, sex, PCs) are specified using the **--covariates** flag
- File must be either tab or space delimited
- The first two columns should be FID and IID, then
- Missing data represented as NA or -9 (for binary traits).

Other required inputs

- LD Reference Data: By default, BridgePRS uses as an LD reference the 1000 genomes AFR and EUR reference samples
- Genotype files: Bridge expects genotypes in Plink Format.
- Specify value of Fixation Index to help the selection of a prior using:
`--fst`
- To select only SNPs that have passed QC, you can include a single column text file using the
`--snp_file`

Exercise: Create Configuration files

File type	
Plink format	caf_genotype.bed, caf_genotype.bim caf_genotype.fam
Phenotypes	caf_test.dat, caf_validation.dat
African summary stats	yoruba.chr1.glm.gz,...,yoruba.chr22.glm.gz
UK Biobank summary stats	ukb.sumstats.gz
1000G Reference Populations: AFR, EUR, EAS, SAS, AMR	1000genomes.bed 1000genomes.bim 1000genomes.fam

Config 1

```

POP= CAF
LDPPOP= AFR
SUMSTATS_PREFIX= yoruba.chr
GENOTYPE_PREFIX= caf_genotype
PHENOTYPE_FILE= caf_test.dat
VALIDATION_FILE= caf_validation.dat

```

Config 2

```

POP= EUR
LDPPOP= EUR
SUMSTATS_PREFIX= ukb.sumstats
GENOTYPE_PREFIX= caf_genotype
PHENOTYPE_FILES= caf_test.dat

```