Lecture 5

Polygenic Risk Scores: Concepts, Calculation, and Challenges

by Dr. Mustafa İsmail Özkaraca

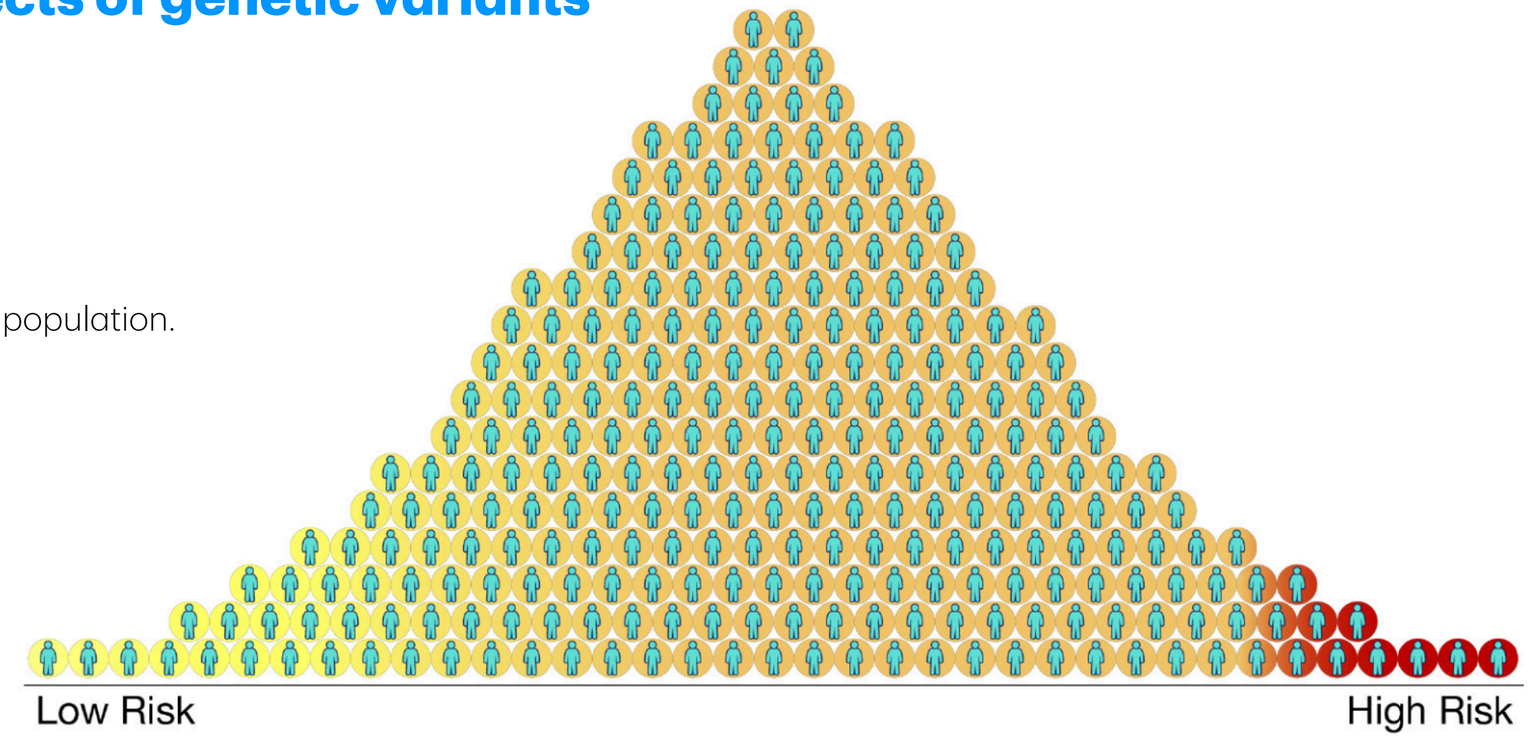
Contents

Understanding Polygenic Risk Score (PRS)?

PRS is a measure of **aggregated effects of genetic variants** on the **risk of developing a disease**.

- Genetic effects are **derived from GWAS**.
- Indicates **relative genetic risk** compared to a reference population.

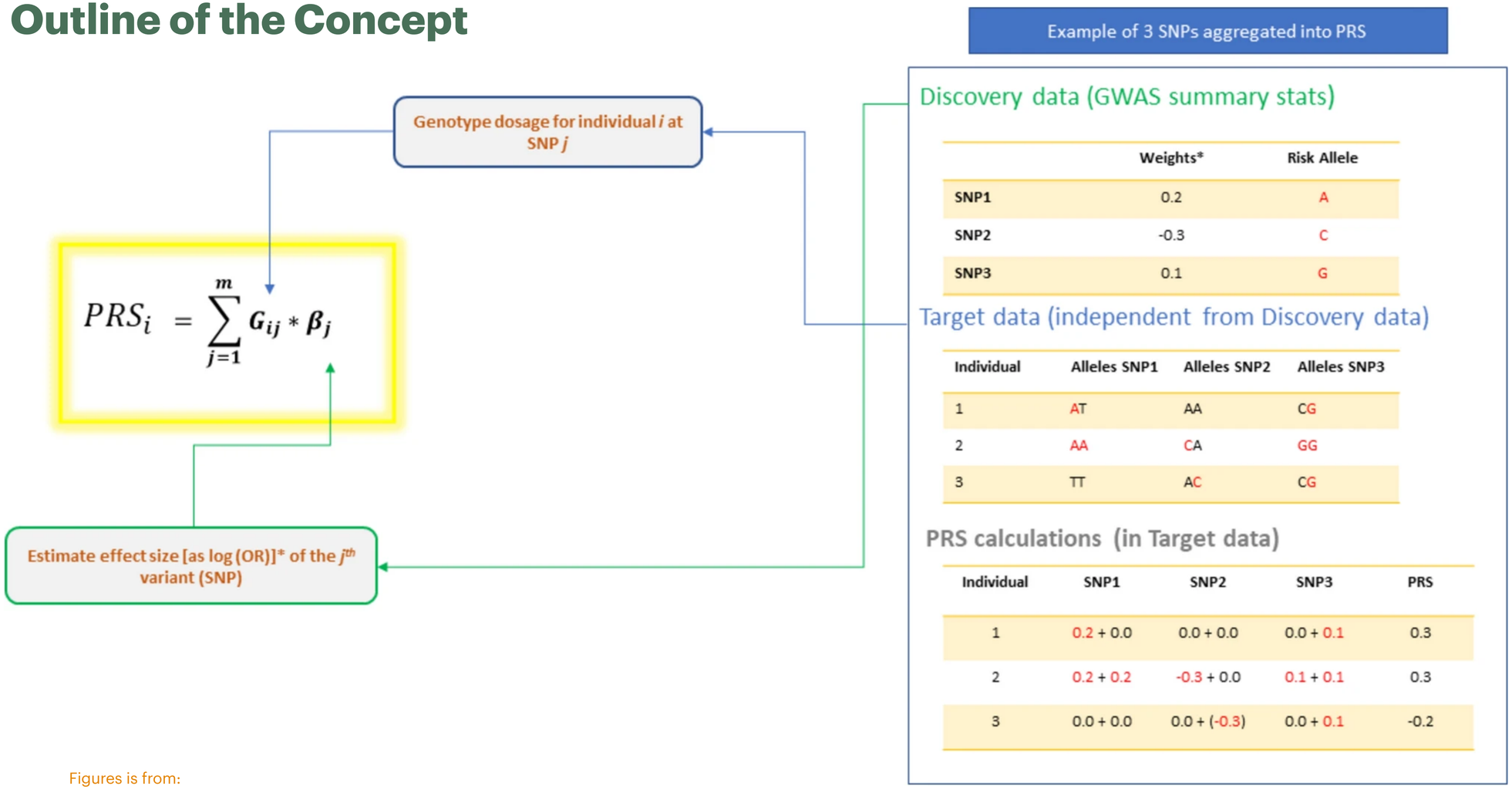
"**Absolute risk** is different. Absolute risk shows the likelihood of a disease occurring. Women who carry a BRCA1 mutation have a 60-80% absolute risk of breast cancer. This would be true even without any comparison to any groups of people." [1]



Figures is from: https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores

Calculating PRS

Outline of the Concept



Example of 3 SNPs aggregated into PRS

Discovery data (GWAS summary stats)

	Weights*	Risk Allele
SNP1	0.2	A
SNP2	-0.3	C
SNP3	0.1	G

Target data (independent from Discovery data)

Individual	Alleles SNP1	Alleles SNP2	Alleles SNP3
1	AT	AA	CG
2	AA	CA	GG
3	TT	AC	CG

PRS calculations (in Target data)

Individual	SNP1	SNP2	SNP3	PRS
1	0.2 + 0.0	0.0 + 0.0	0.0 + 0.1	0.3
2	0.2 + 0.2	-0.3 + 0.0	0.1 + 0.1	0.3
3	0.0 + 0.0	0.0 + (-0.3)	0.0 + 0.1	-0.2

LD ALERT !!

Figures is from:

Ndong Sima, C.A.A., Step, K., Swart, Y. et al. Methodologies underpinning polygenic risk scores estimation: a comprehensive overview. Hum. Genet. 143, 1265–1280 (2024).

Calculating PRS

C + T Approach

```
plink \
    --bfile EUR.QC \
    --clump-p1 1 \
    --clump-r2 0.1 \
    --clump-kb 250 \
    --clump Height.QC.Transformed \
    --clump-snp-field SNP \
    --clump-field P \
    --out EUR
```

Each of the new parameters corresponds to the following

Parameter	Value	Description
clump-p1	1	P-value threshold for a SNP to be included as an index SNP. 1 is selected such that all SNPs are include for clumping
clump-r2	0.1	SNPs having r^2 higher than 0.1 with the index SNPs will be removed
clump-kb	250	SNPs within 250k of the index SNP are considered for clumping
clump	Height.QC.Transformed	Base data (summary statistic) file containing the P-value information
clump-snp-field	SNP	Specifies that the column `SNP` contains the SNP IDs
clump-field	P	Specifies that the column `P` contains the P-value information

A more detailed description of the clumping process can be found here

Note

The r^2 values computed by `--clump` are based on maximum likelihood haplotype frequency estimates

Figures is from: https://choishingwan.github.io/PRS-Tutorial/plink/

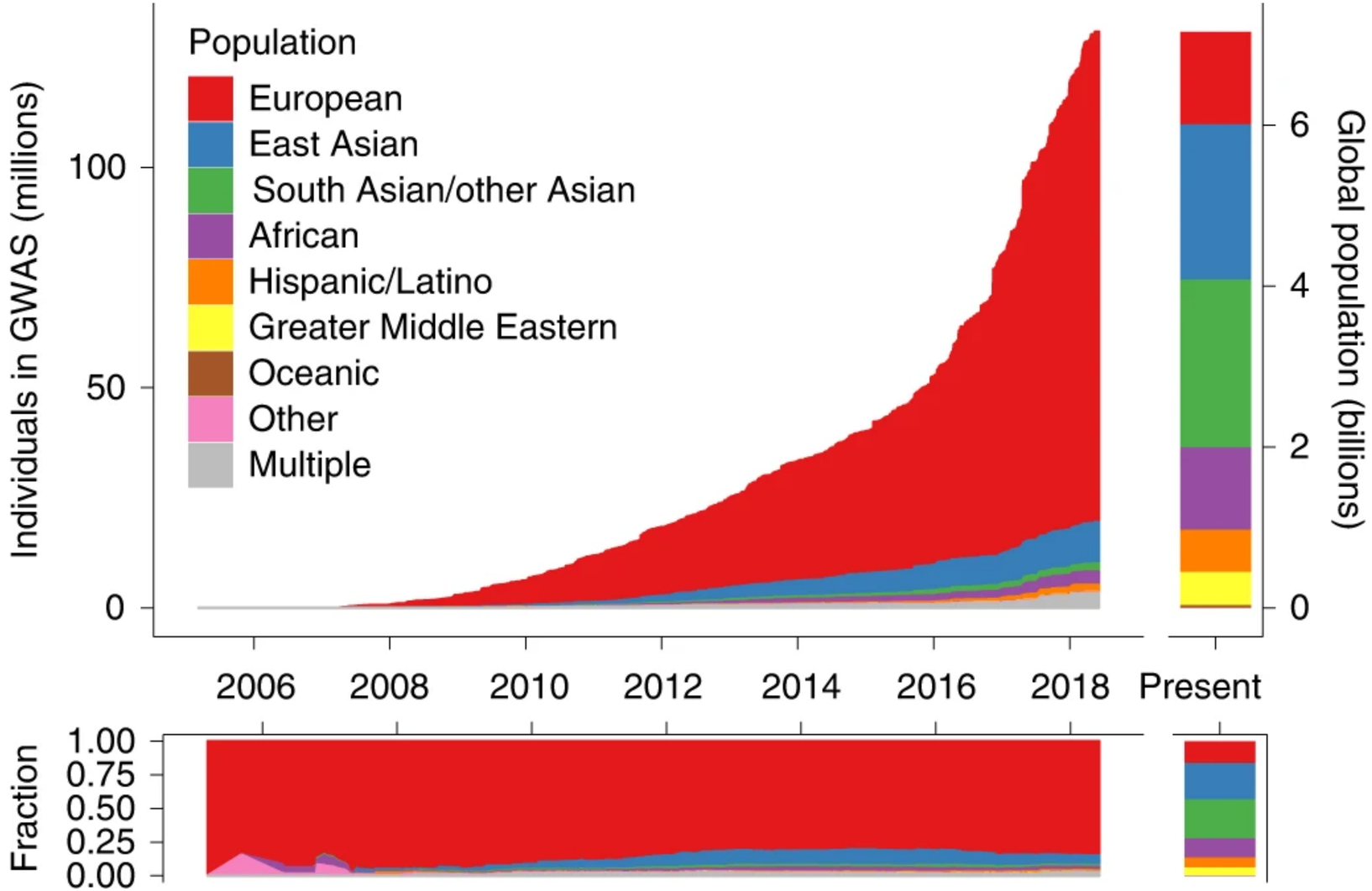
PRS Tool	Methodology[1]	Data requirements	Implementation	Source	Website/GitHub repository
PLINK (2007)	C + T (manual thresholding)	GWAS summary statistics	PLINK framework	(Purcell et al., 2007)	https://www.cog-genomics.org/plink
PLINK2 (2015)				(Chang et al., 2015)	https://www.cog-genomics.org/plink2
PRSice (2015)	C + T	GWAS summary statistics	R and C++	(Euesden et al., 2015)	https://github.com/choishingwan/PRSice
LDpred (2015)	Bayesian regression	GWAS summary statistics	Python based	(Vilhjálmsson et al., 2015)	https://github.com/bvilhjal/ldpred
Lassosum (2017)	Lasso penalized regression	GWAS summary statistics or individual-level data	R package: *lassosum*	(Mak et al., 2017)	https://github.com/tshmak/lassosum
PRSice-2 (2019)	C + T	GWAS summary statistics	R and C++	(Choi and O'Reilly, 2019)	https://choishingwan.github.io/PRS-Tutorial/prsice/
PRS-CS (2019)	Bayesian regression with continuous shinkage priors	GWAS summary statistics and LD reference panel	Python based	(Ge et al., 2019)	https://github.com/getian107/PRScs
SBayesR (2019)	Bayesian multiple regression	GWAS summary statistics	GCTB software	(Lloyd-Jones et al., 2019)	http://cnsgenomics.com/software/gctb
LDpred-funct (2019)	Bayesian regression	GWAS summary statistics	Python package	(Márquez-Luna et al., 2021)	https://github.com/carlaml/LDpred-funct
LDpred2 (2020)	Bayesian regression	GWAS summary statistics and LD reference panel	R package: *bigsnpR*	(Privé, Arbel, and Vilhjálmsson, 2021)	https://privefl.github.io/bigsnpr/articles/LDpred2.html
Lassosum2 (2021)	Lasso penalized regression	GWAS summary statistics or individual-level data	R package: *bigsnpR*	(Privé, Arbel, Aschard, et al., 2021)	https://privefl.github.io/bigsnpr/reference/snp_lassosum2.h
PRS-CSx (2022)	Bayesian regression with shared continuous shrinkage priors	Multiple GWAS summary statistics	Python based	(Ruan et al., 2022)	https://github.com/getian107/PRScsx
BridgePRS (2023)	Bayesian regression	GWAS summary statistics	Shell script	(Hoggart et al., 2023)	https://github.com/clivehoggart/BridgePRS
PRSet (2023)	Pathway-based	GWAS summary statistics	R and C++	(Choi et al., 2023)	Currently under active development but can be downloaded fo free under the PRSice website
GAUDI (2024)	Fused lasso	GWAS summary statistics	R, Python, Shell script	(Sun et al., 2024)	https://github.com/quansun98/GAUDI

[1]C+T: (LD-) Clumping and Thresholding; lasso: Least Absolute Shrinkage and Selection Operator

Table is from:

Ndong Sima, C.A.A., Step, K., Swart, Y. et al. Methodologies underpinning polygenic risk scores estimation: a comprehensive overview. Hum. Genet. 143, 1265–1280 (2024).

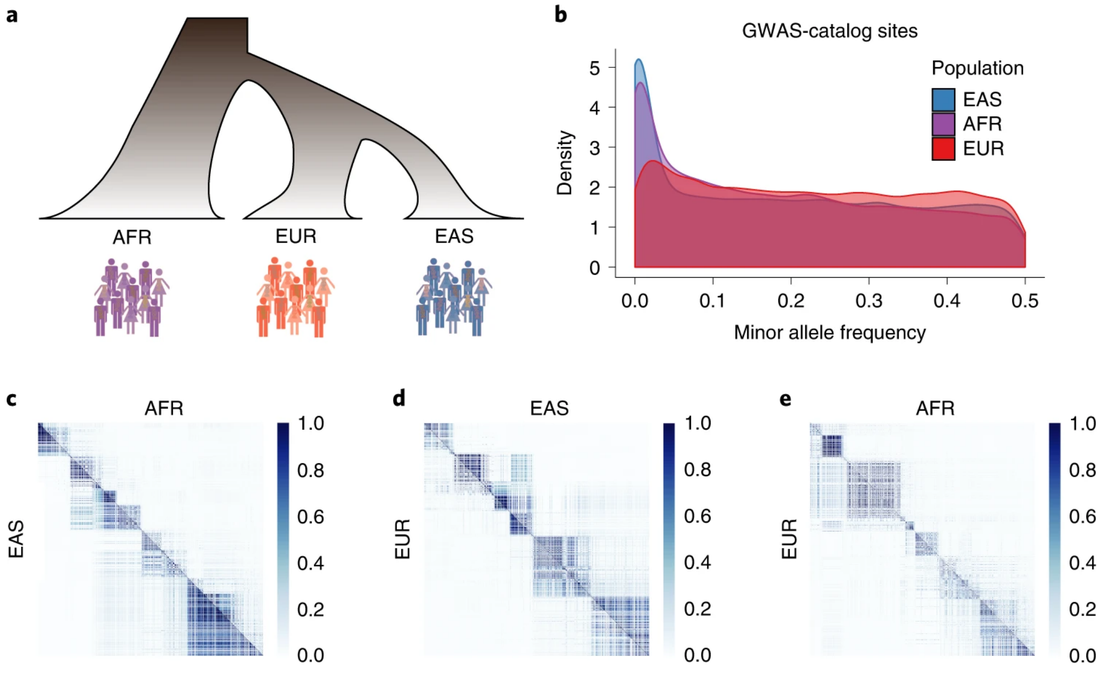
Population Bias

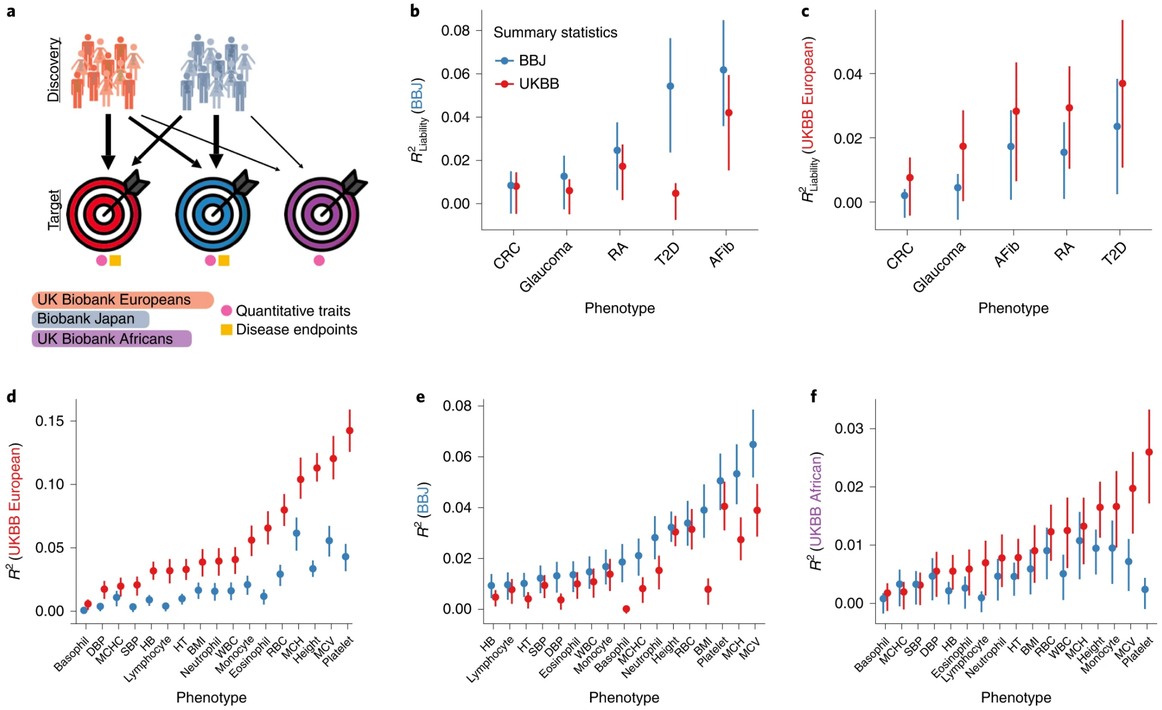


Figures is from:

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

Population Bias





Figures are from:

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

What’s Next

1. Understanding Heritability

2. Early Methods: Twin Studies

3. Modern Approaches: SNP-Based Heritability

4. Challenges in Heritability Studies