

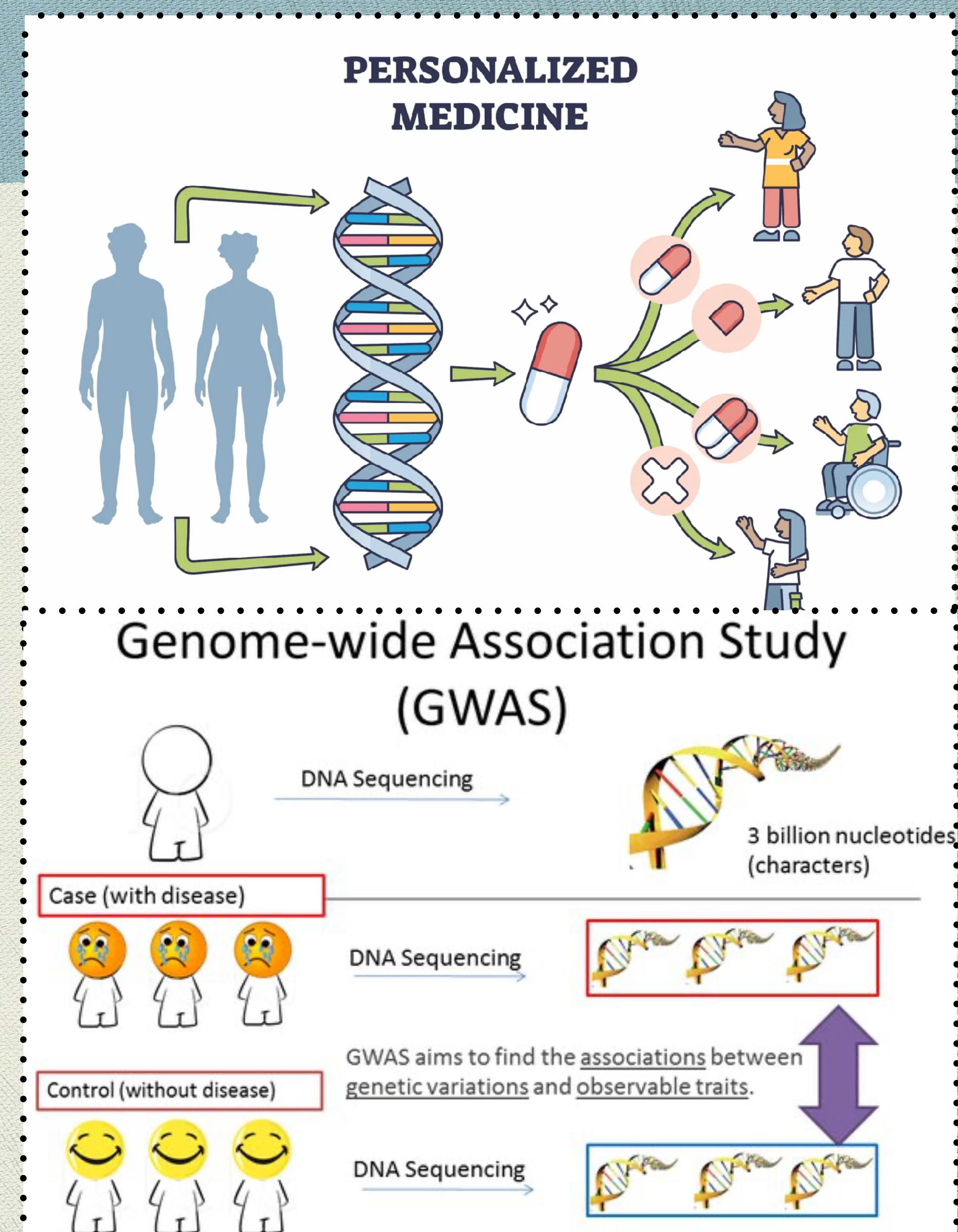


# Polygenic Risk Score Analysis

Bharath Kumar Swargam  
MS Data Science  
UMBC

# Background

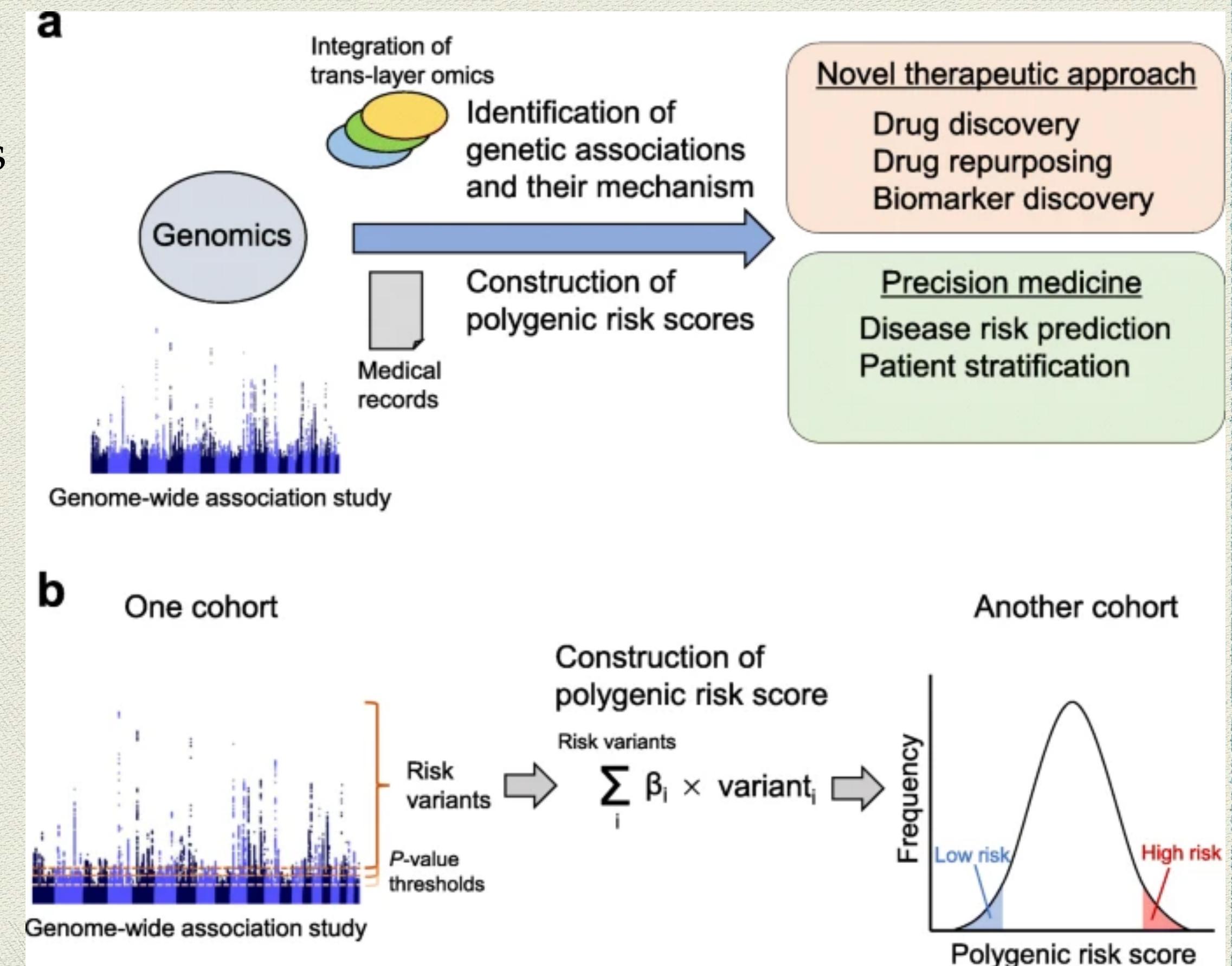
- Prediction of disease risks/ trait is essential part of personalized medicine
- Common diseases are influenced by multiple genetic variants
- Traditional genetic studies focuses on identifying mutations/ genes like Sodium voltage-gated channel alpha subunit 9 gene (SCN9A) for pain
- Genome-Wide Association Study (GWAS) - Single Nucleotide Polymorphisms (SNPs) associated with the disease/ trait
- Polygenic Risk Score aggregates small effects from many variants to better capture genetic influences on complex traits



# What is Polygenic Risk Score (PRS)?

3

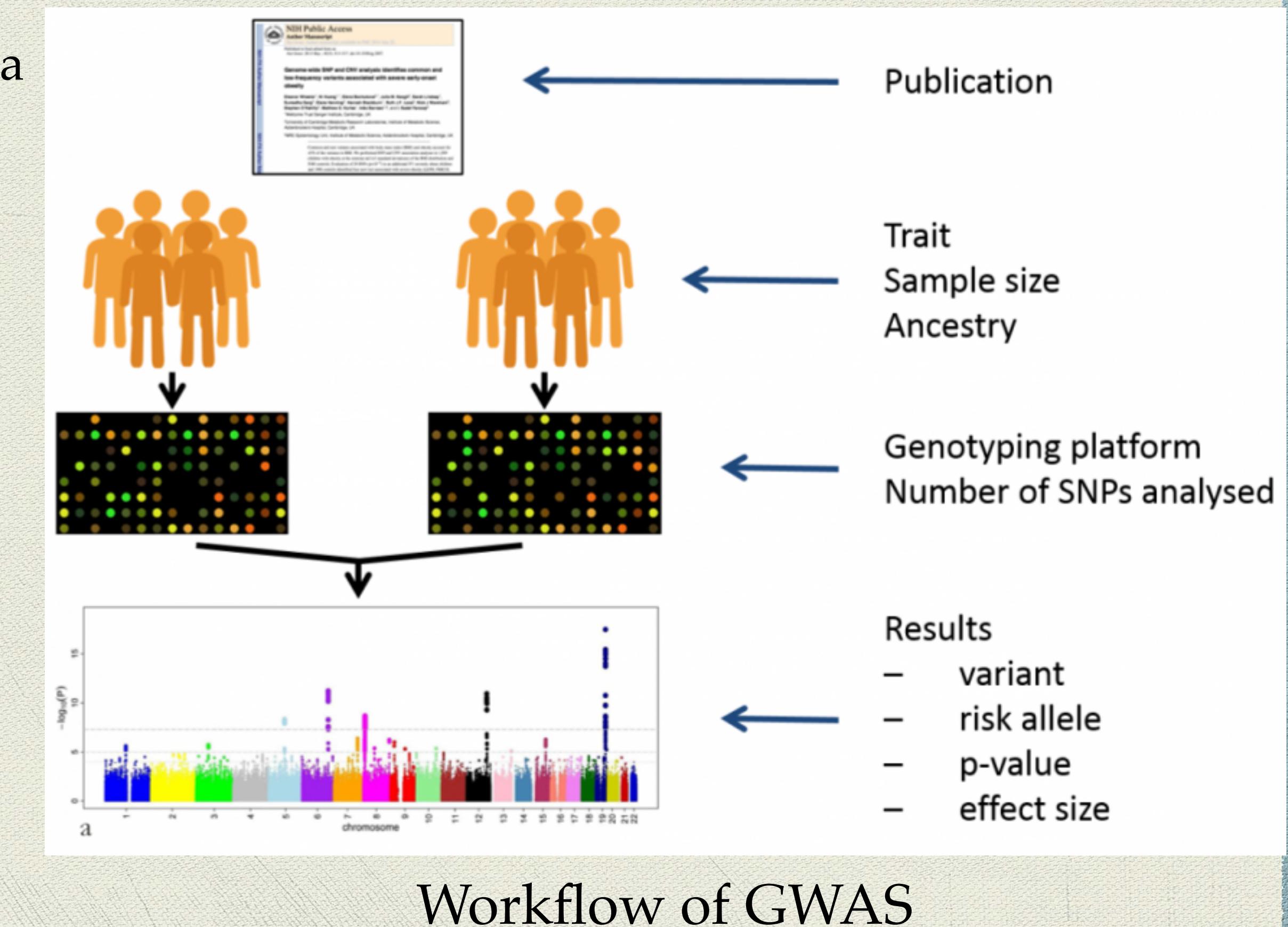
- ◆ A polygenic risk score (PRS) is the sum of an individual's effect alleles, weighted by their effect sizes derived from GWAS data
- ◆ It is a single value estimate of an individual's genetic liability to a trait or disease
- ◆ It is sensitive to ethnic background
- ◆ Unclear which method is best among the available, categorized into three approaches



# Basic Terminology

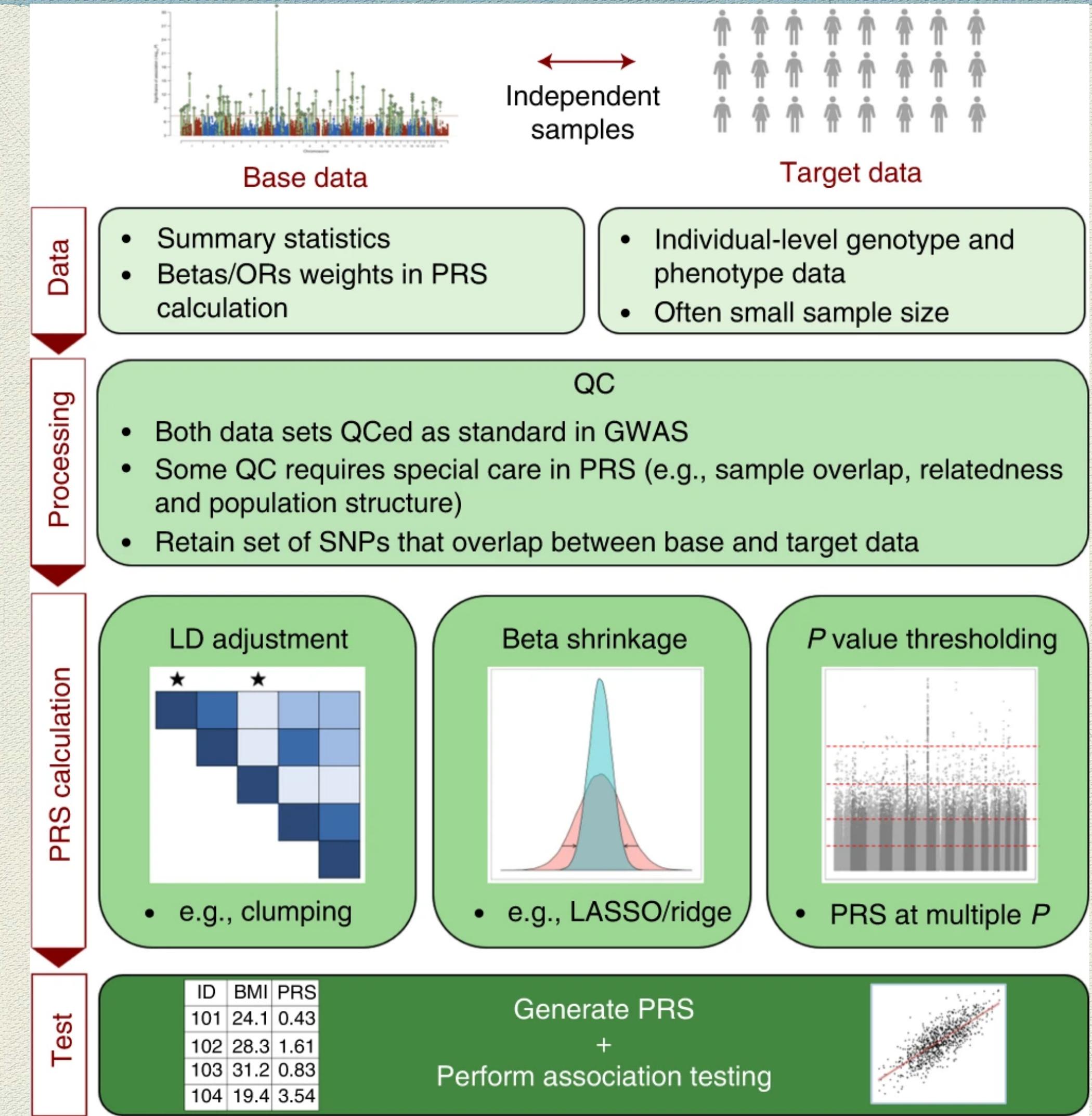
4

- ◆ Summary statistics: results obtained after conducting a GWAS
- ◆ Variant: the Single Nucleotide Polymorphism
- ◆ Risk Allele : Allele of a SNP that increases the association/ risk of a trait/ disease
- ◆ P-value: Probability that the observed association between the variant and trait happened by chance
- ◆ Effect Size: A measure of the strength of the relationship between the genetic variant and a trait



# Workflow for Polygenic Risk Score Calculation

5



# Basic example of calculating PRS

6

## 1. Discovery GWAS summary statistics

	SNP 1	SNP 2	SNP 3	SNP 4
Effect allele	C	A	C	T
Weight	0.2	-0.3	0.1	0.2

## 2. Target sample genotypes

IID	SNP 1	SNP 2	SNP 3	SNP 4
1	CT	AA	CA	TG
2	CT	AA	CA	GG
3	TT	TT	CC	TT
4	CC	AT	AA	TG

## 3. Polygenic score

IID	SNP 1		SNP 2		SNP 3		SNP 4		PGS
1	1*0.2	+	2*-0.3	+	1*0.1	+	1*0.2	=	-0.1
2	1*0.2	+	2*-0.3	+	1*0.1	+	0*0.2	=	-0.3
3	0*0.2	+	0*-0.3	+	2*0.1	+	2*0.2	=	0.6
4	2*0.2	+	1*-0.3	+	0*0.1	+	1*0.2	=	0.3

# Data

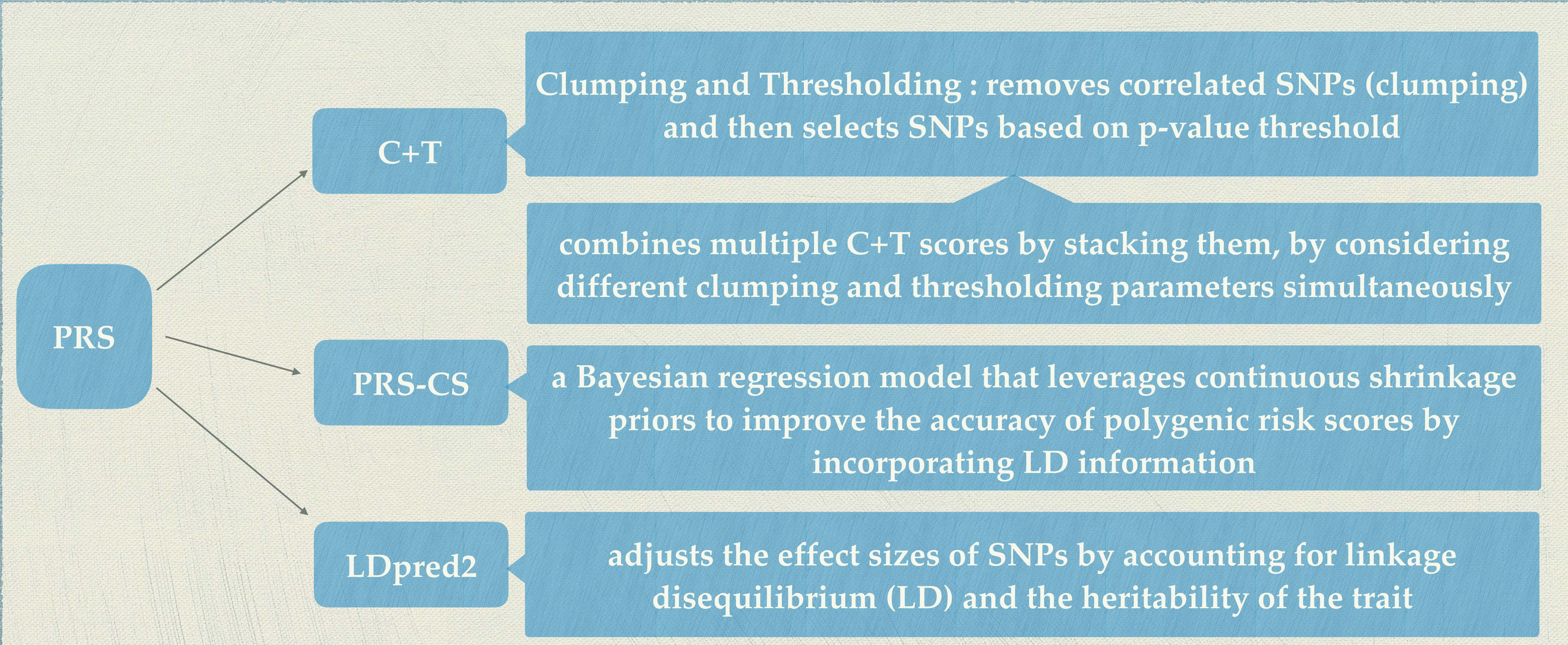
Trait Number	Trait
1	Chest pain or discomfort
2	Back pain for 3+ months
3	Headaches for 3+ months
4	Leg pain on walking
5	Medication for pain relief, constipation, heartburn: Aspirin
6	Ibuprofen (e.g. Nurofen)
7	Paracetamol
8	Omeprazole (e.g. Zanprol)
9	Medication any
10	Pain type(s) experienced in last month: Headache
11	Neck or shoulder pain
12	Back pain
13	Stomach or abdominal pain
14	Hip pain
15	Knee pain
16	Pain all over the body
17	Pain any

## Target Data

- ◆ 140 Individuals
- ◆ Phenotype and genotype ( binary PLINK files)
- ◆ Pain - target phenotype
- ◆ Sex - covariate
- ◆ Data is collected across the knee surgery
- ◆ European Ancestry

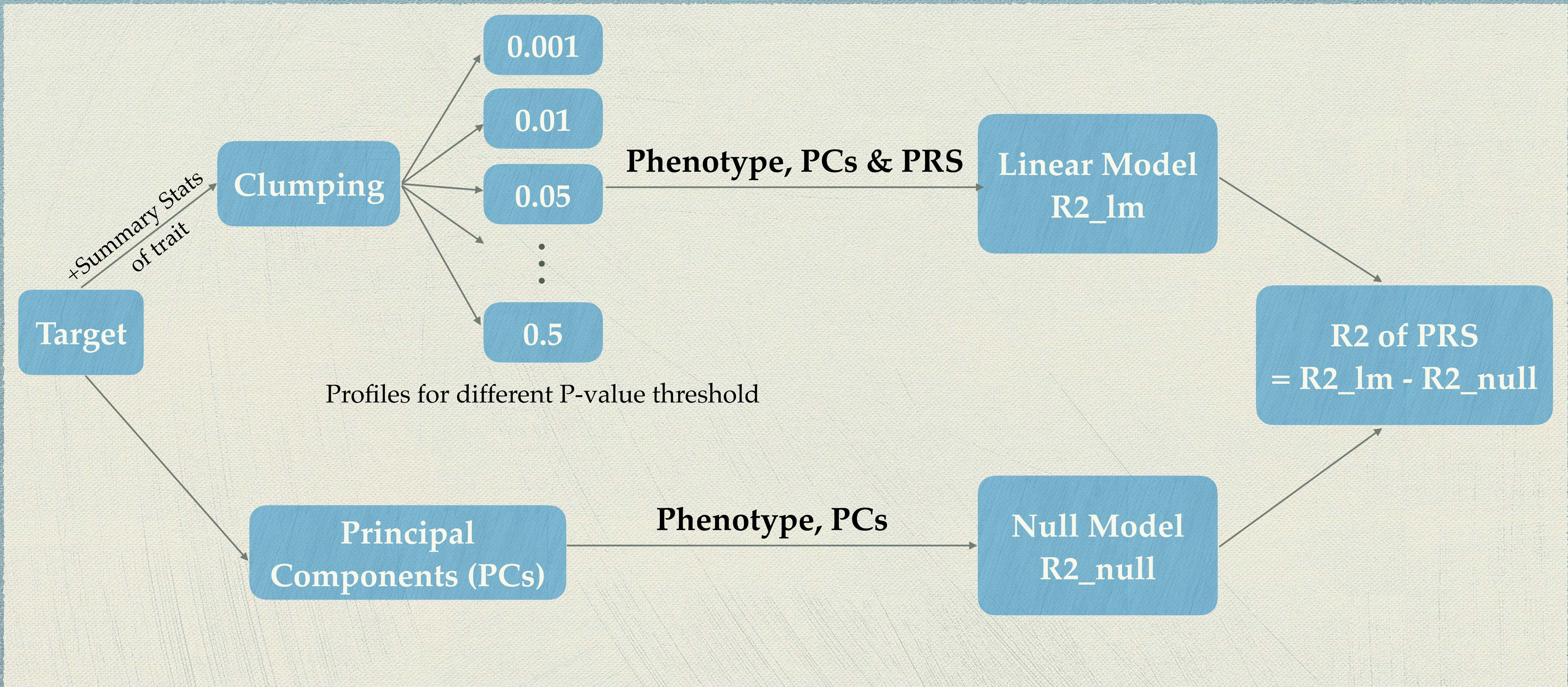
# Methods of obtaining PRS

8



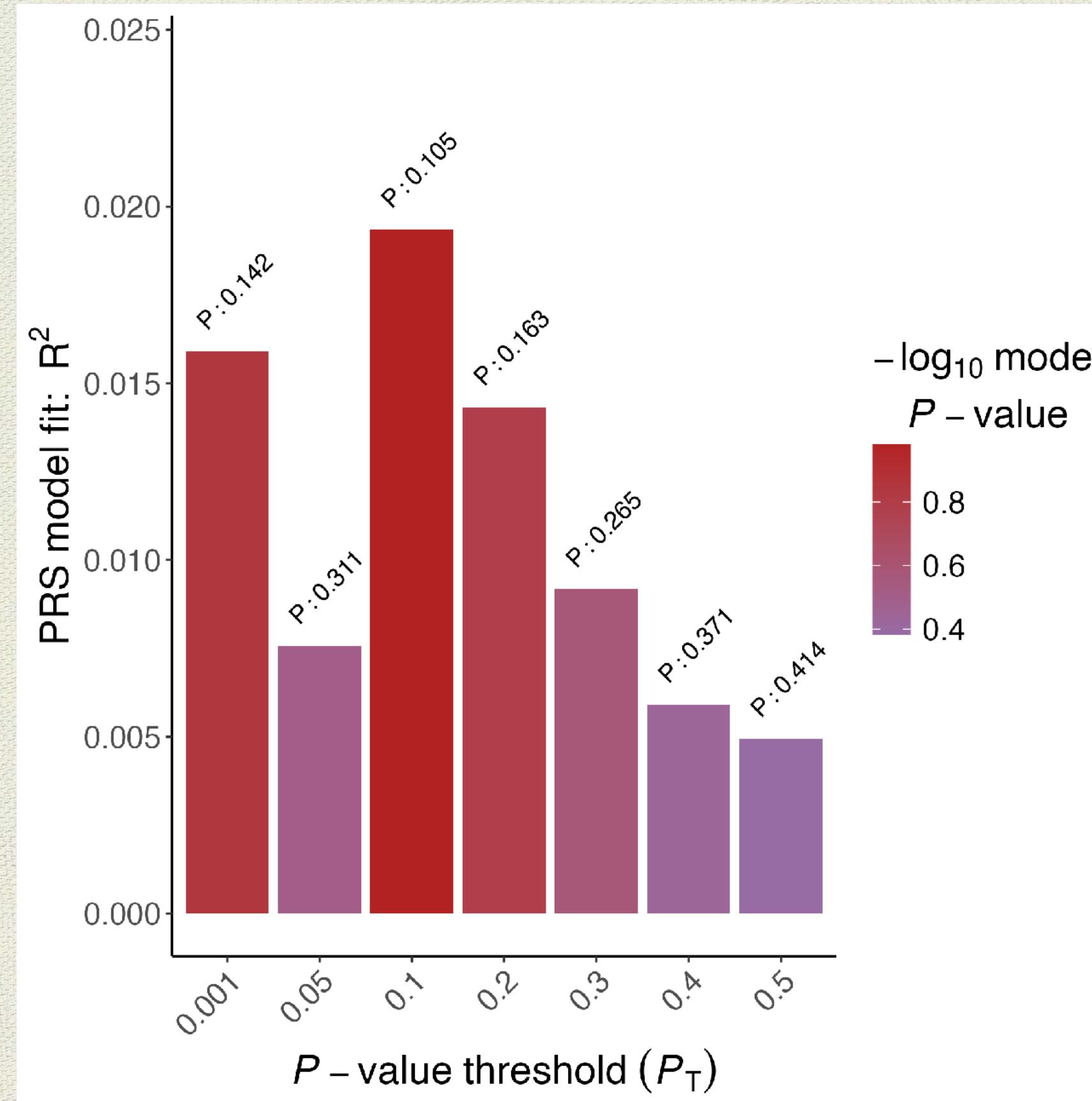
# C+T method

9

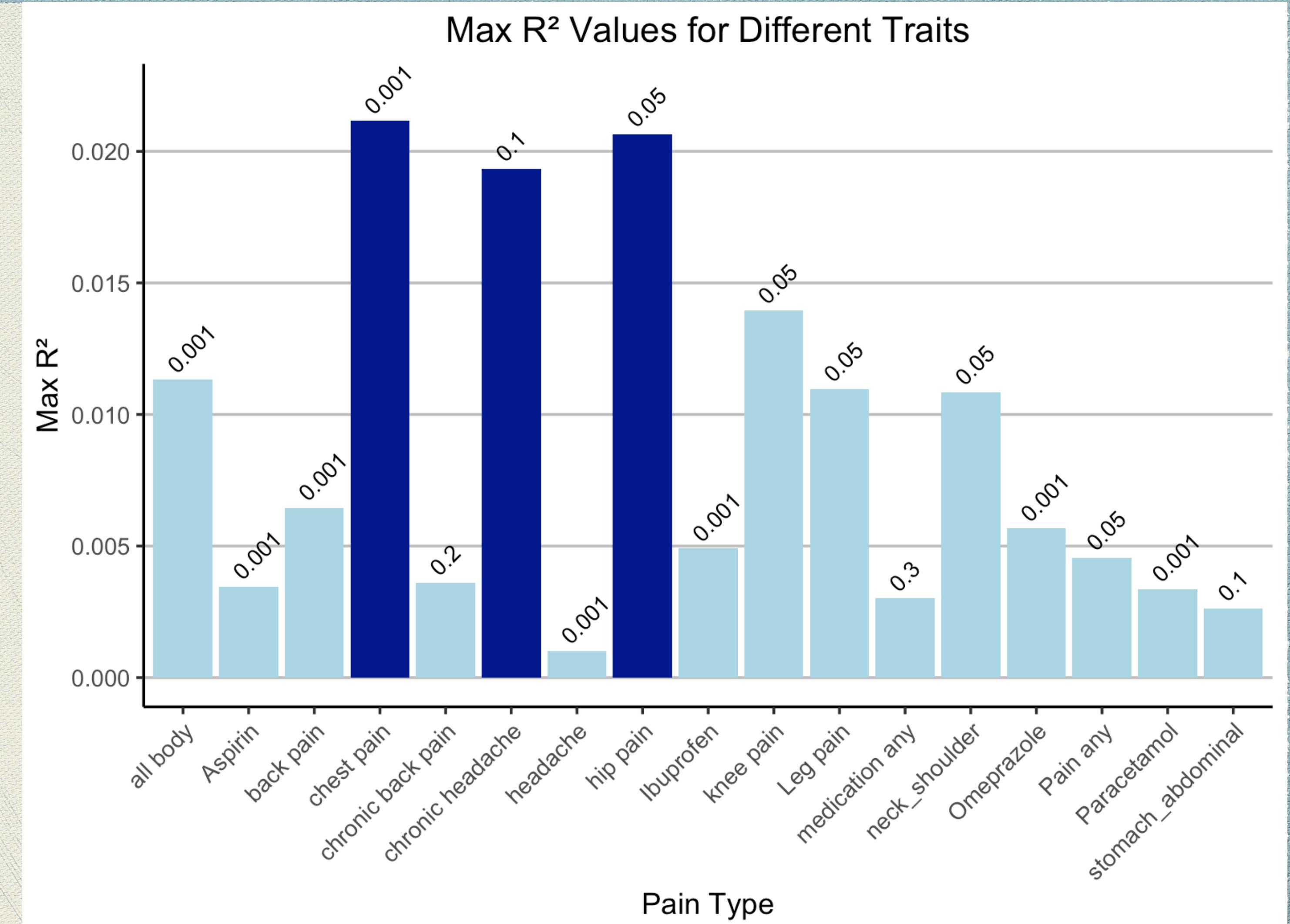


# Results

10

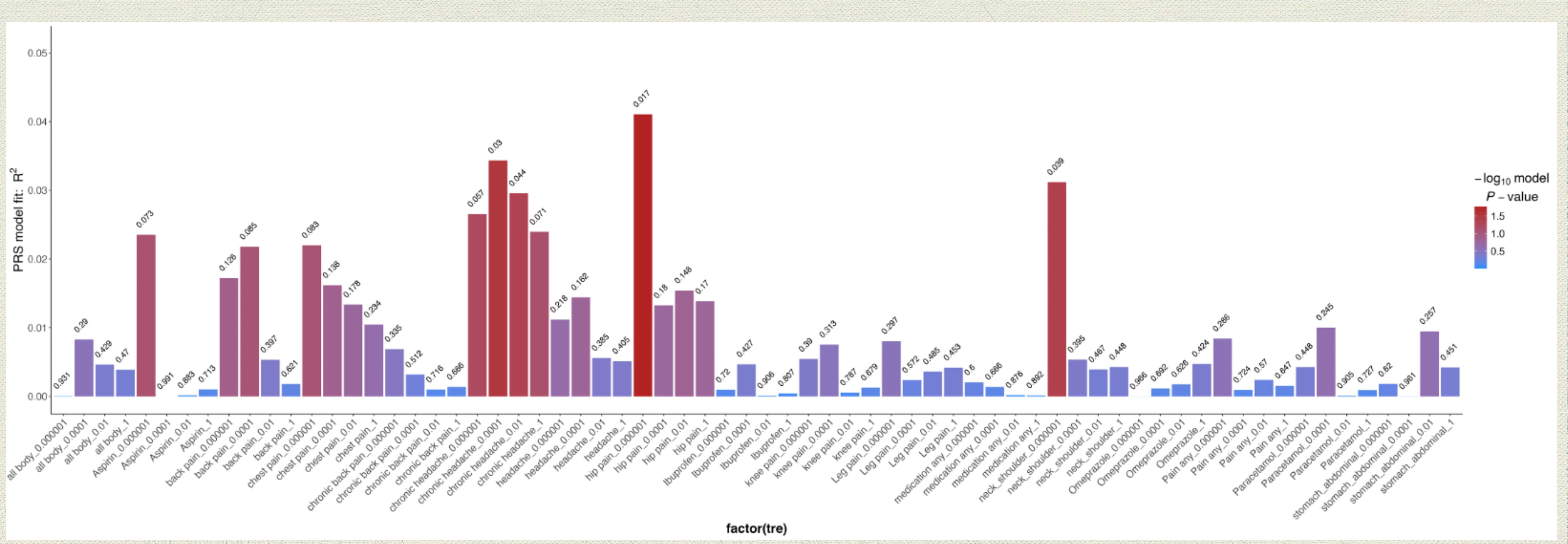


Bar plot of R<sup>2</sup> Vs P-value for chronic headache



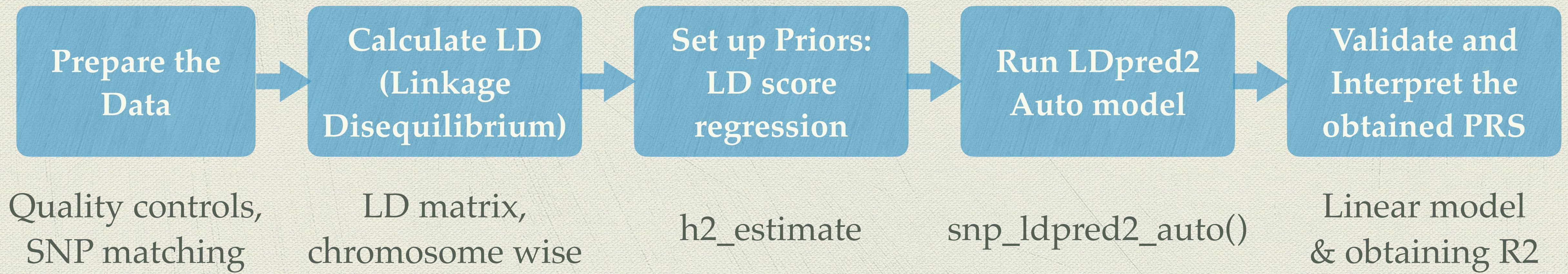
# PRS-CS Results

11



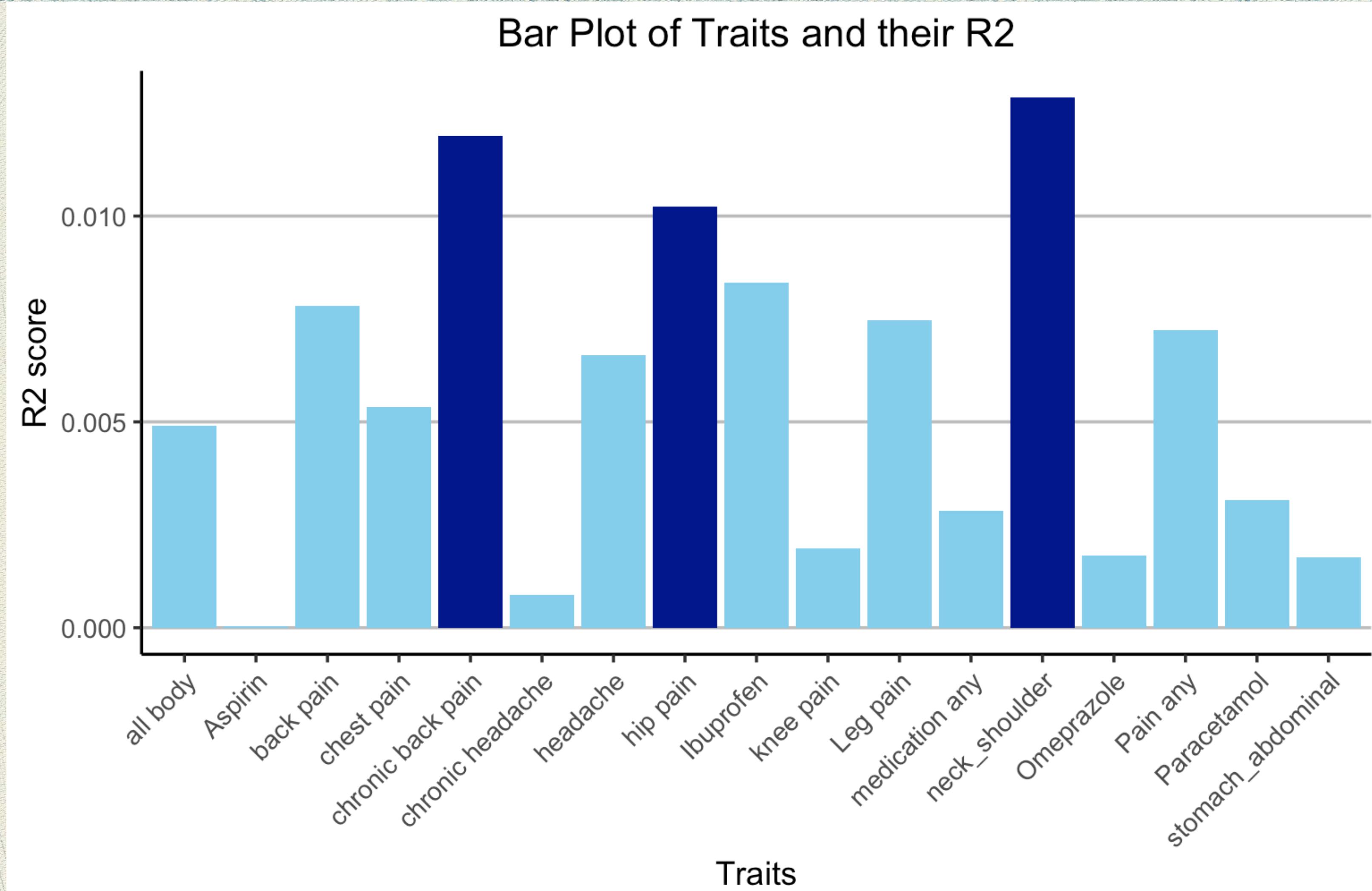
# LDpred2 workflow

12



# Results of LDpred2

13



# Work to be done

14

- ◆ Same trend with the three different approaches
- ◆ Combine all the Polygenic Risk Scores to build a single model
- ◆ Develop a single model for multiple traits instead of different models for each trait : Multi-trait PRS
- ◆ Calculate Polygenic Risk Score with different target phenotypes and addition of other predictors
- ◆ Focus on traits with higher heritability, and more correlated with the target phenotype

# Learnings

15

Genome-Wide  
Association Studies

Polygenic Risk  
Scores



- ◆ How to work with data in healthcare?
- ◆ Challenging part : Quality control steps, dealing with data

# Acknowledgements

16



- ◆ Evelina Mocci
- ◆ Dr. Seth Ament
- ◆ Ament Lab members
- ◆ Dr. Michelle Giglio



Thank you!

Questions?

Any Feedback?