McGill initiative in Computational Medicine

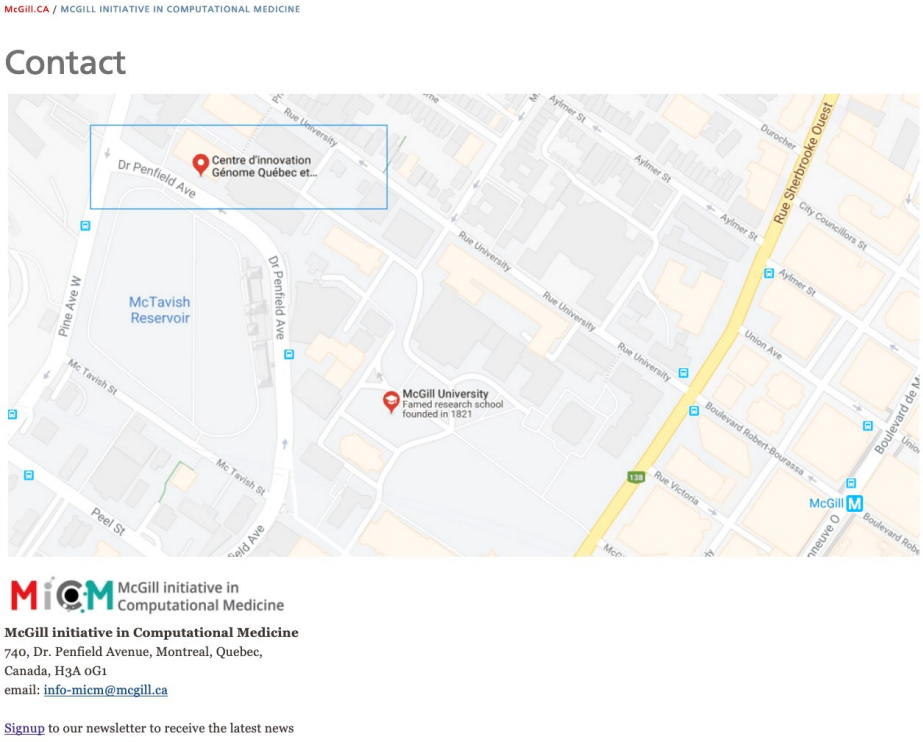
GWAS & Polygenic Risk Score

Workshop Lead: Yuan Ding

Date: 2025-08-14

McGill initiative in Computational Medicine

Mission : aims to deliver inter-disciplinary research programs and empower the use of data in health research and health care delivery



https://www.mcgill.ca/micm

Learning outcomes

1. Fundamental understanding of GWAS
2. Fundamental understanding of PRS
3. Access and interpret GWAS summary statistics.
4. Compute PRS and interpret results.
5. Understand the pitfalls and limitations of GWAS and PRS studies.

Outline of this workshop

Part 1. Genome-wide association studies (GWAS)

Part 2. Polygenic risk scores (PRS)

Part 3. Tutorial for PRS of coronary artery disease

Outline of this workshop

Part 1. Genome-wide association studies (GWAS)

Part 2. Polygenic risk scores (PRS)

Part 3. Tutorial for PRS of coronary artery disease

Genetic variants

- What are genetic variants?

Genetic variants

- SNP

```
I WILL TAKE A CAR
I WILL TAKE A CAB
```

- Indels

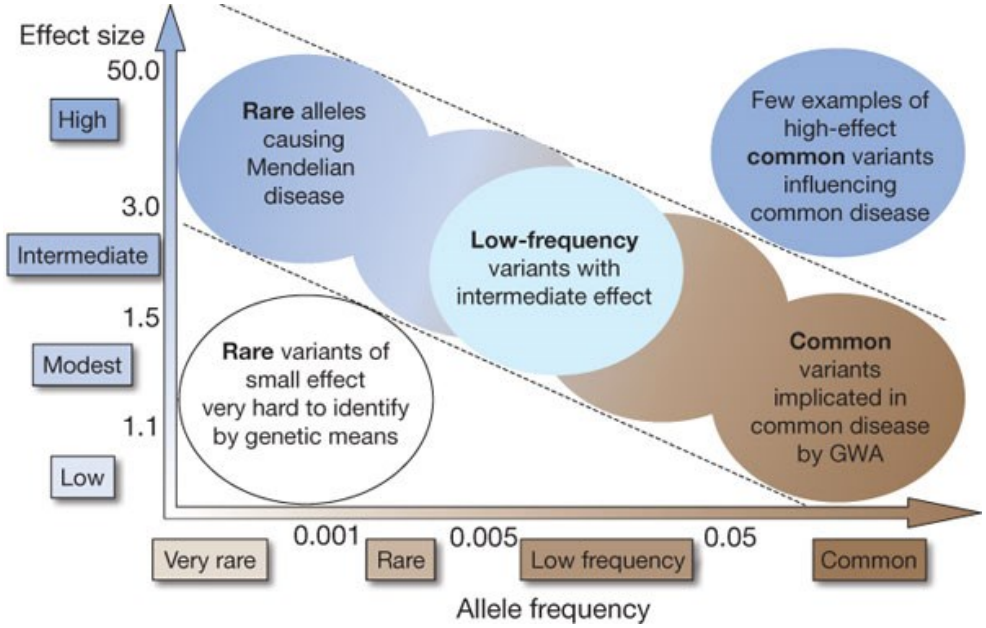
```
I WILL TAKE A CAR
I WILL TAKE A CARBONARA
```

- CNVs

```
I WILL TAKE A CAR
I WILL TAKE A CAR A CAR A CAR
```

Contribution of genetic variants

Not all variants affect phenotypes equally



Manolio et al. 2009. *Nature*

Mendelian diseases vs complex traits

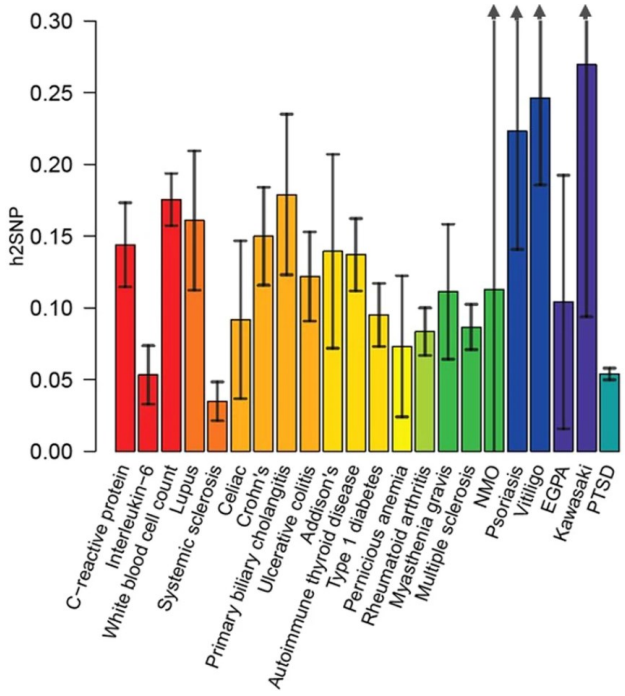
Mendelian traits	Complex traits
• Large effect size • Variants are often deleterious • Clear inheritance pattern	• Across many genes • Variants are often less deleterious • Small effect size • No distinct inheritance pattern

Mendelian diseases vs complex traits

Mendelian diseases	Complex diseases
• Cystic fibrosis • Phenylketonuria • Color blindness • Hemophilia	• Obesity • Type 2 diabetes • Type 1 diabetes • Multiple sclerosis • Coronary artery disease • Schizophrenia

What is SNP-based heritability?

- How much phenotypic variance can be explained by SNPs
- High SNP-based heritability: common variants play a big role in the genetic architecture of the trait



Maihofer et al. 2024. *Translational Psychiatry*

What is SNP-based heritability?

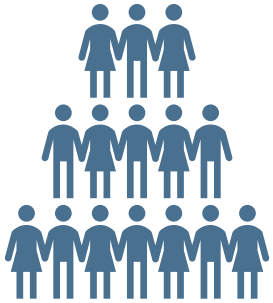
- SNP heritability does not capture variance explained due to:
  - Extremely rare variants
  - Structural variants
  - Epigenetic modifications
  - Gene-gene and gene-environment interactions

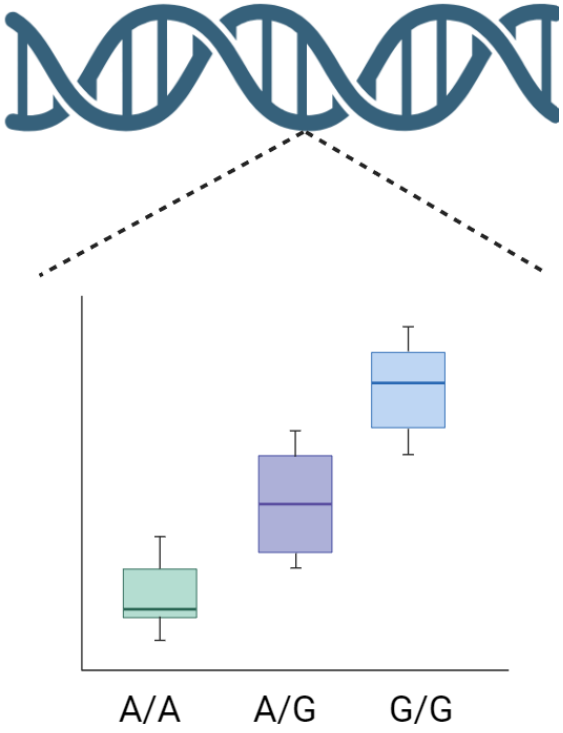
Capturing variants across the genome

Whole-genome sequencing	Genotyping
• DNA is fragmented, sequenced, and aligned • Annotate aligned reads to identify SNPs	• DNA hybridizes to microarray chips with probes to known SNPs • When DNA is hybridized (binds to probe), it generates a fluorescent signal • The intensity of the signals is processed • Variants are annotated against reference genome

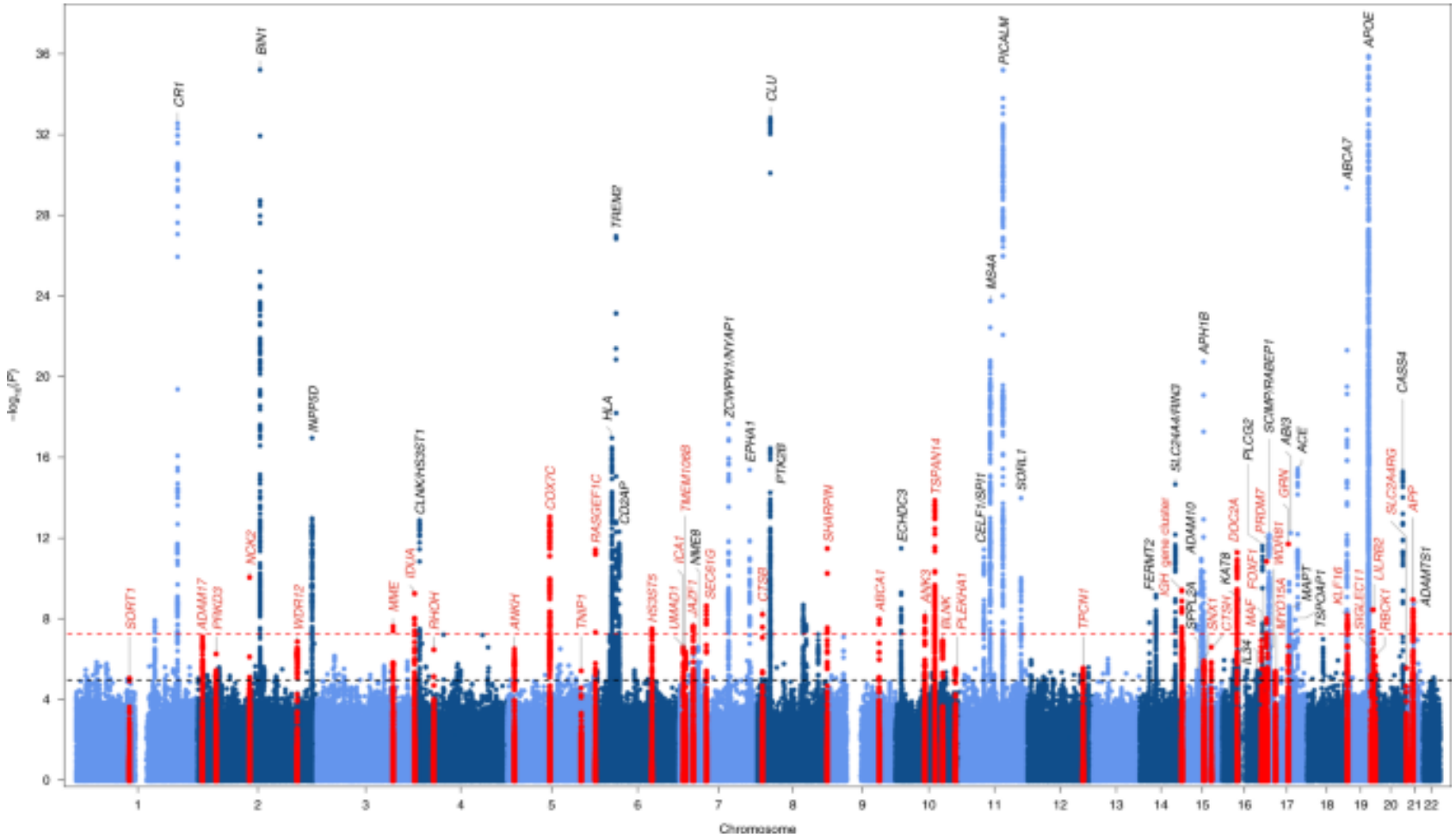
Genome-wide association studies

- Goal: test the effect of genetic variants on a trait/disease
- Continuous traits: linear regression
- Binary traits: logistic regression





Genome-wide association studies



Bellenguez et al. 2022. *Nature Genetics*

Genome-wide significance

- P < 0.05 cannot be used (multiple testing burden)
- Adjust for number of independent **LD blocks** across the entire genome
- There are roughly 1 million blocks of SNPs (**LD blocks**) independent of each other
- Apply Bonferroni P-value correction: P = 0.05 / 1,000,000 = **5x10e-8**

GWAS summary statistics

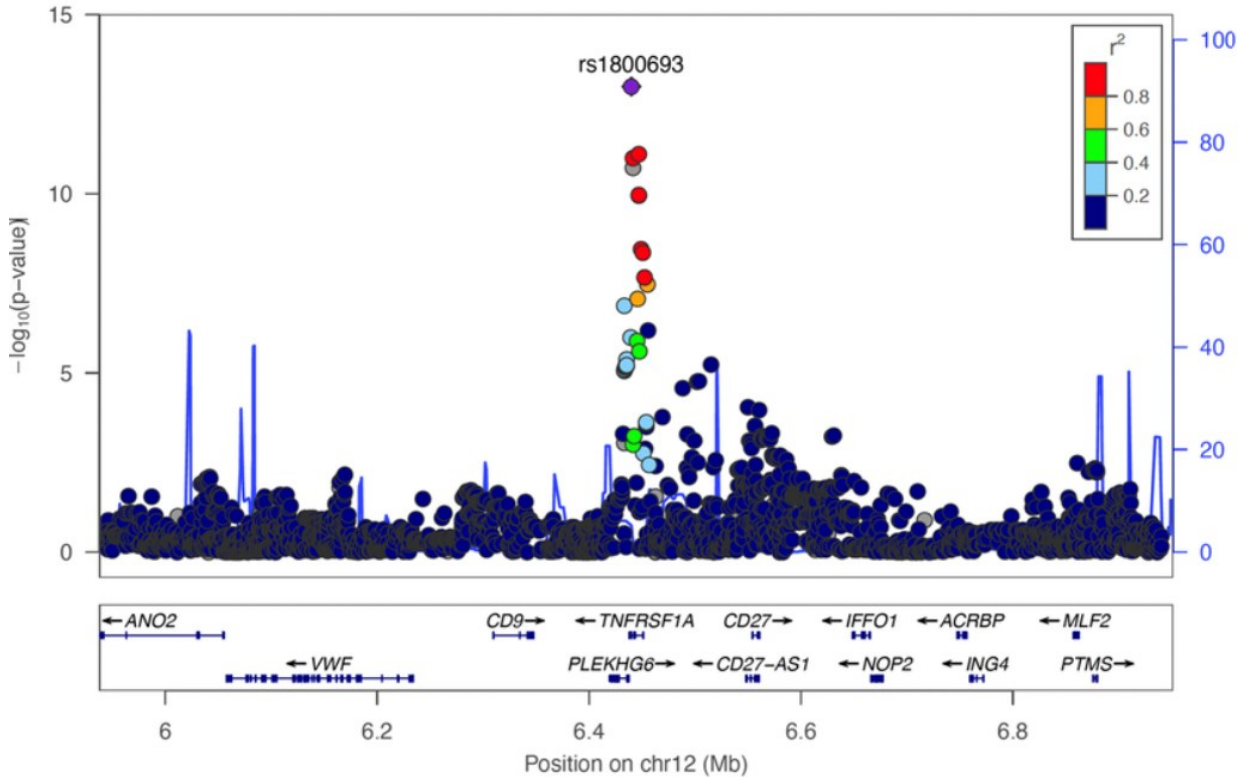
- Each SNP has an effect size and P-value
- For binary trait: odds ratio = e ^ beta
- Each copy of C allele on rs1800693 increases the odds of multiple sclerosis by about 14%.

Chr	BP (hg19)	Effect Allele	Other Allele	Beta	SE	P-value
12	6440009	C	T	0.127	0.0166	1.01E-13

IMSGC et al. 2019. *Science*

Pitfalls of GWAS (1)

- Pitfall: raw summary statistics association cannot be used directly
- True risk variant will “tag” other variants through LD

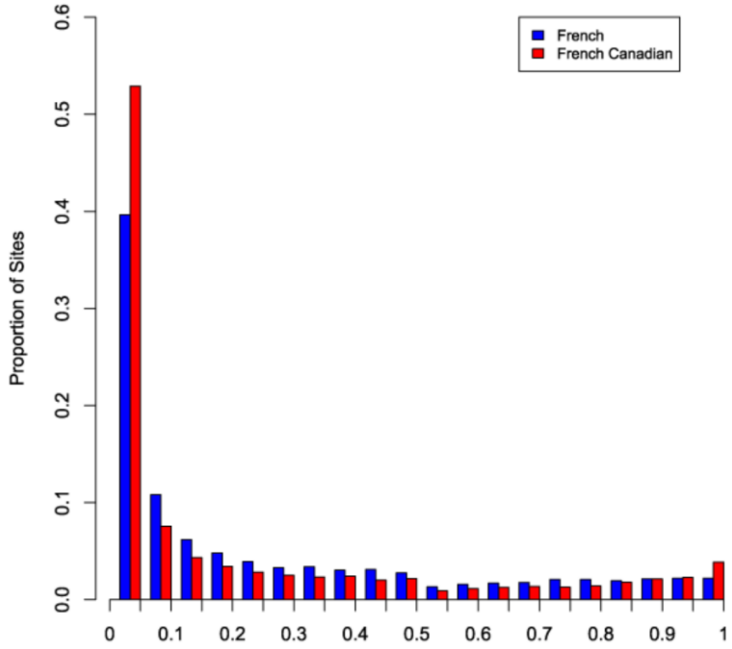


Kim & Patsopoulos et al. 2022. *Seminars in Immunopathology*

Pitfalls of GWAS (2)

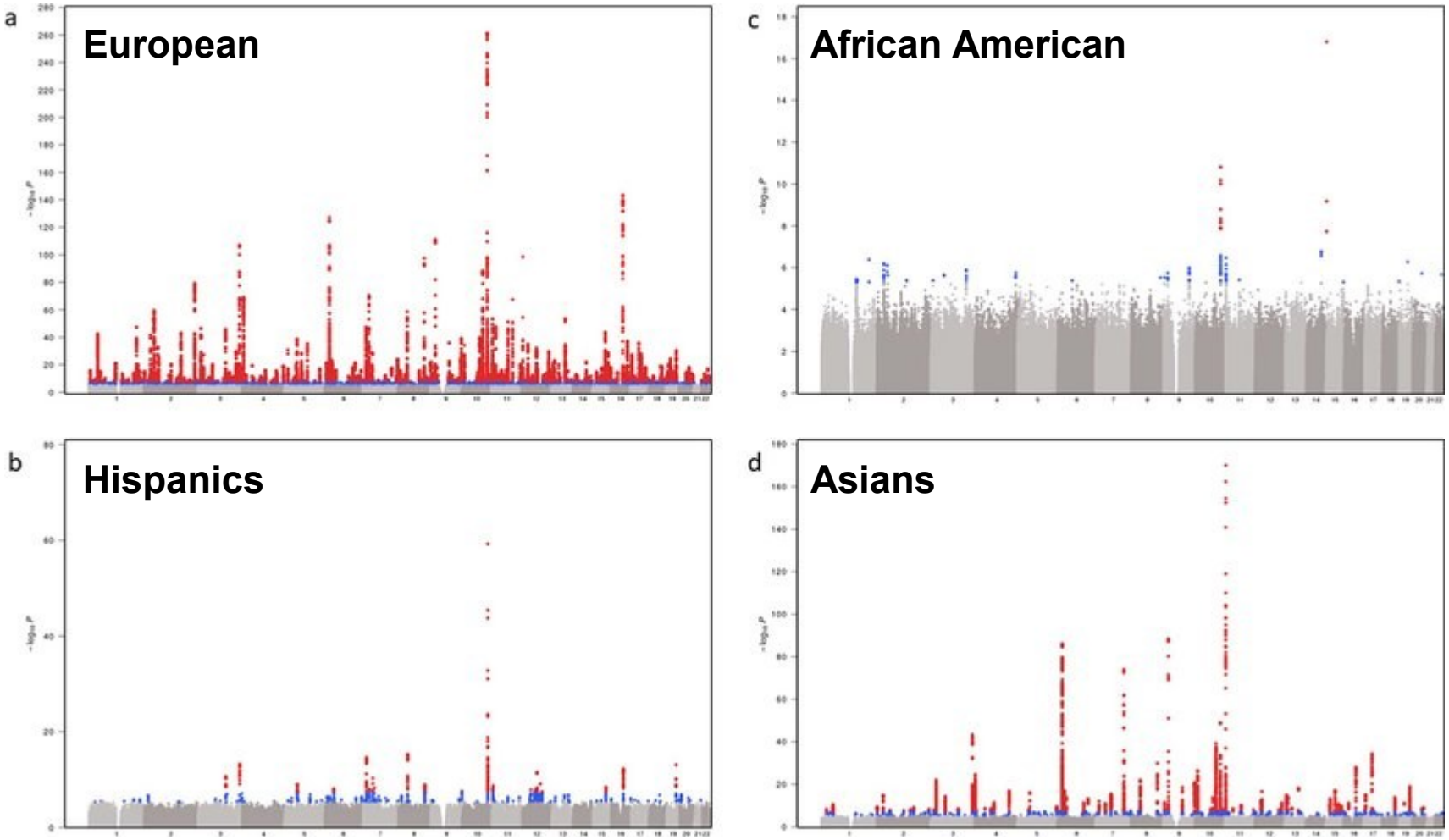
- Allele frequencies can vary between different ancestries

- Migrations, genetic drift, selection pressures:
  - Differences in allele frequencies
  - Differences in how genetic variants are inherited together (LD)



Casals*, Hodgkinson* et al. 2013. *PLOS Genetics*

Pitfalls of GWAS (2)

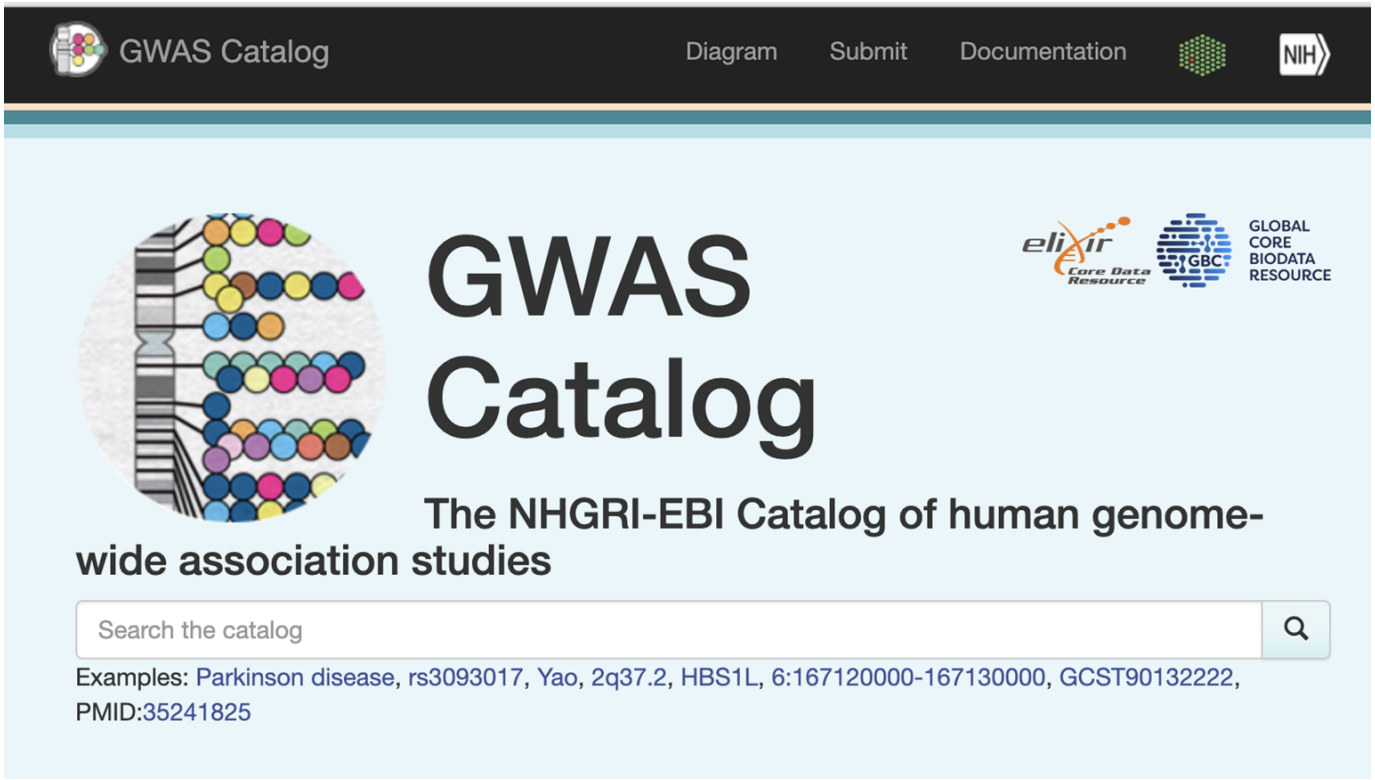


Vujkovic et al. 2020. *Nature Genetics*

Hands-on Activity: GWAS Sumstats

Where can I search for GWAS summary statistics?

Visit https://www.ebi.ac.uk/gwas/ to look through GWAS summary statistics.



Summary of GWAS

1. Genome-wide association studies identify SNPs that are associated to traits.

2. LD prevents direct use of GWAS for risk stratification and inference of causal gene

3. GWAS performed in European populations may not generalize to other ancestries.

4. Open sources (like GWAS catalog) contain summary statistics for GWAS.

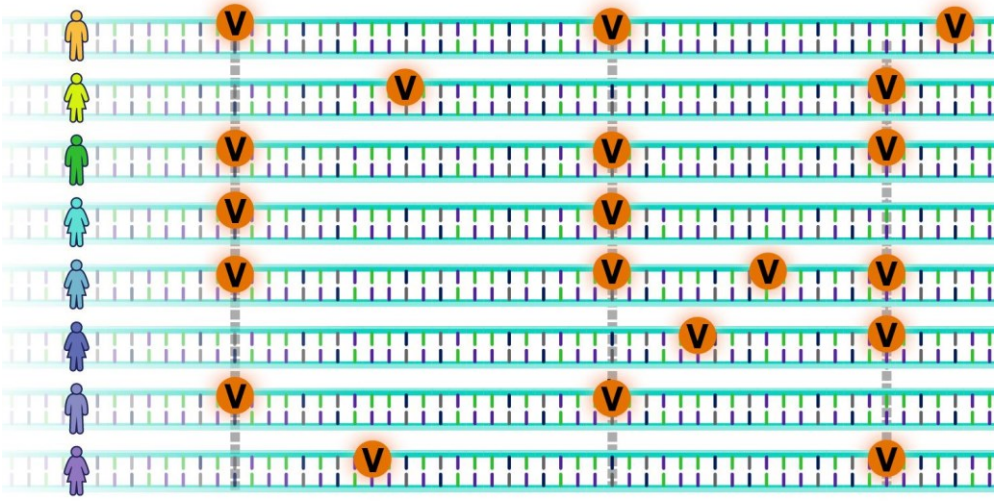
Outline of this workshop

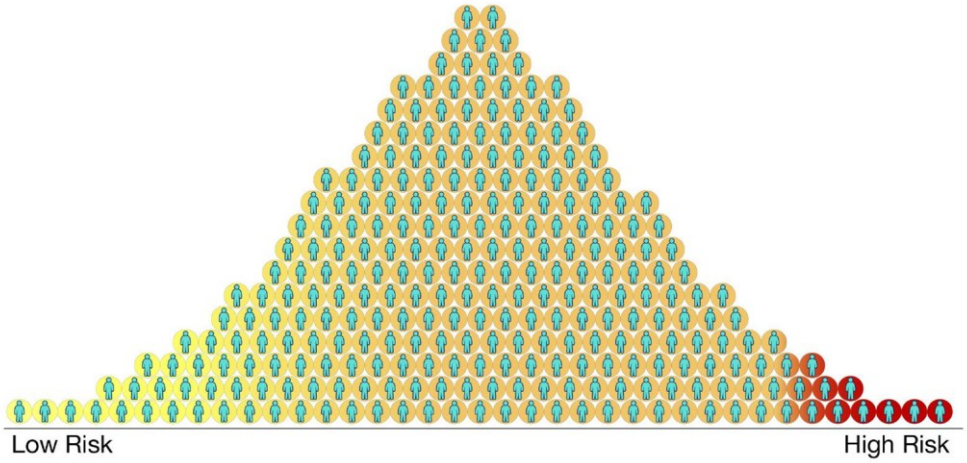
Part 1. Genome-wide association studies (GWAS)

Part 2. Polygenic risk scores (PRS)

Part 3. Tutorial for PRS of coronary artery disease

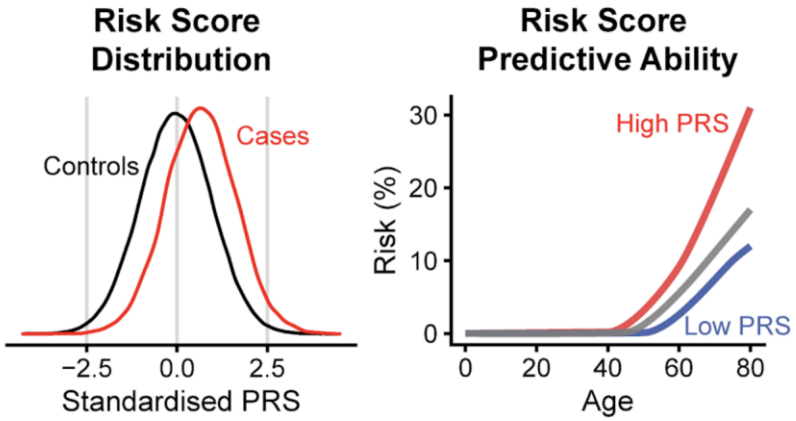
Polygenic risk scores





Polygenic risk scores

Goal: early screening, stratification of high-risk individuals



Polygenic risk scores

Target data
Individual-level genotype

Process genetic data to keep high-quality common SNPs.

Ancestry/population stratification check.

Training data
GWAS summary statistics

Samples from GWAS ≠ sample from target data

Matched ancestry with target data.

Well-powered, large sample size.

Linkage disequilibrium (LD) reference panel

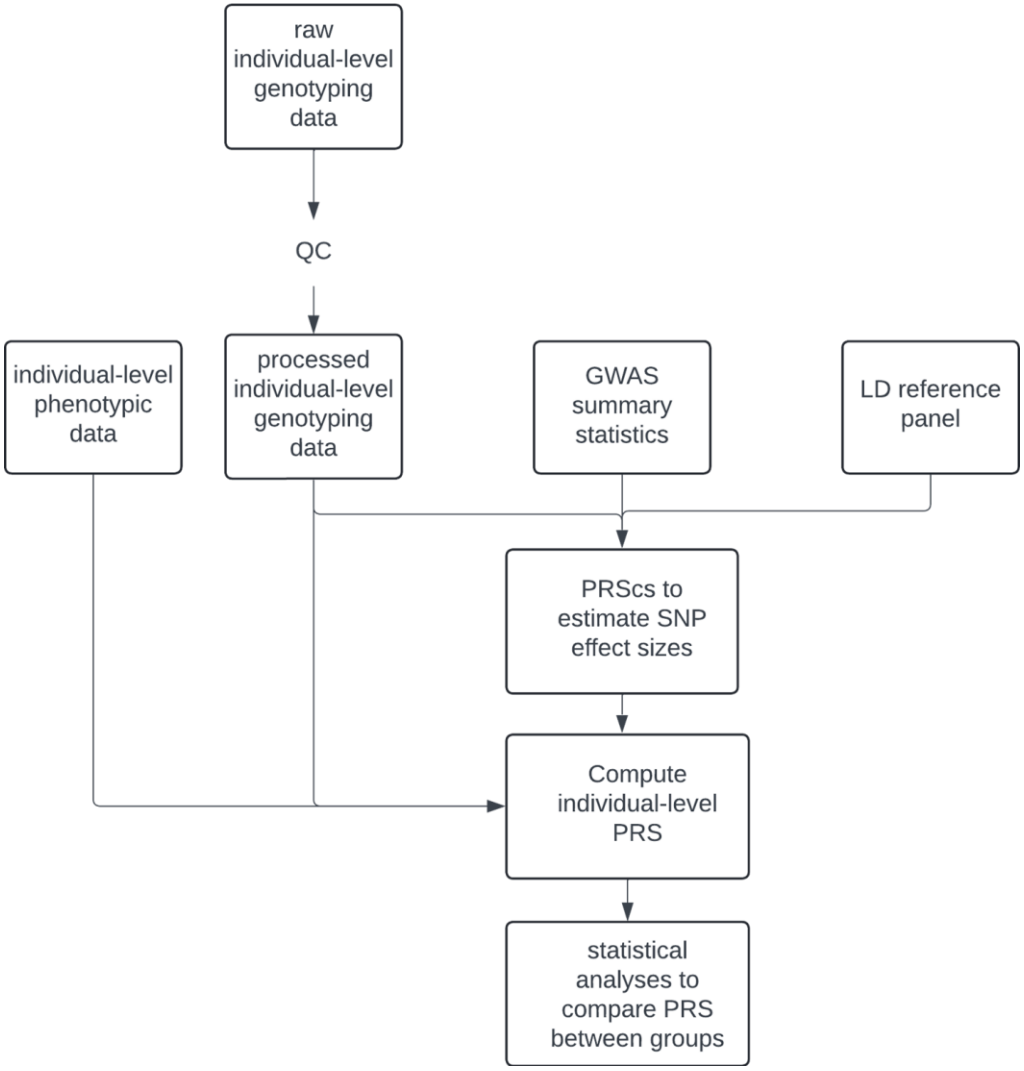
Correlation between SNPs.

Ancestry of reference panel = ancestry of target data = ancestry of base data

PRS strategies

Clumping and thresholding	**LD prediction**	**Bayesian estimation of SNP effect sizes**
• *Ex.: PRSice* • Uses independent SNPs that are significant at a p-value threshold	• *Ex.: LD pred* • Predict effect sizes of SNPs based on LD correlation • Needs an LD matrix (correlation between all SNPs)	• *Ex.: PRScs* • LD reference panel to adjust for the correlation between SNPs • Adjust SNP effect sizes before computing PRS

PRS flowchart



PRS flowchart

- Removes SNPs that are missing in more than a certain proportion of samples
- Remove SNPs that are occur at extremely low frequencies
- Remove SNPs that violate Hardy-Weinberg equilibrium
- Out of all SNPs that are correlated, select only one (LD pruning)
- Remove individuals who have more than a certain portion of missing genotype

Outline of this workshop

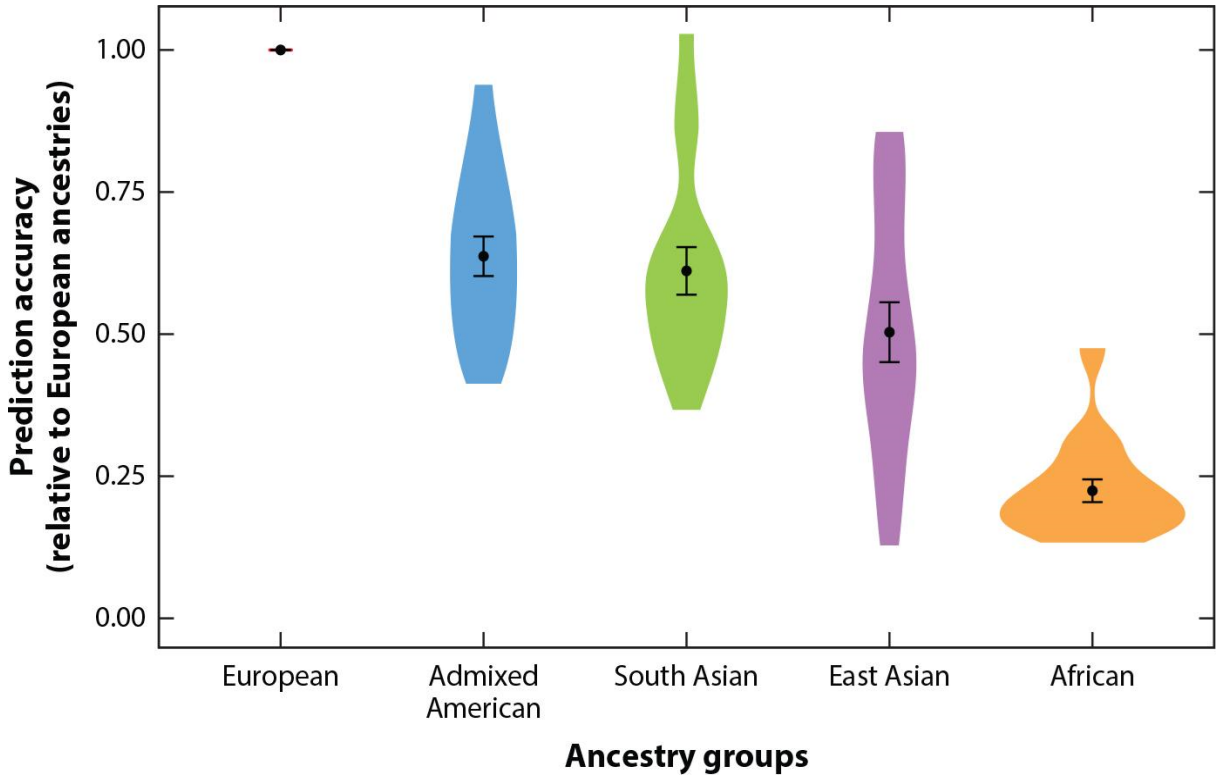
Part 1. Genome-wide association studies (GWAS)

Part 2. Polygenic risk scores (PRS)

Part 3. Tutorial for PRS of coronary artery disease

Pitfalls of PRS

- Limited clinical utility for non-European populations
- Exacerbate health disparities



Wang et al. 2022. *Annual Review of Biomedical Data Science*

Summary of PRS

1. PRS: estimated genetic susceptibility for a disease

2. PRS weights are typically derived from GWAS summary statistics

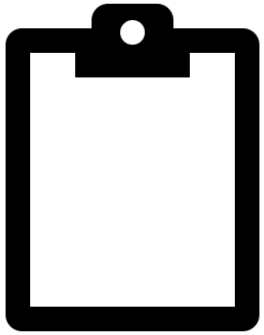
3. PRS trained on one ancestry does not perform well on another

Thank you for attending!

1

Scan the QR code to confirm you attended today's workshop.

2



Fill out the feedback survey in the next 72h.

3

Get recognition for this workshop on your co-curricular record.