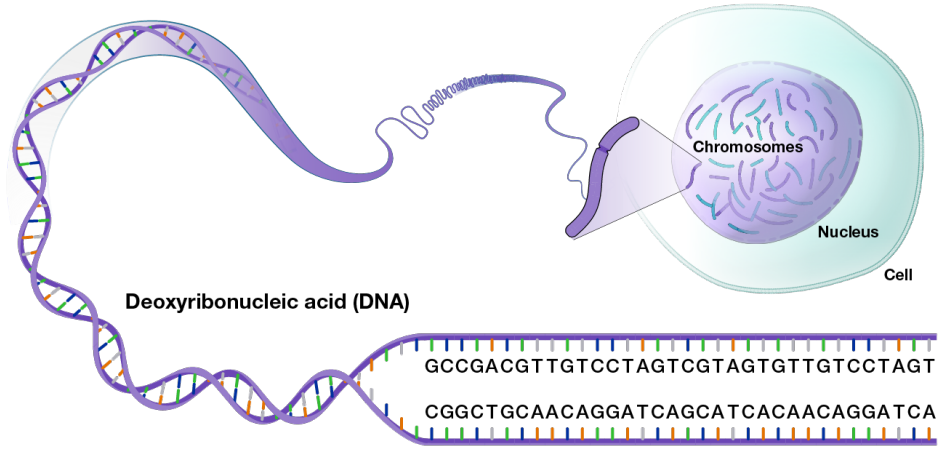
Introduction to Polygenic Risk Score

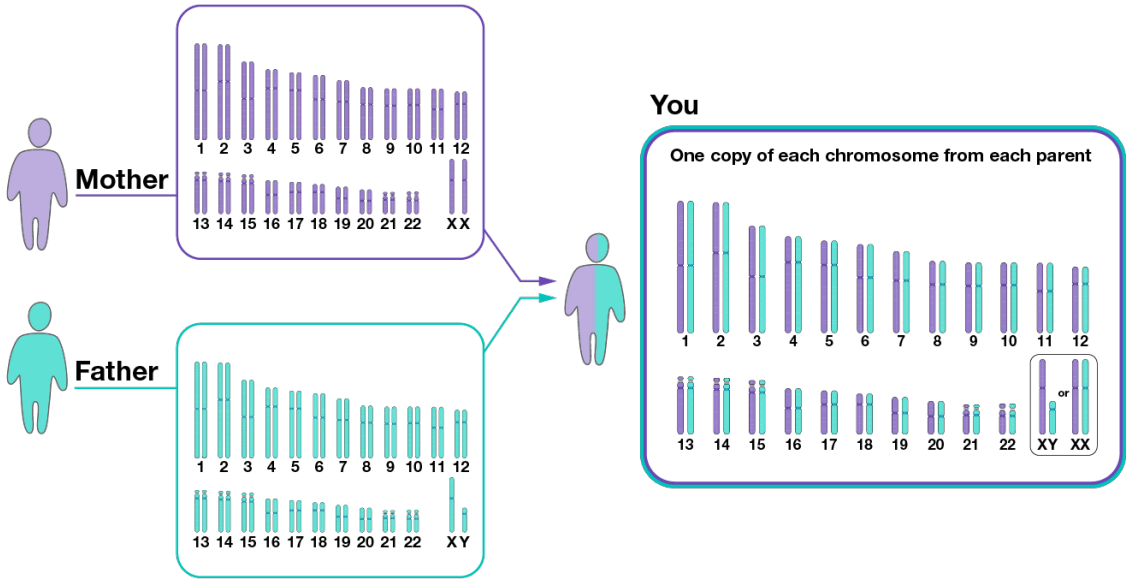
Lin, Chien-Wei

Taiwan Biobank

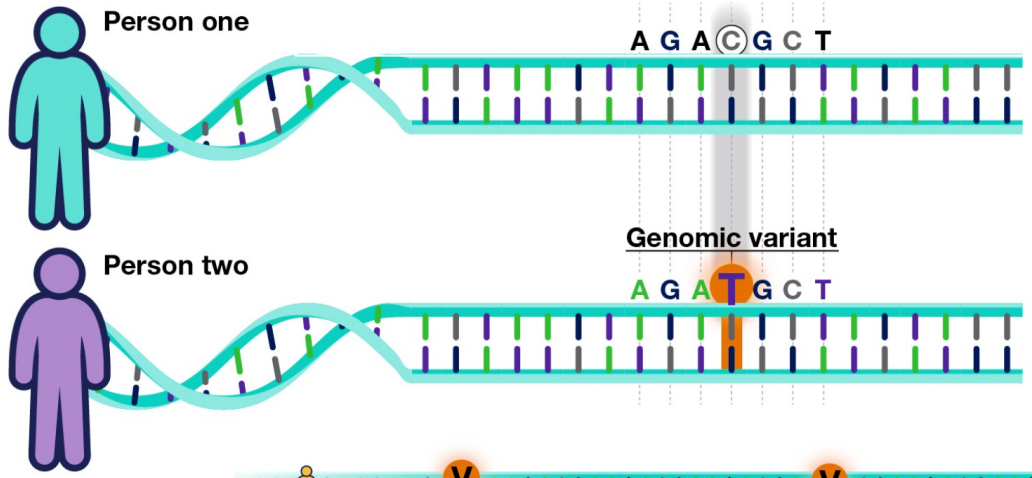
National Human Genome Research Institute

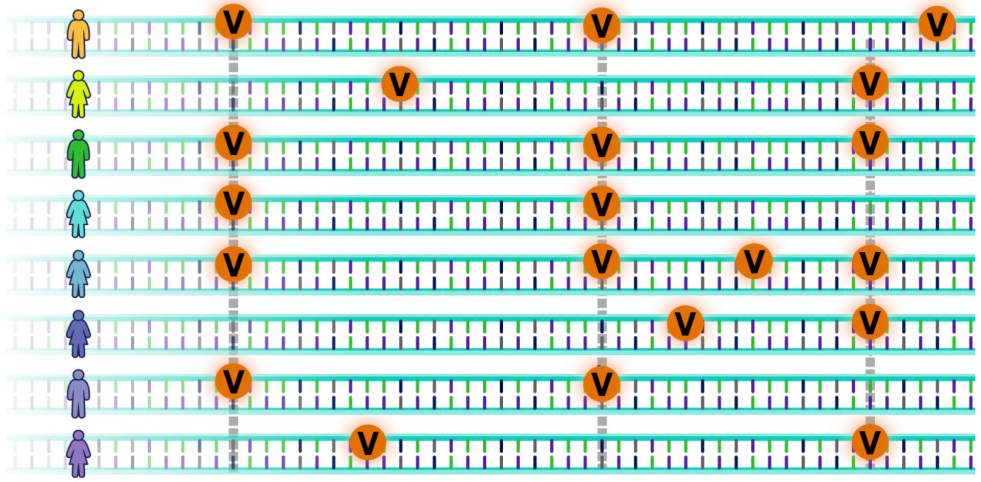
Human genomic variation





Single-Nucleotide Variant (SNV)

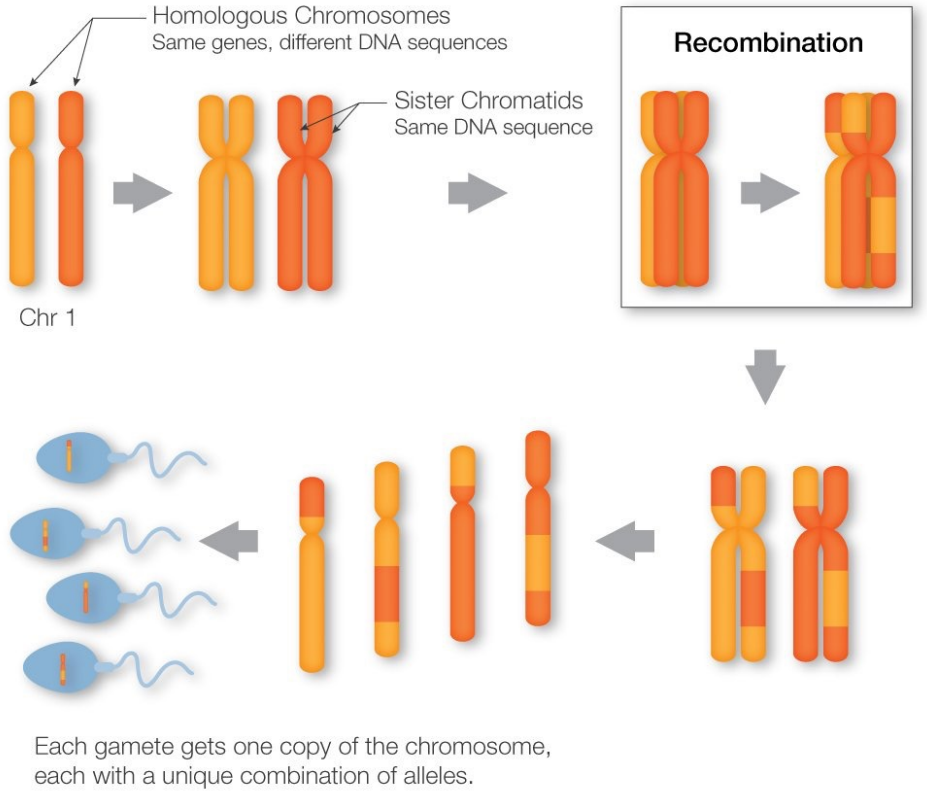


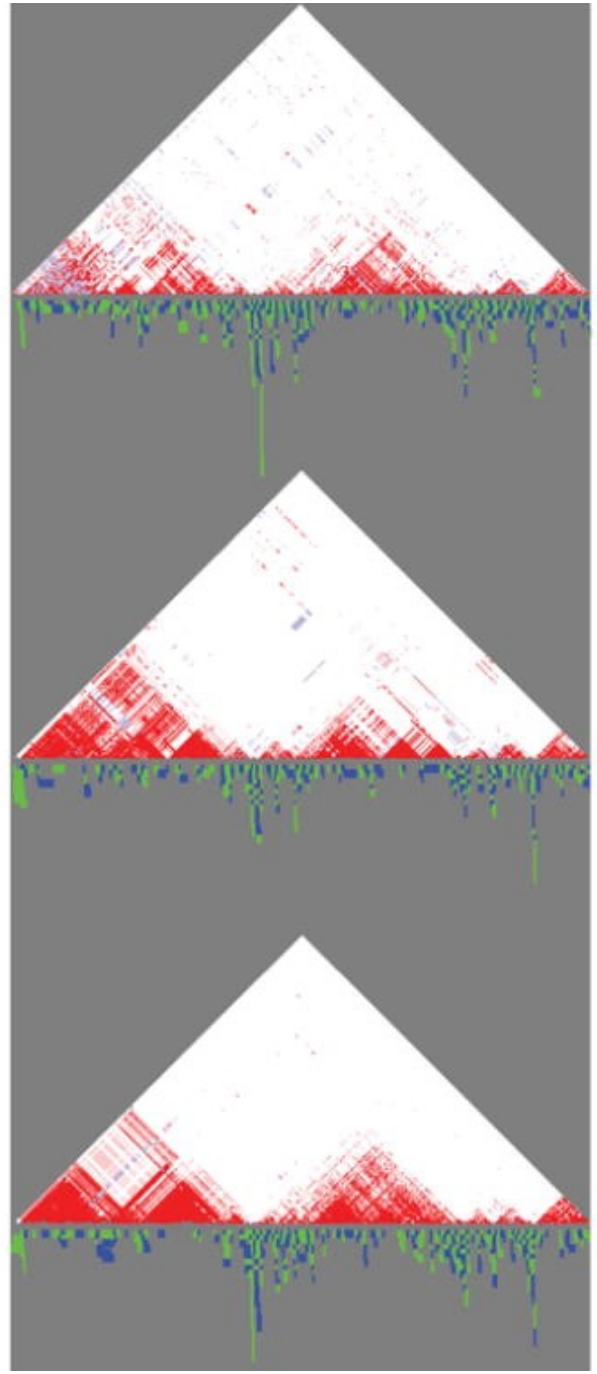


Linkage disequilibrium (LD)

LD, usually expressed as r^2, is the nonrandom association of alleles at different loci in a given population

- $r^2 = 1$: the alleles are completely correlated
- $r^2 = 0$: the alleles are in linkage equilibrium

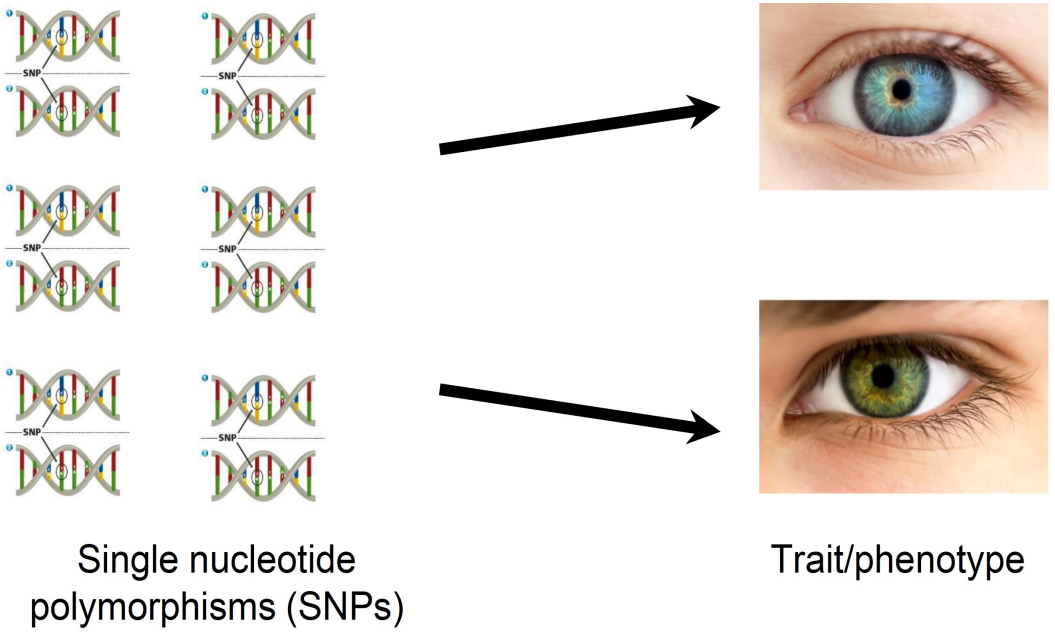




Nature. 2003;426(6968):789-96.

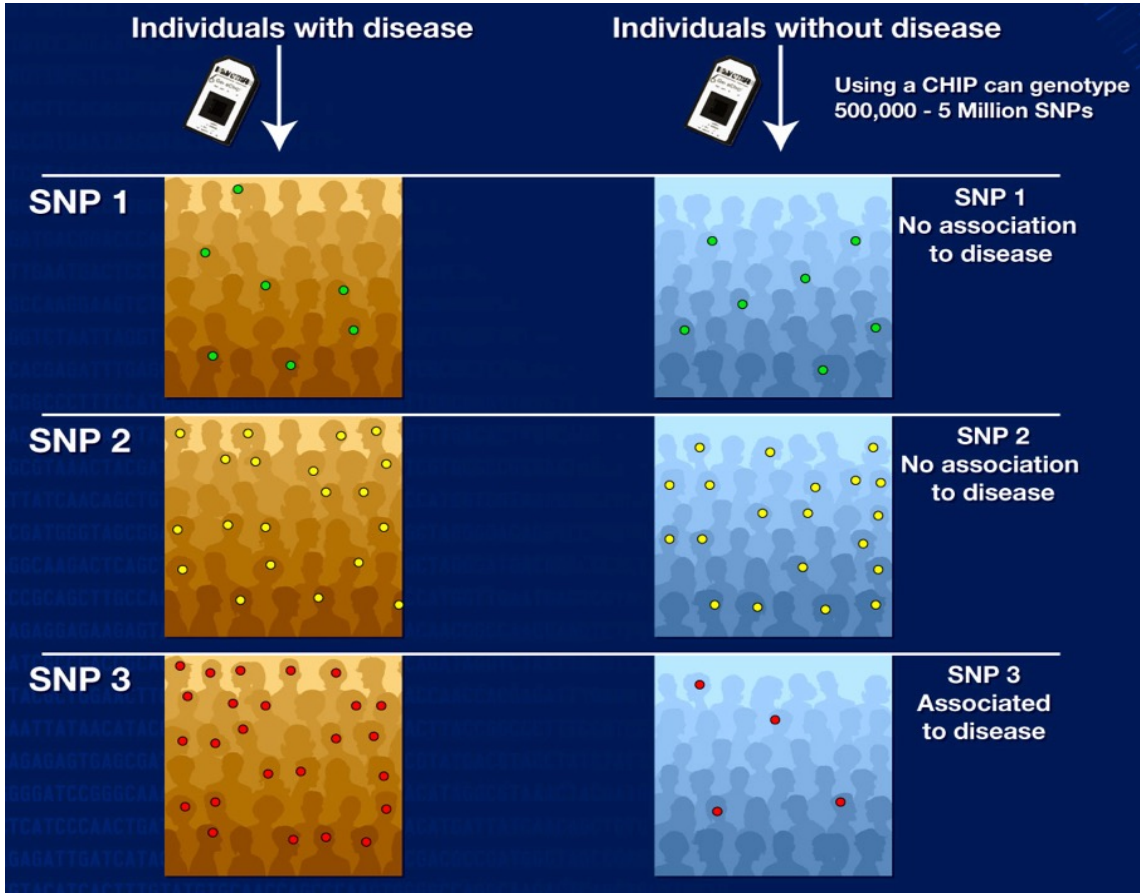
Genome-wide association study (GWAS)

GWAS involves scanning markers across the genomes of many people to find genetic variations associated with a particular disease



Single nucleotide polymorphisms (SNPs)

Trait/phenotype



NIH National Human Genome Research Institute

Beyond GWAS

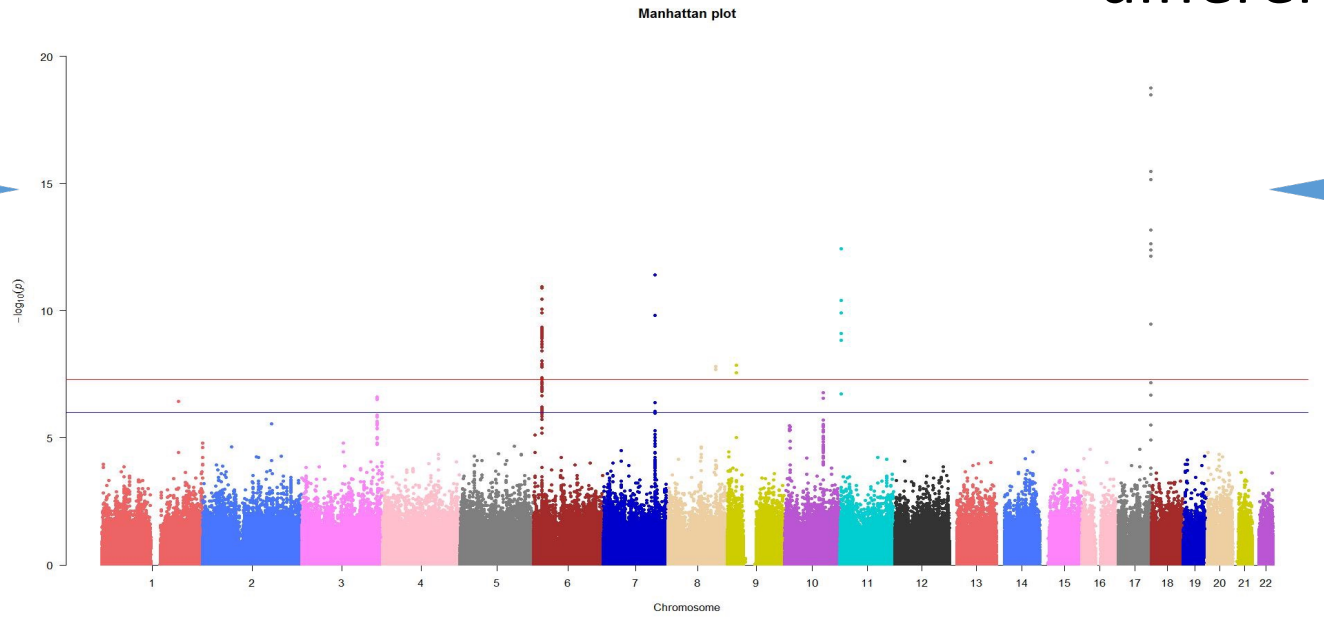
Mendelian randomization

GWAS meta-analysis

Analyzing genetic correlation between different traits

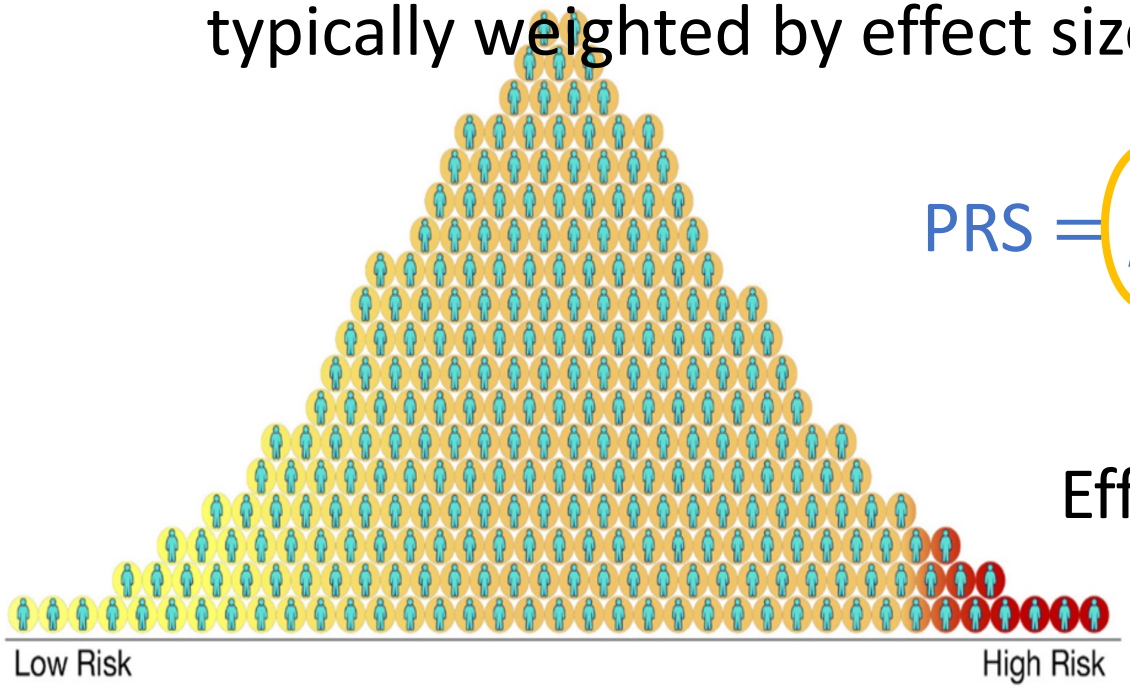
Polygenic risk score analysis

Estimation of SNP heritability



Polygenic Risk Score (PRS)

- PRS uses information from external GWAS results to predict risk in our own study sample
- PRS is the sum of trait-associated risk alleles across many genetic loci, typically weighted by effect sizes from a GWAS



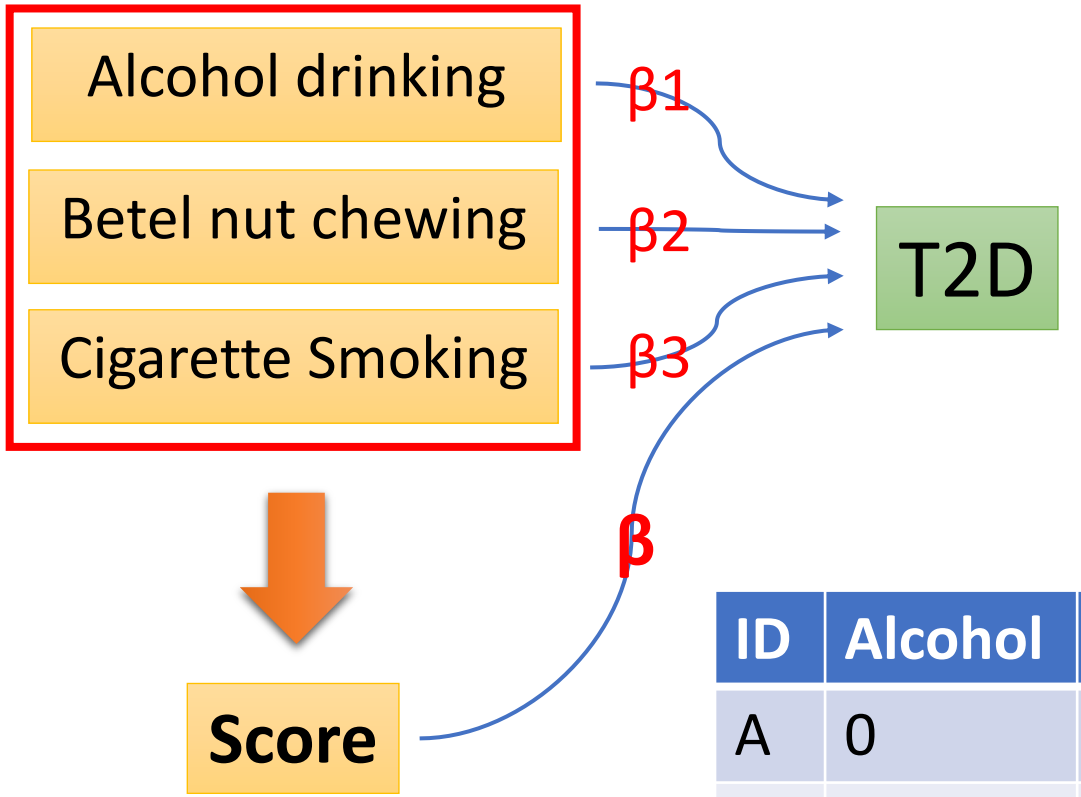
$$\text{PRS} = \beta_1 SNP_1 + \beta_2 SNP_2 + \cdots + \beta_n SNP_n$$

Effect size

Number of risk alleles

Number of variants

PRS construction



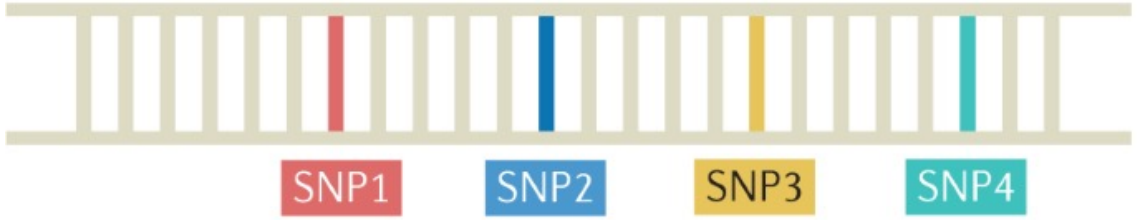
ID	Alcohol	Betel nut	Cigarette	Score
A	0	0	0	0
B	1	0	1	2
C	1	1	1	3

ID	Alcohol	Betel nut	Cigarette	Score
A	0	0	0	0 x β1 + 0 x β2 + 0 x β3
B	1	0	1	1 x β1 + 0 x β2 + 1 x β3
C	1	1	1	1 x β1 + 1 x β2 + 1 x β3

PRS construction

1 GWAS summary statistics

Allele	A	C	T	A
Effect	+1.5	−0.5	+2.0	−1.5



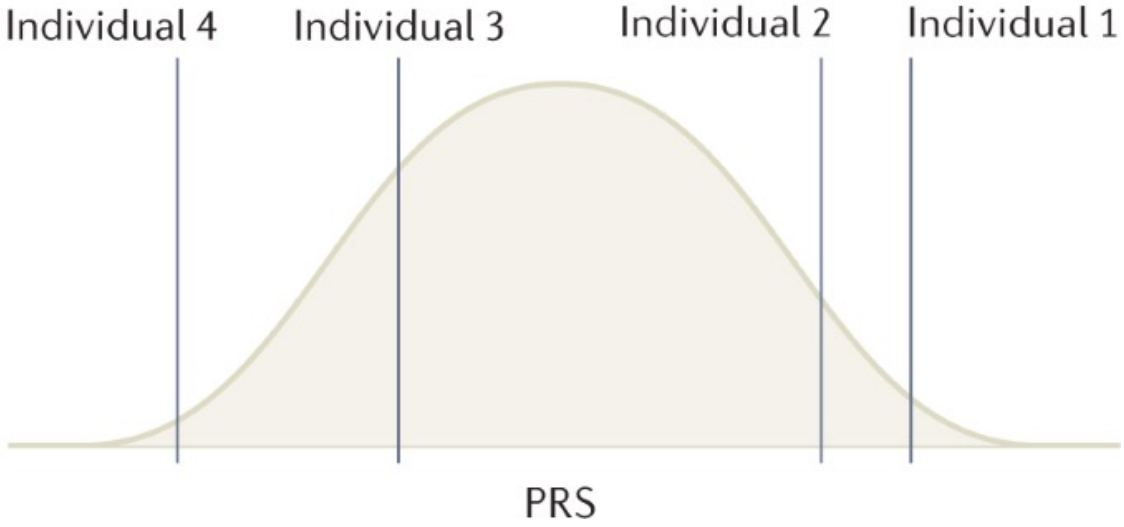
2 Genotype data

	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA

3 Polygenic risk score

Individual 1	1.5	−	0.5	+	4.0	−	0.0	= **5.0**
Individual 2	1.5	−	0.0	+	2.0	−	1.5	= **2.0**
Individual 3	0.0	−	1.0	+	2.0	−	1.5	= **−0.5**
Individual 4	0.0	−	1.0	+	0.0	−	3.0	= **−4.0**

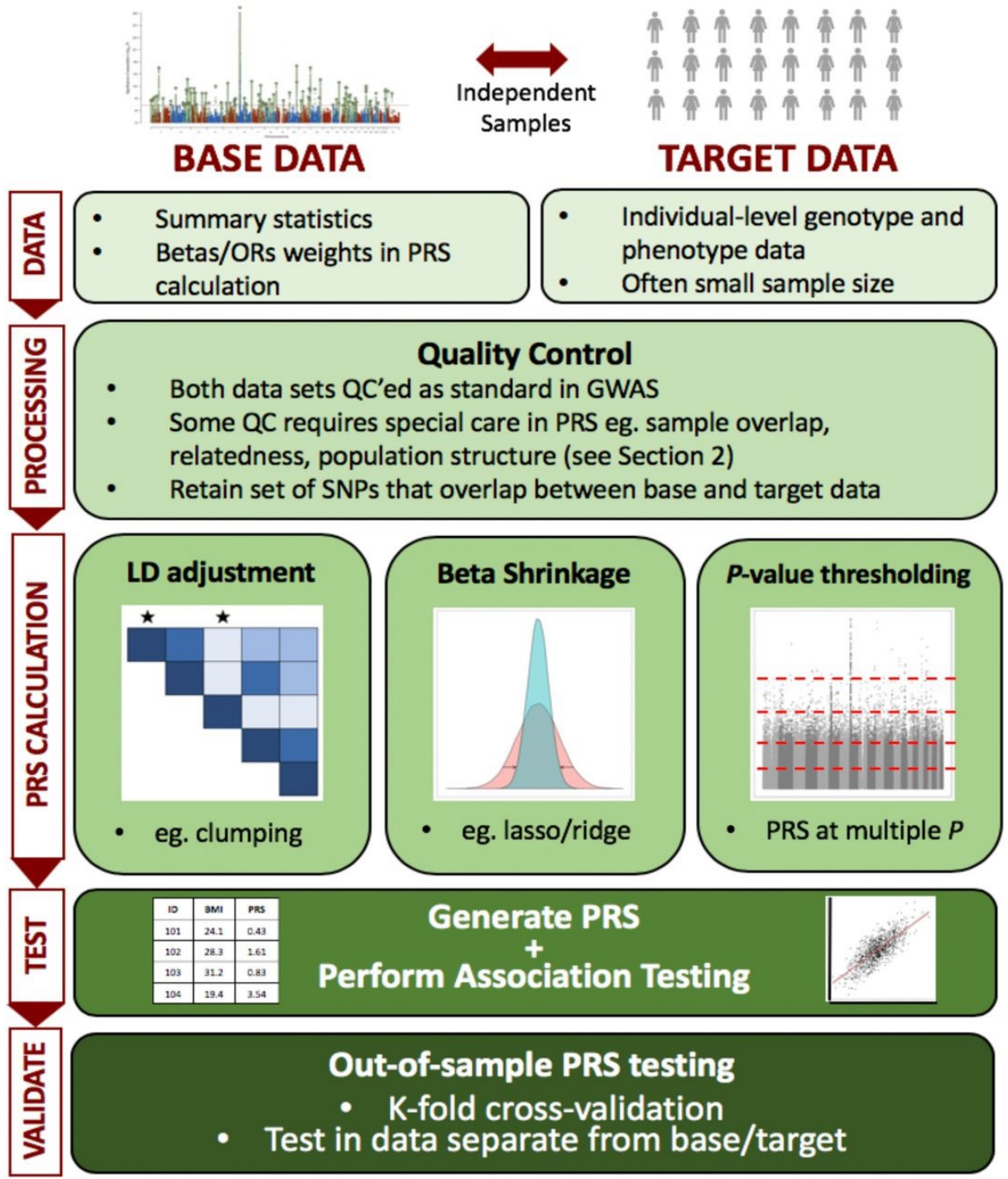
4 PRS distribution



Nature Reviews Methods Primers volume 1, Article number: 59 (2021)

PRS analysis workflow

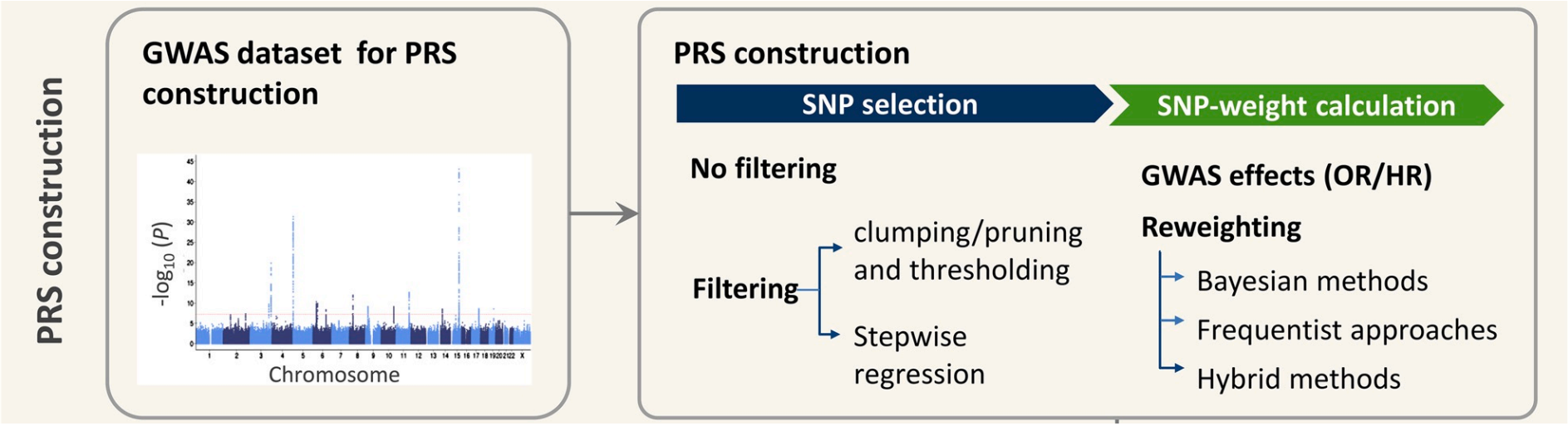
- Summary statistics
  - Estimated association coefficients for every variant analyzed in a GWAS
- Base data
  - Large GWAS providing information on association coefficients
- Target data
  - Relatively smaller studies used to determine tuning parameters or/and to report final model performance



Nat Protoc. 2020 Sep;15(9):2759-2772.

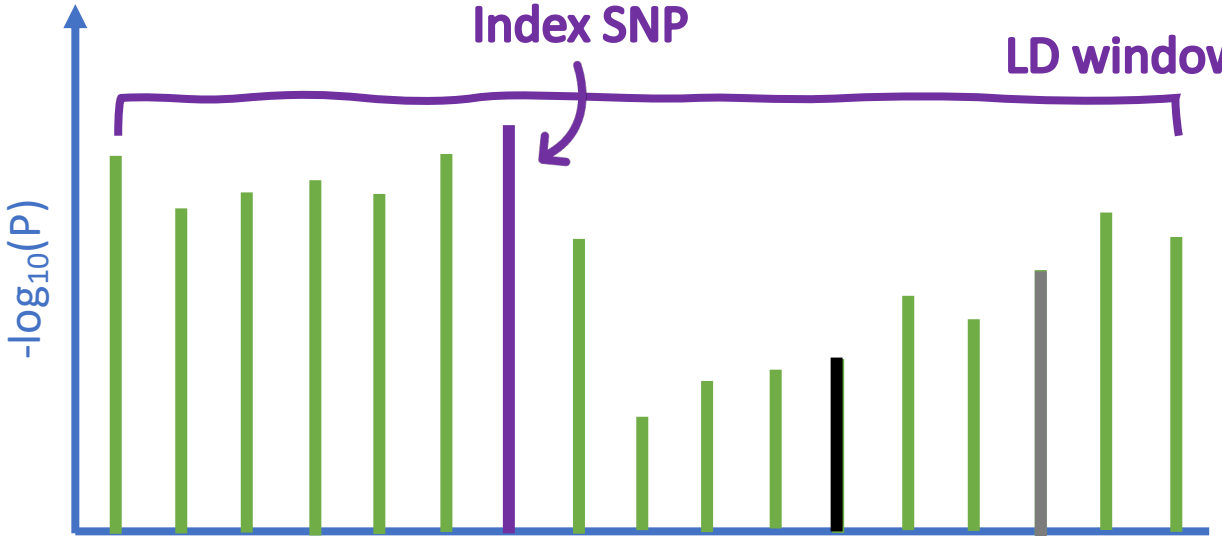
Challenges in PRS development

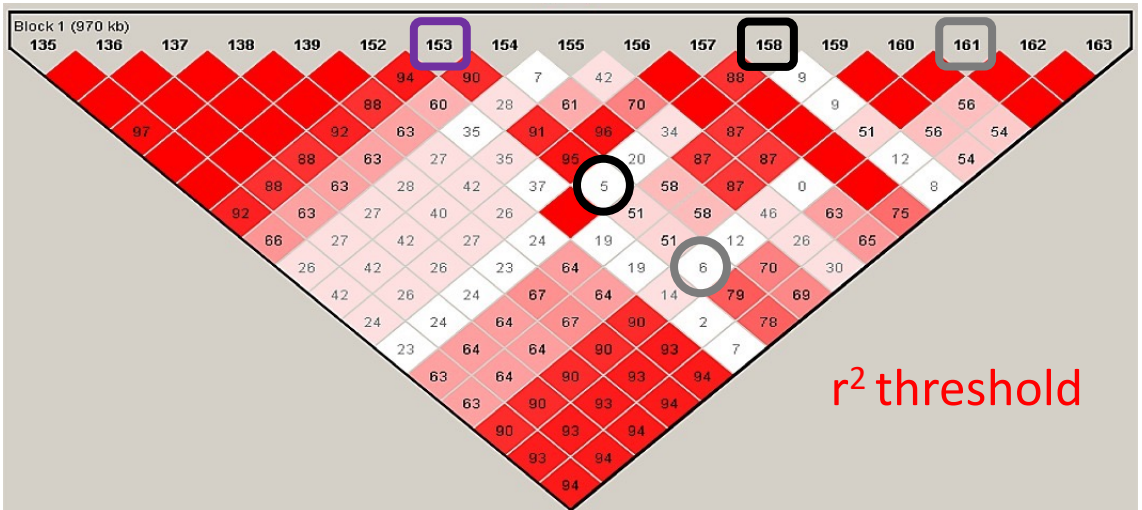
- Variant selection
  - Determine which variants to include
- Weight calculation
  - The trait-associated weights to assign to the selected variants



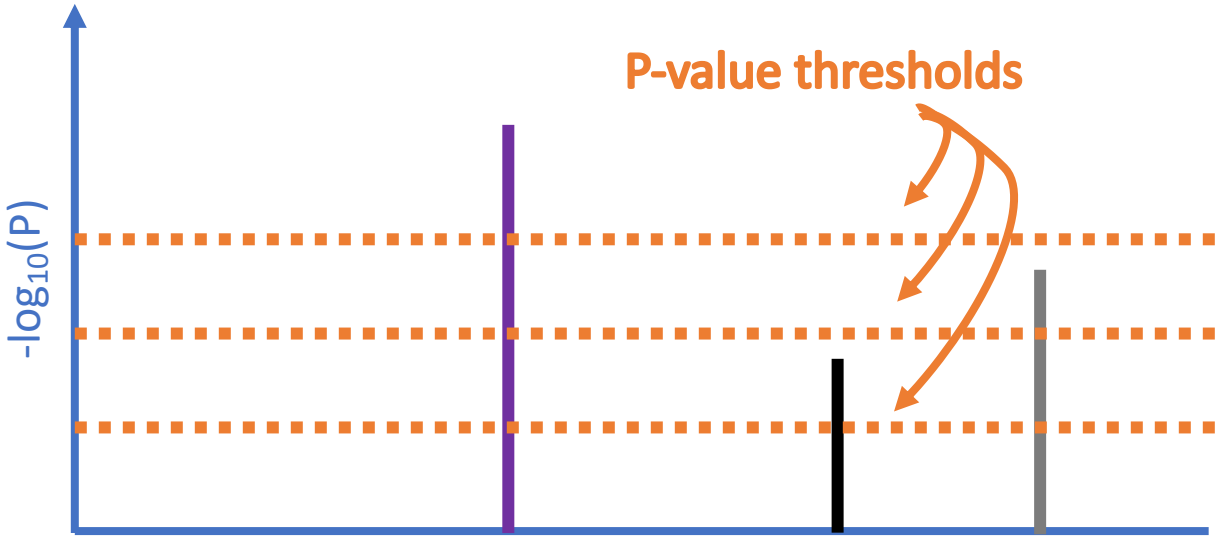
Med Rev (2021). 2022 Feb 14;1(2):129-149.

Clumping and thresholding (C+T)

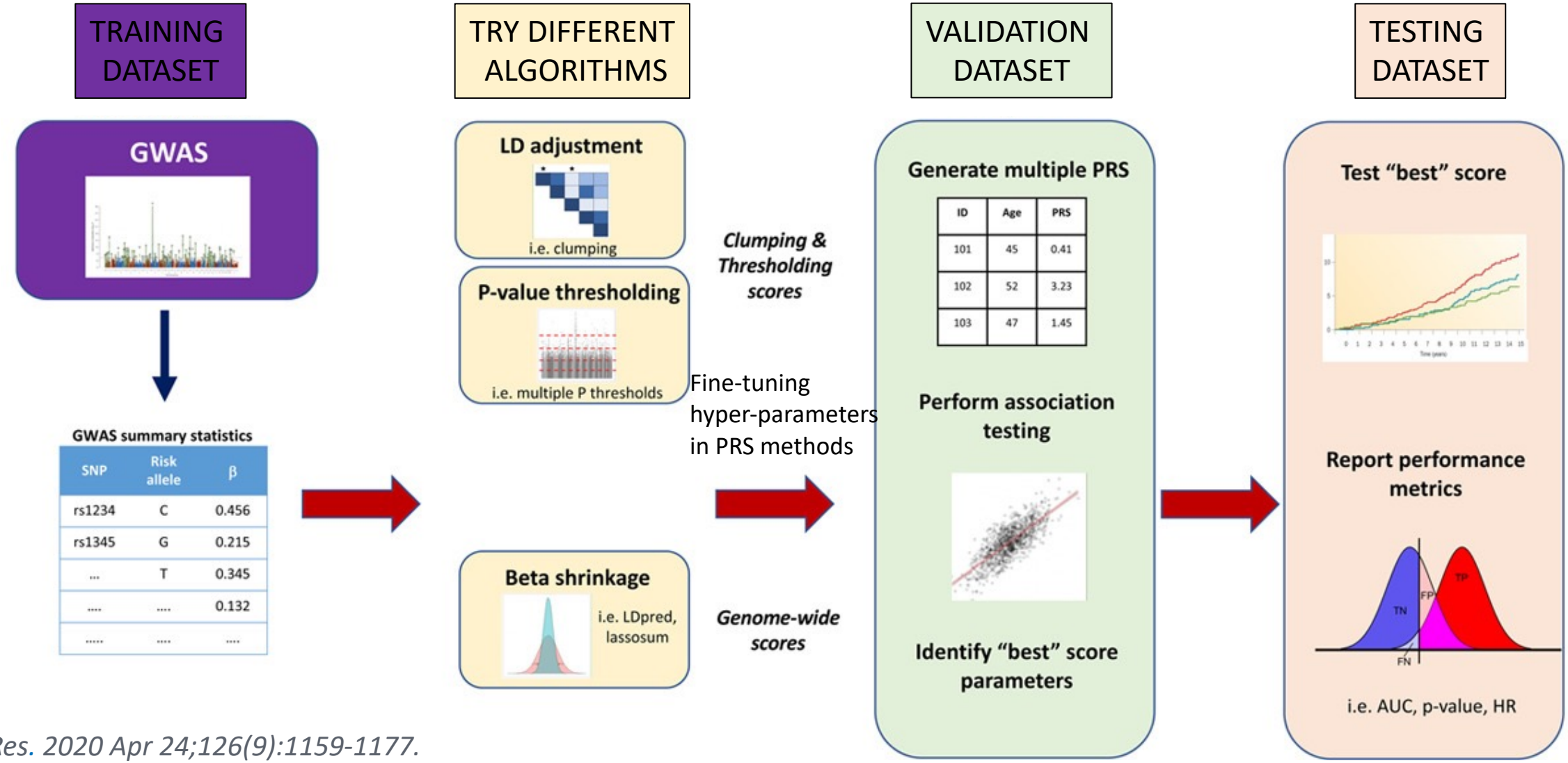




$$Score1 = \beta SNP + \beta SNP + \beta SNP$$
$$Score2 = \beta SNP + \beta SNP$$
$$Score3 = \beta SNP$$



Development of PRS



Circ Res. 2020 Apr 24;126(9):1159-1177.

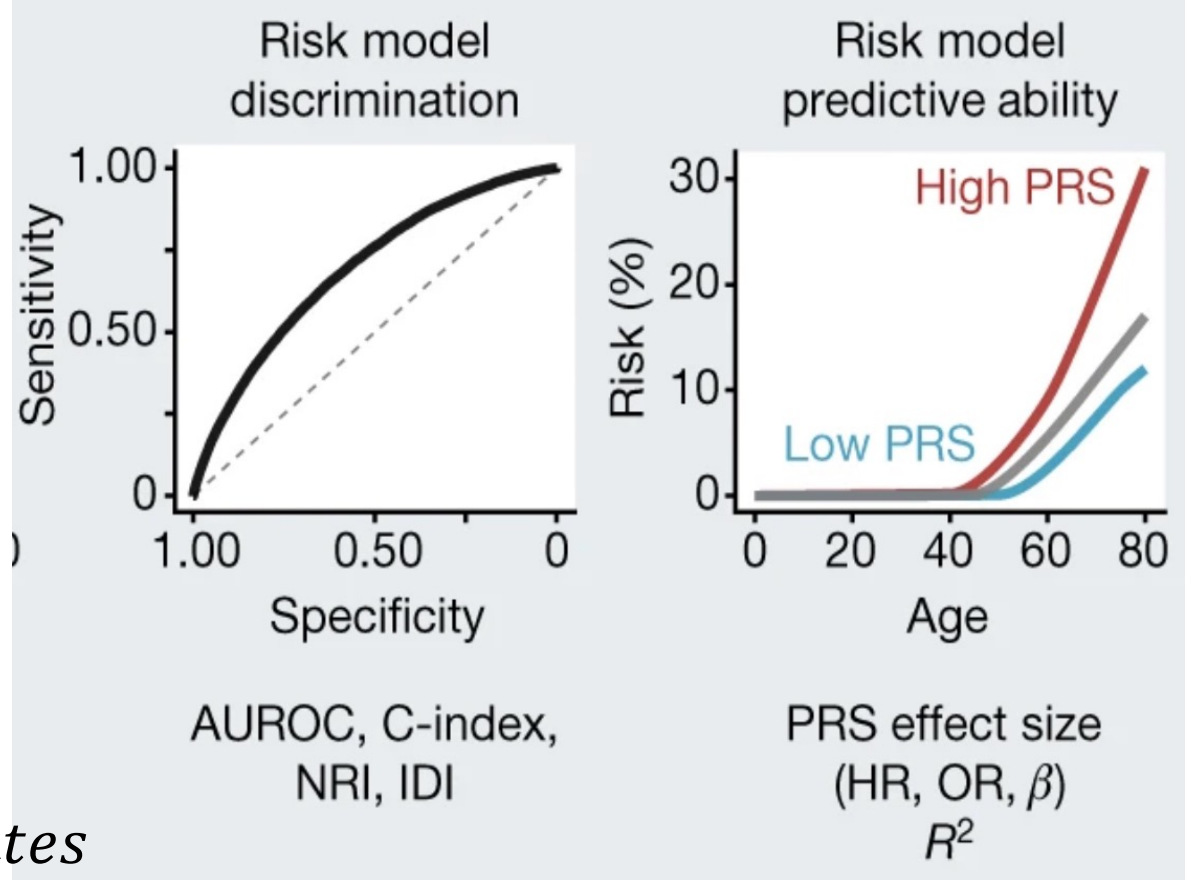
Evaluating PRS

- Accuracy of the PRS can be assessed by various metrics
  - Quantitative traits: R^2
  - Binary traits: AUC

- Comparing the difference in phenotypic variance explained by the PRS in two models

$$full\ model:\ Phenotype \sim PRS + Covariates$$

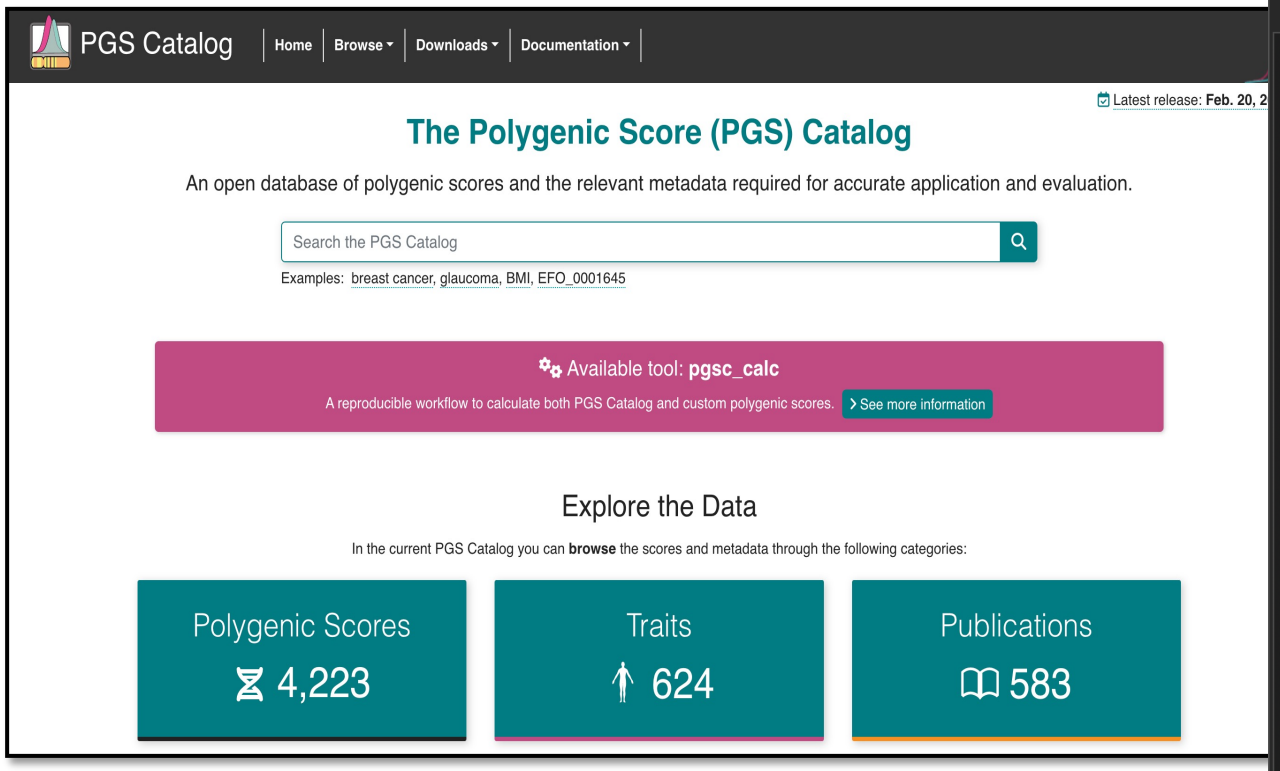
$$reduced\ model:\ Phenotype \sim Covariates$$

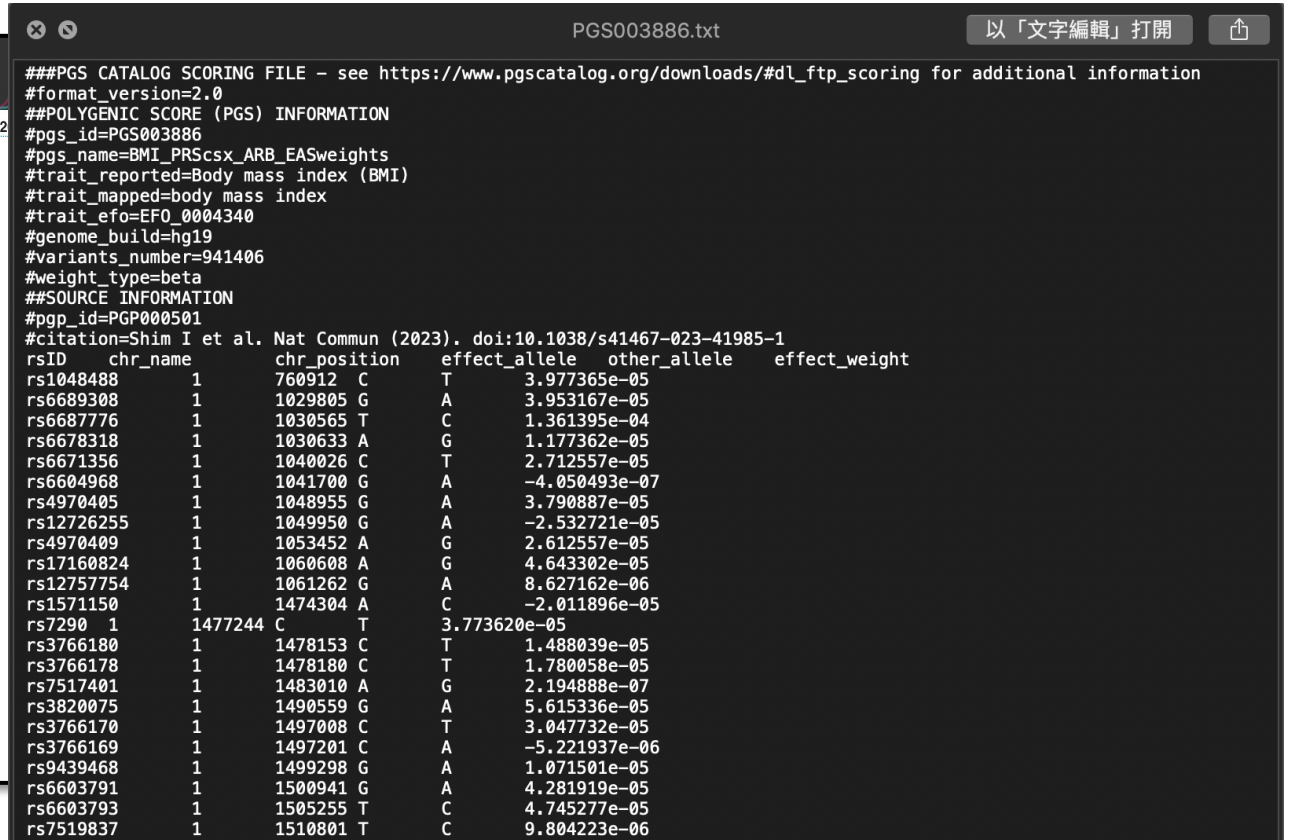


Nature. 2021 Mar;591(7849):211-219.

Open resource

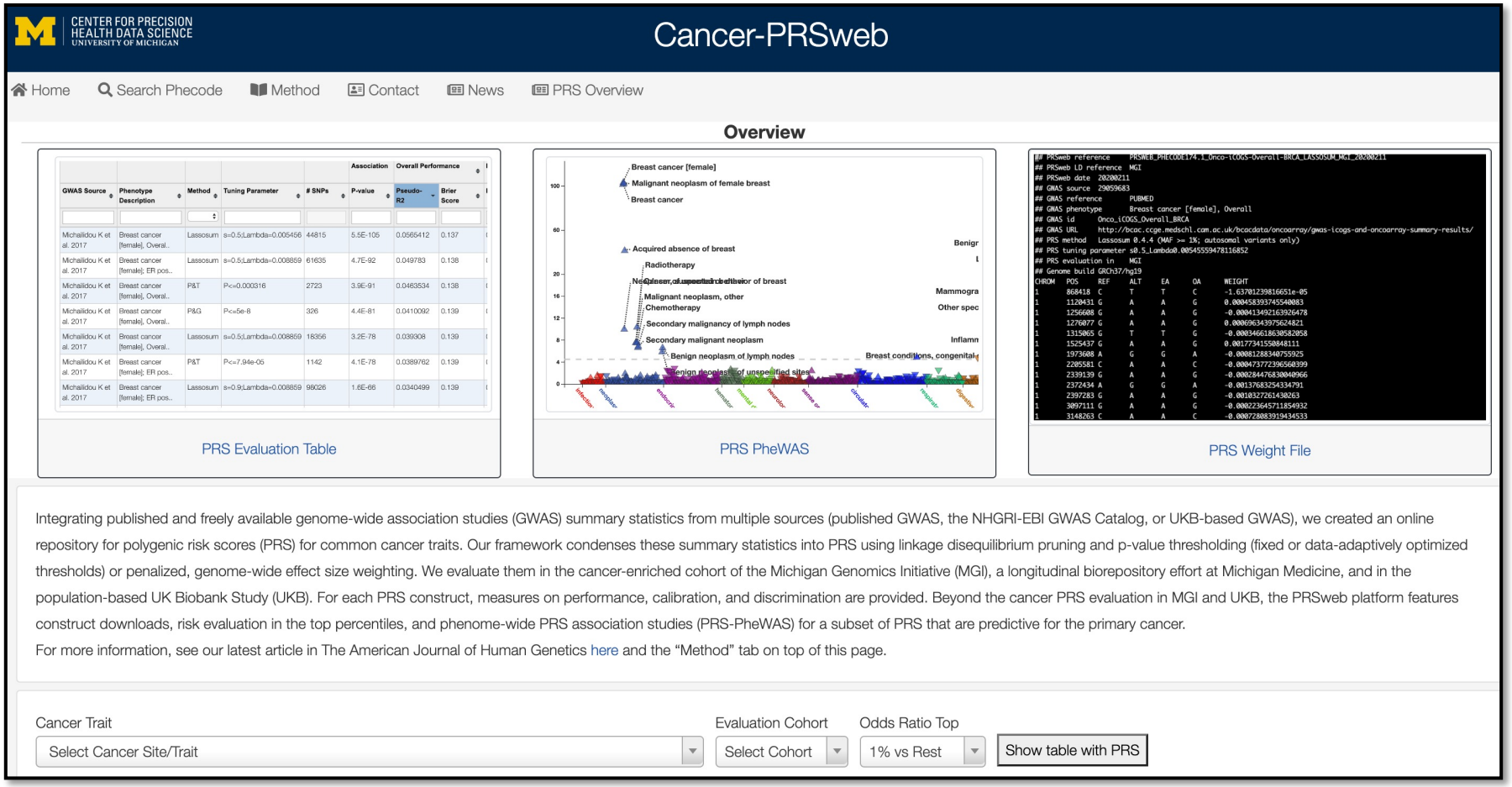
- The Polygenic Score (PGS) Catalog (https://www.pgscatalog.org)
- Scoring info and relevant metadata necessary to apply PRS



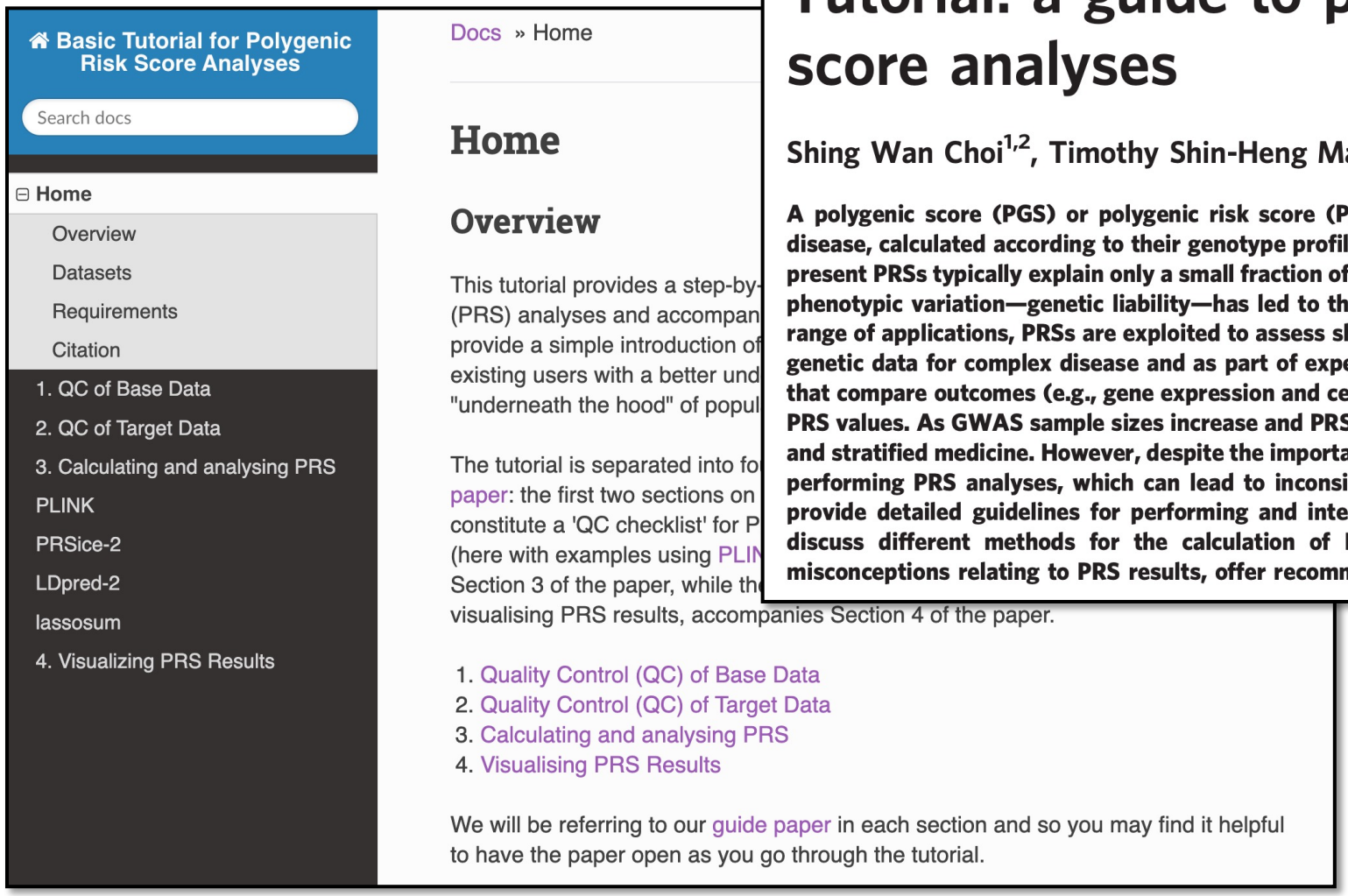


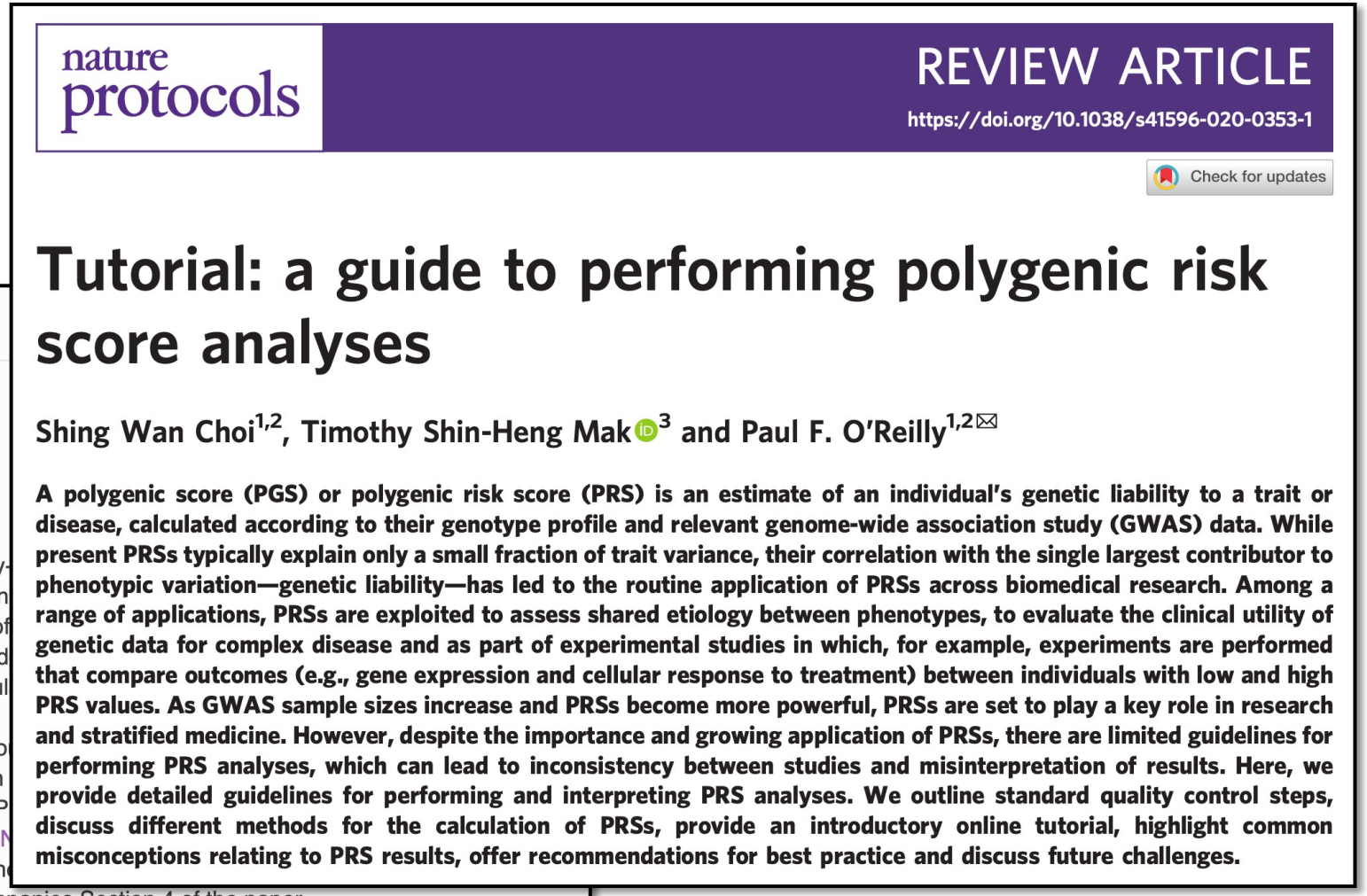
Open resource

- Cancer-PRSweb (https://prsweb.sph.umich.edu:8443)



PRS tutorial






THANK YOU