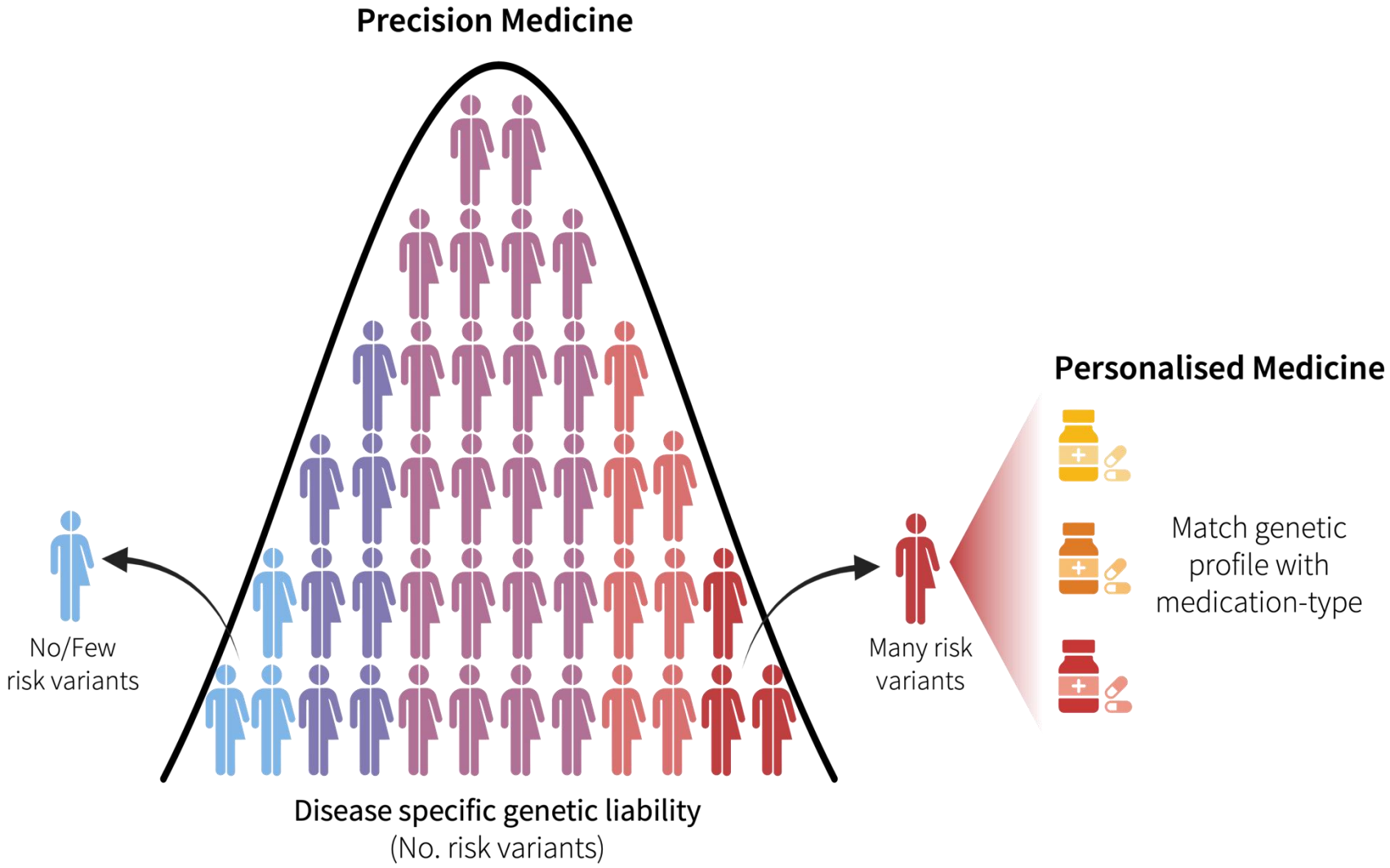
# Polygenic Scores (PGS)

#7



# LETS GET STARTED

# AGENDA

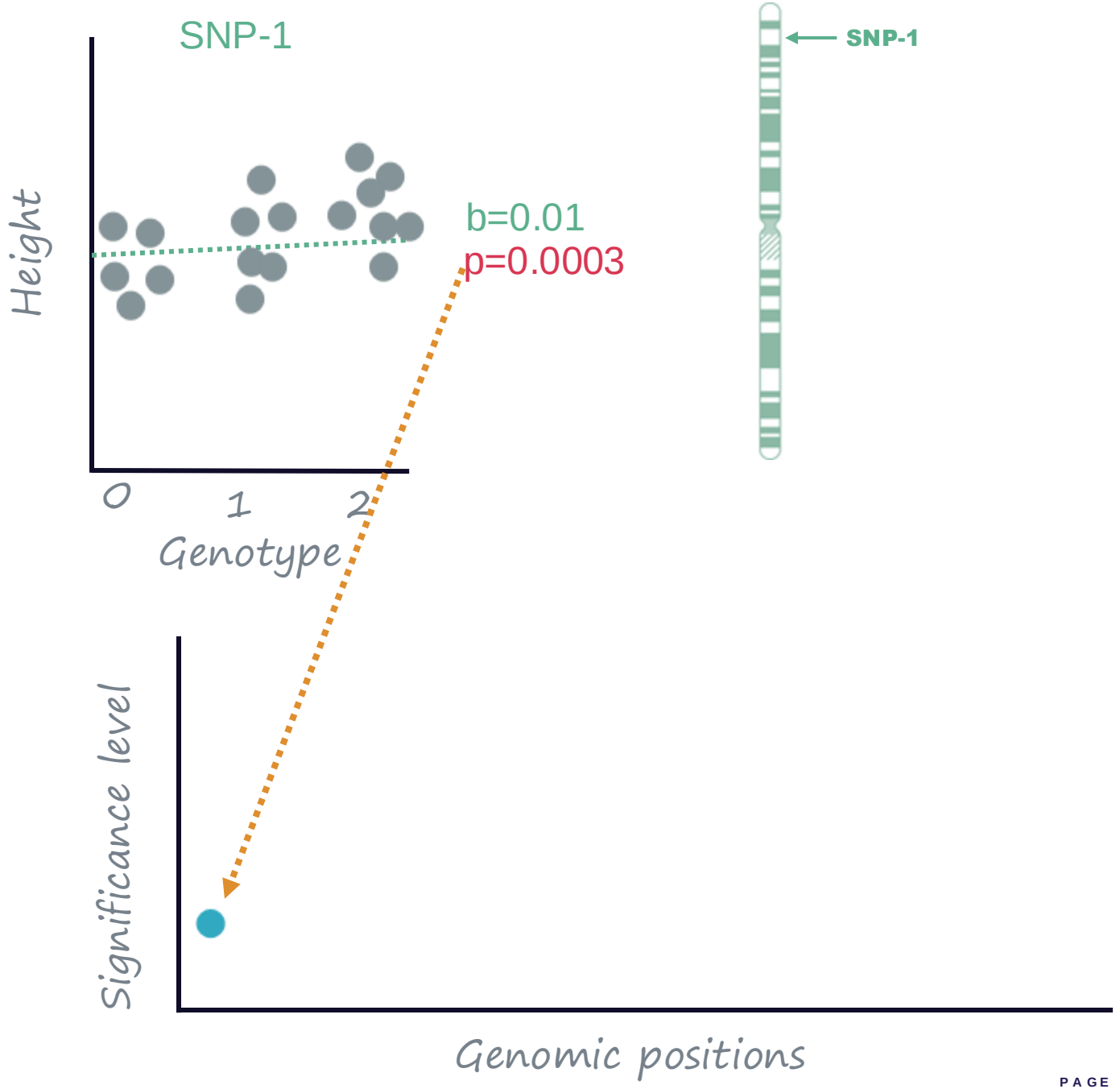
| | |
|---|---|
| **08:15 – 08:45** | Recap [*GWAS + R exercise from last*] |
| **08:45 – 08:50** | Break |
| **08:50 – 09:20** | Lecture [*PGS*] |
| **09:20 – 09:30** | Break |
| **09:30 – 10:30** | Exercise 1 + 2 [+ *joint discussion* ] |
| **10:30 – 10:40** | Break |
| **10:40 – 11:30** | Group work |
| **11:30 – 11:55** | Presentation of group work |
| **11:55 – 12:00** | Evaluation at Moodle |

# AGENDA

| | |
|---|---|
| **08:15 – 08:45** | Recap [*GWAS + R exercise from last*] |
| **08:45 – 08:50** | Break |
| **08:50 – 09:20** | Lecture [*PGS*] |
| **09:20 – 09:30** | Break |
| **09:30 – 10:30** | Exercise 1 + 2 [*+ joint discussion* ] |
| **10:30 – 10:40** | Break |
| **10:40 – 11:30** | Group work |
| **11:30 – 11:55** | Presentation of group work |
| **11:55 – 12:00** | Evaluation at Moodle |

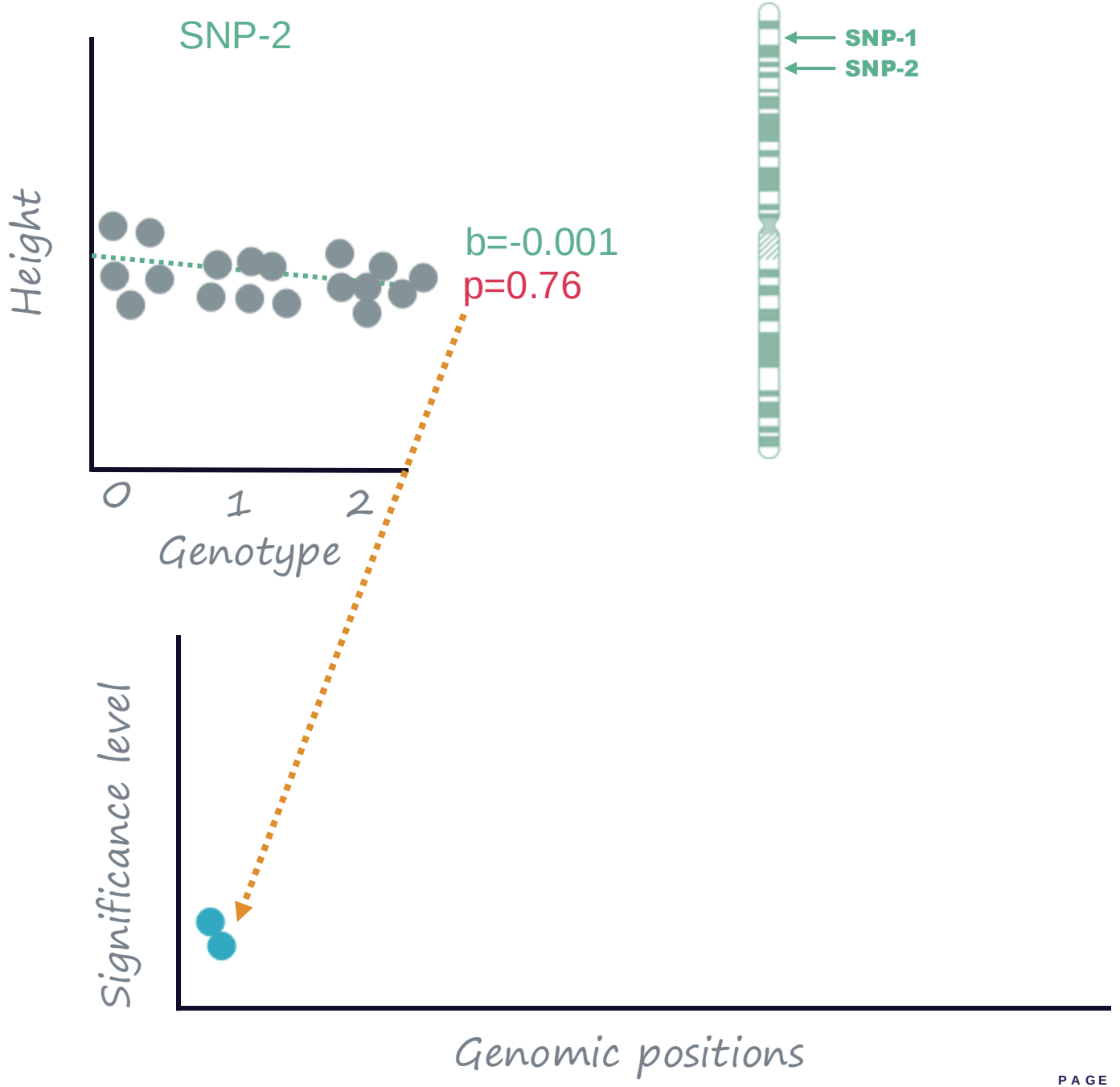
# GWAS RECAP

**Which SNPs associate with height?**



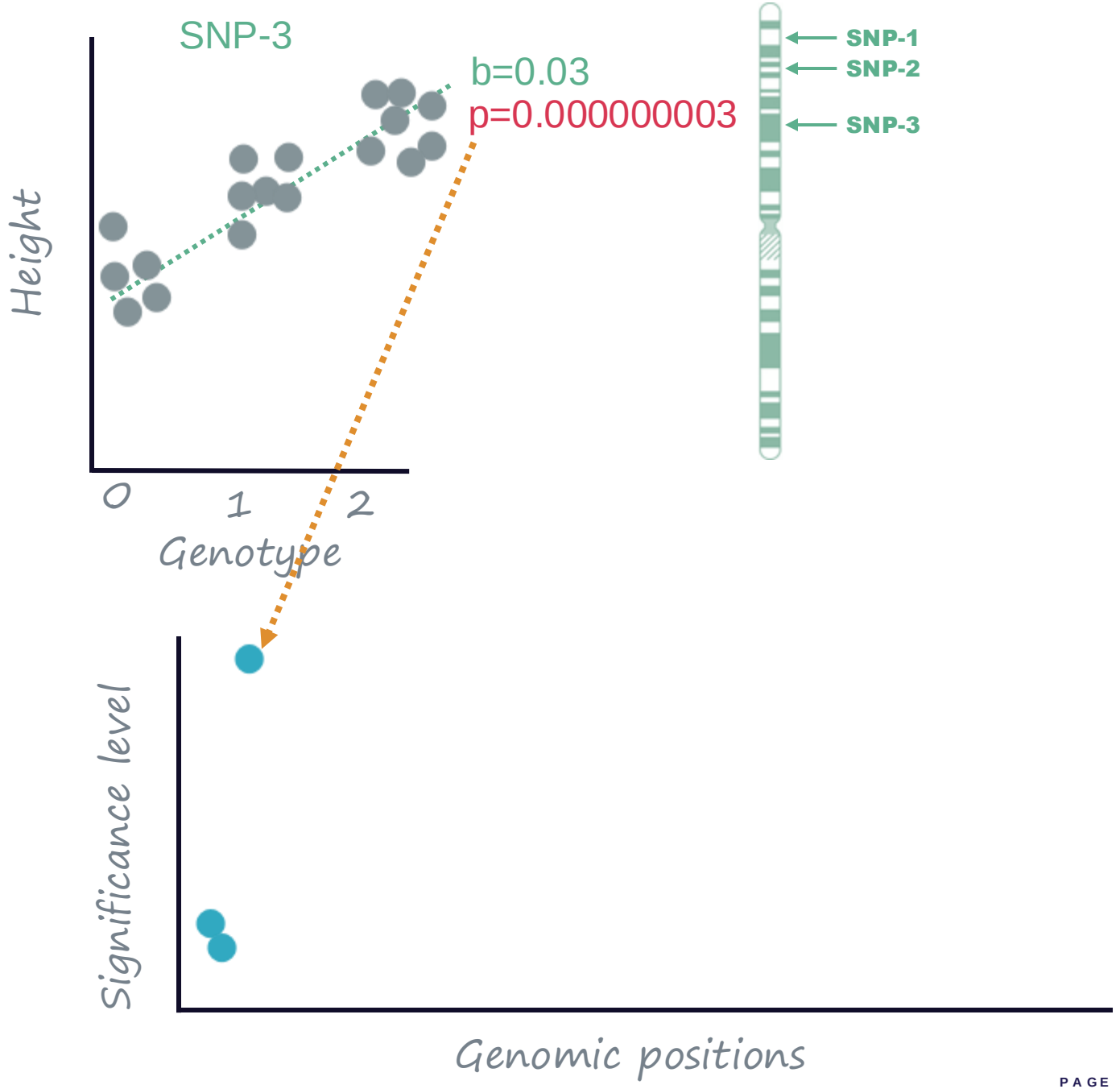
# GWAS RECAP

**Which SNPs associate with height?**



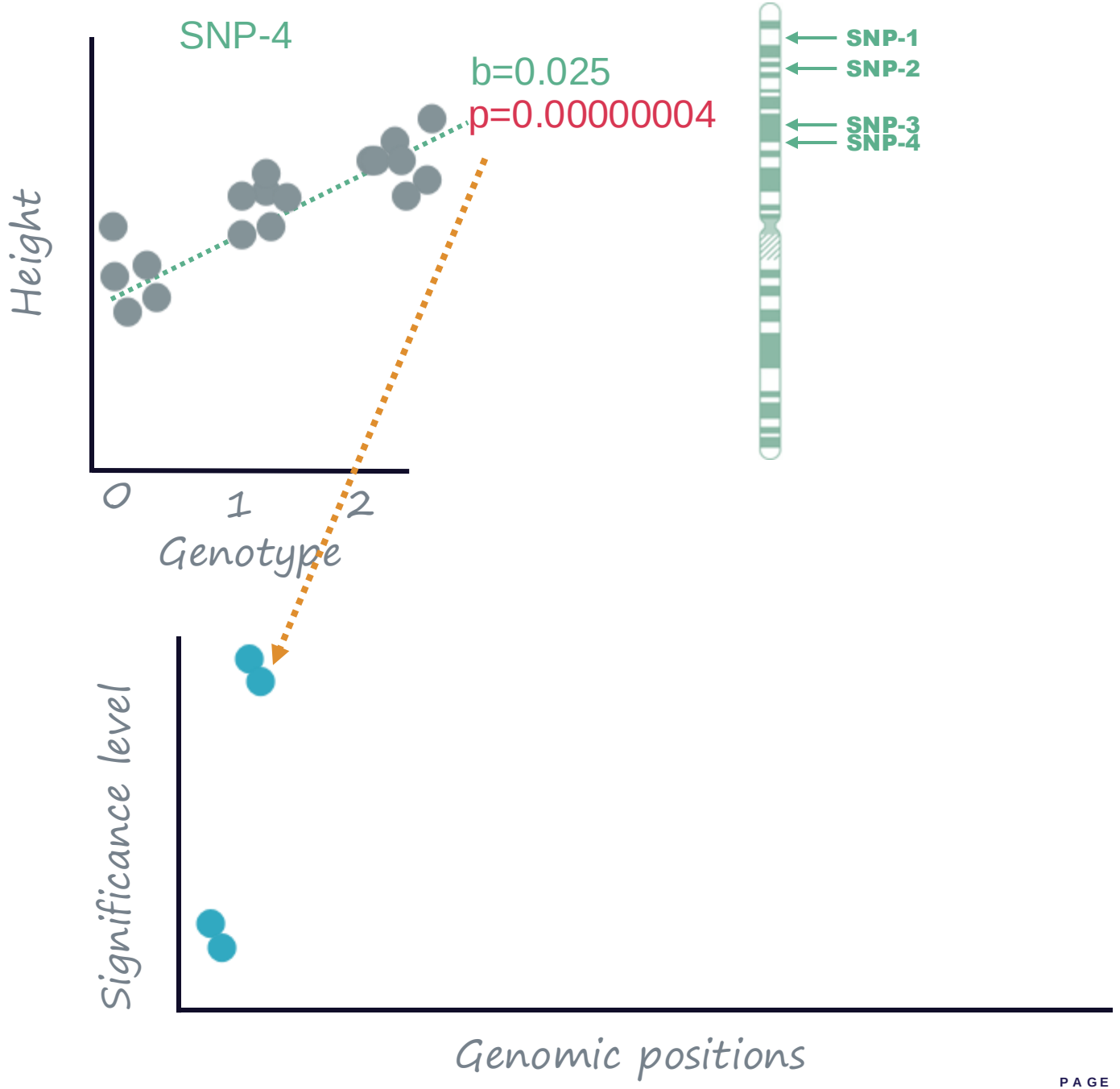
# GWAS RECAP

**Which SNPs associate with height?**



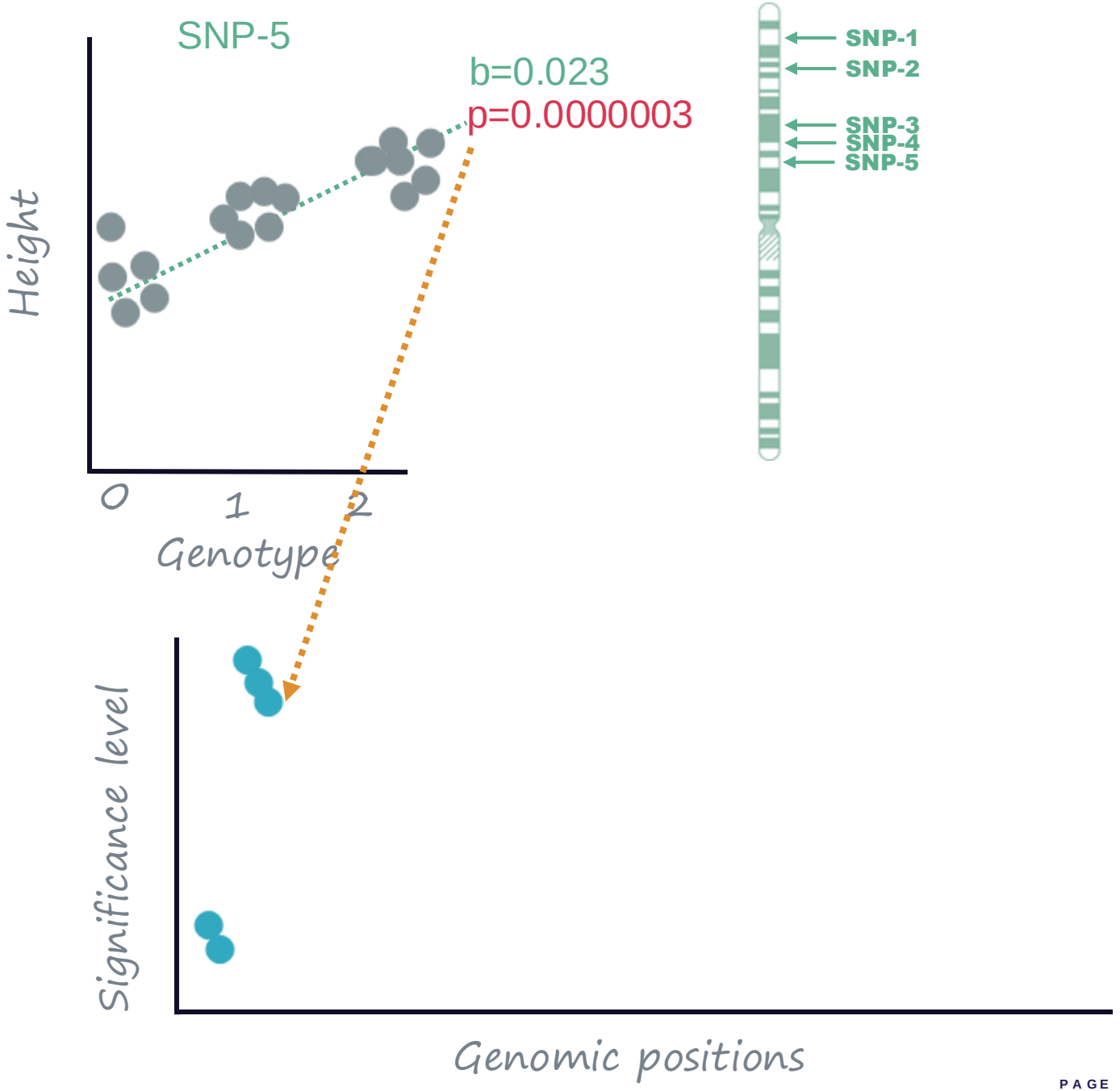
# GWAS RECAP

**Which SNPs associate with height?**



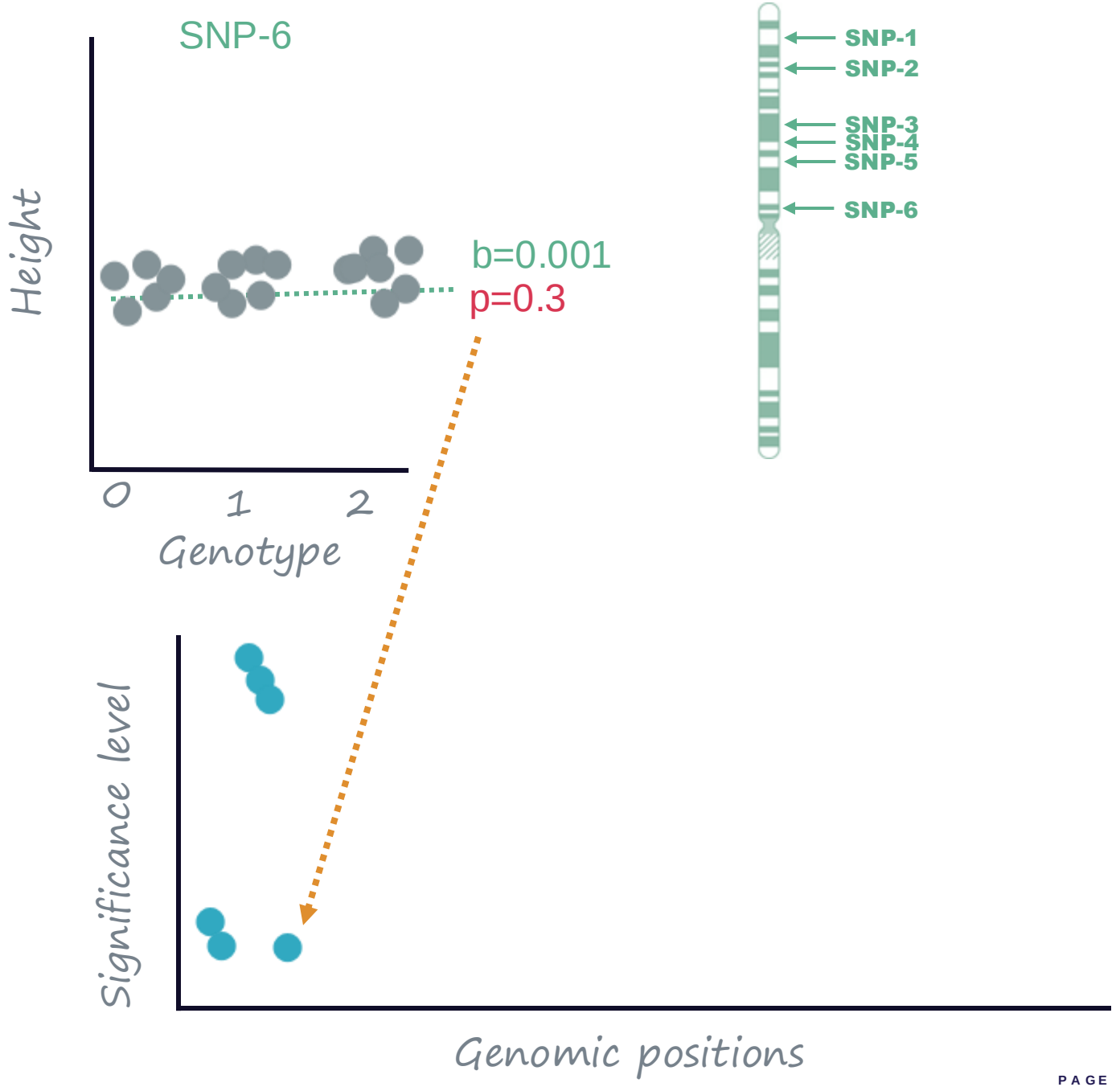
# GWAS RECAP

**Which SNPs associate with height?**



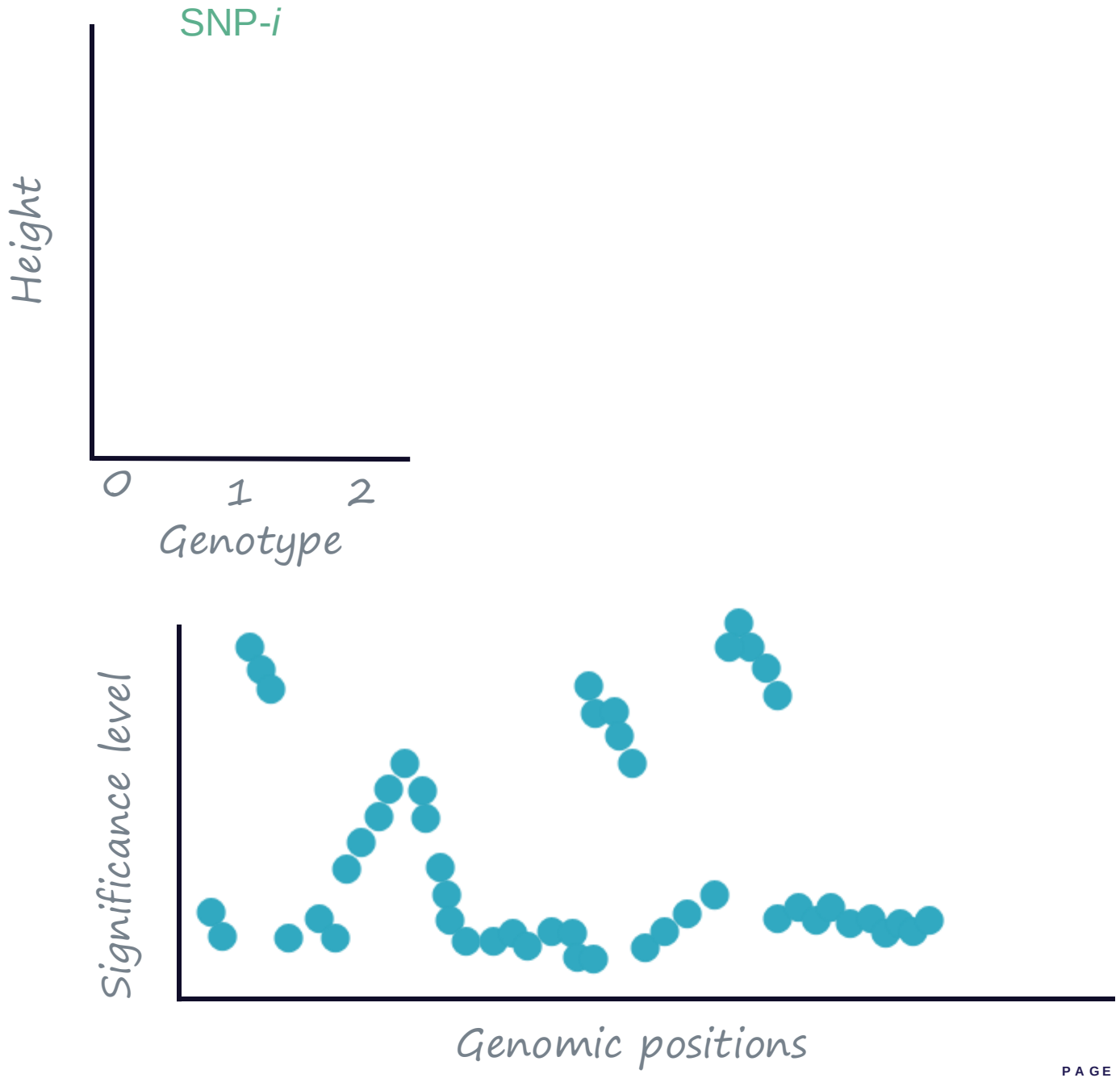
# GWAS RECAP

**Which SNPs associate with height?**



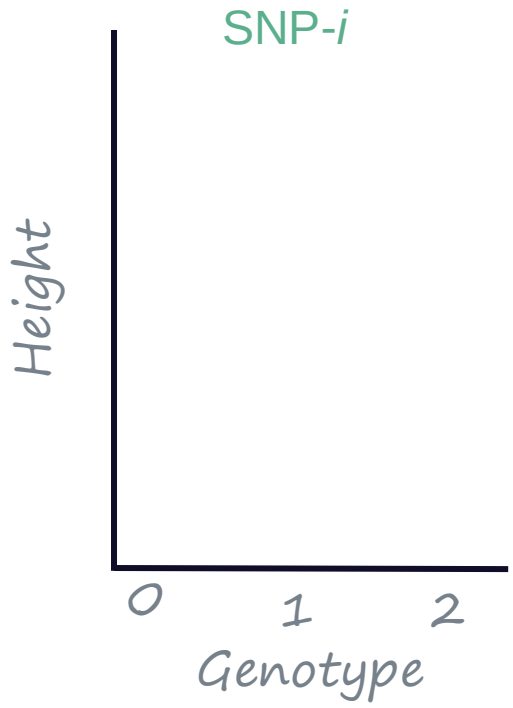
# GWAS RECAP

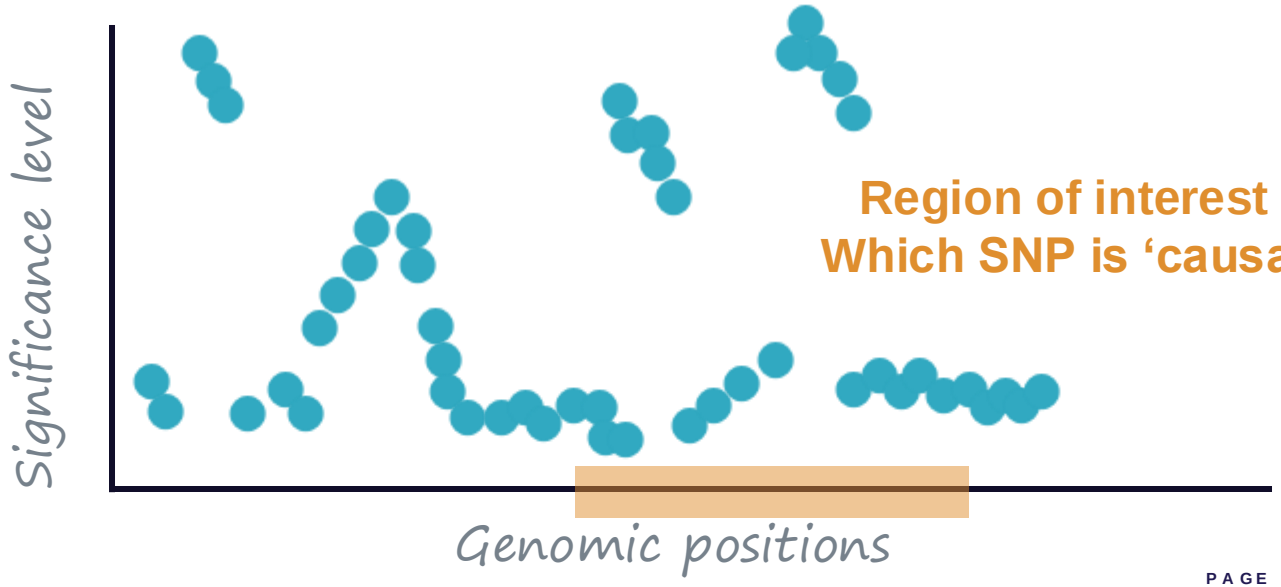
**Which SNPs associate with height?**



# GWAS RECAP

**Which SNPs associate with height?**



# Linkage disequilibrium (LD) – what is it?

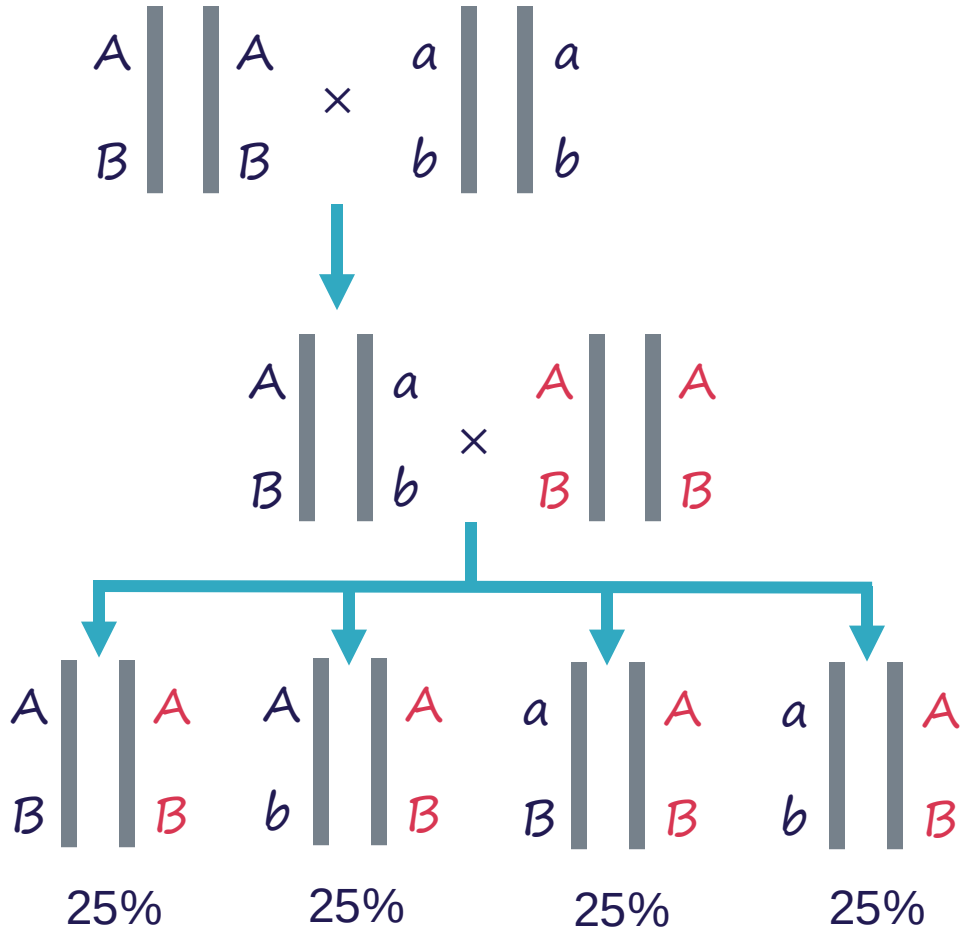
# LD REVISITED

**Mendel's law of independent assortment**

- Genes do not influenec each other with regard to sorting of alleles into gameets

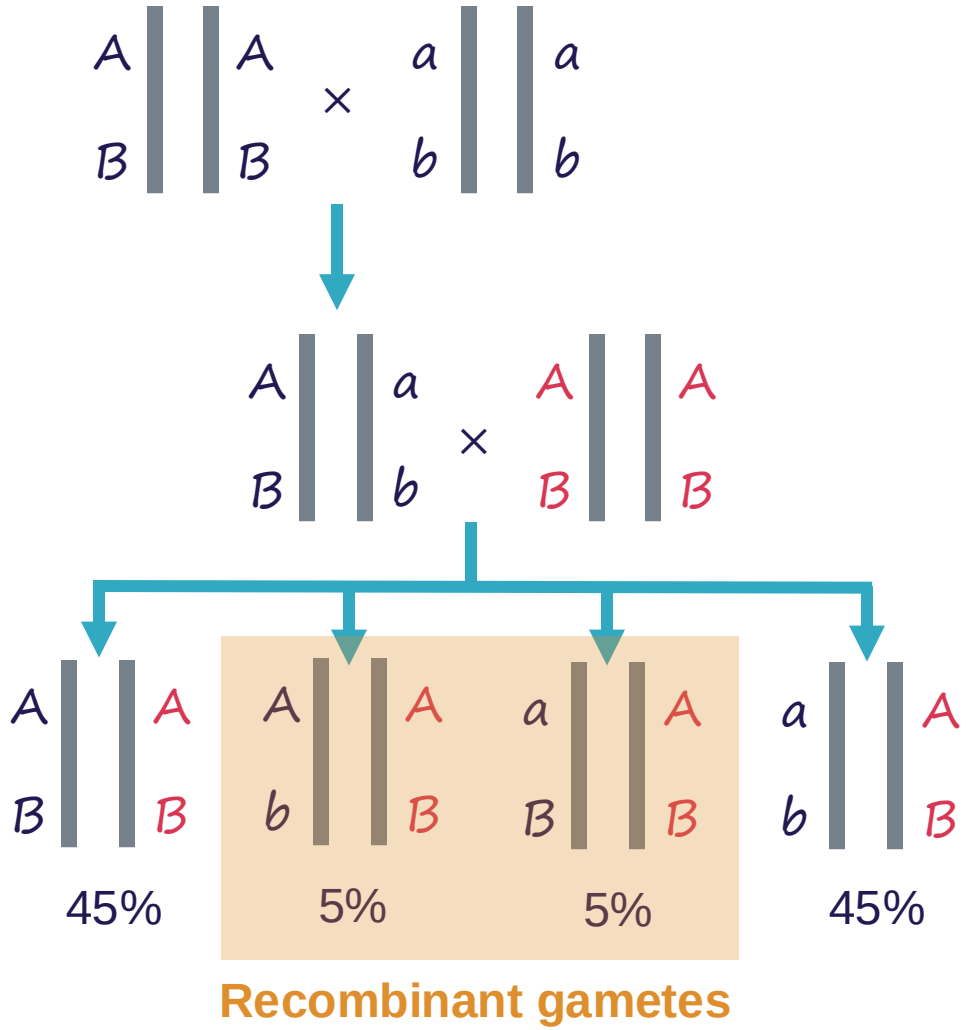
- Every possible combination of alleles for every gene is equally likely to occur

# INDEPENDENT ASSORTMENT




**Independent assortment**

# NON-INDEPENDENT ASSORTMENT



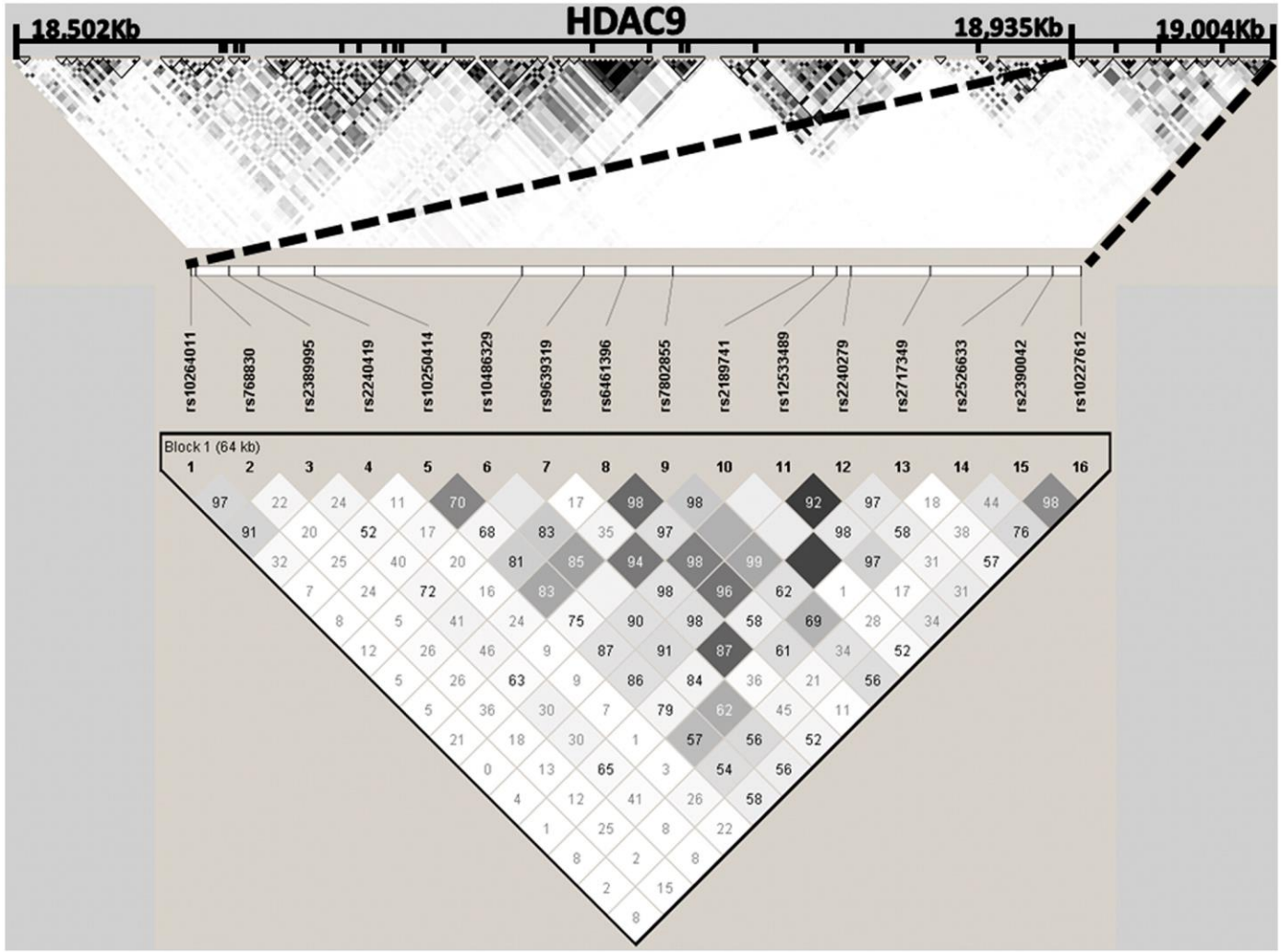

**Non-independent assortment**

# LD REVISITED

Non-independent assortment / Non-random association of loci within the population

Parameter of the entire population

LD can be “measured” between pairs of SNPs (correlation or D)



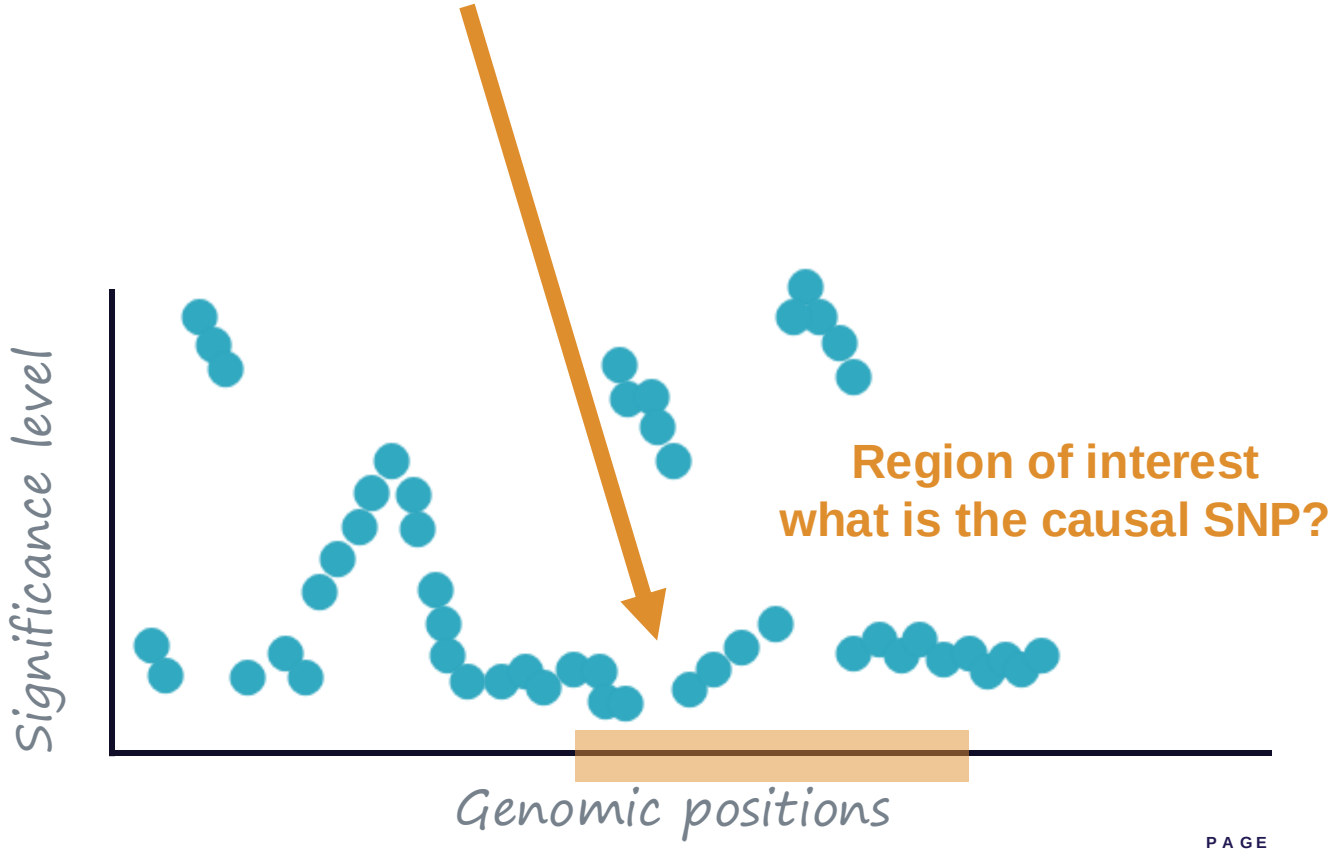
# GWAS RECAP

**Which SNPs associate with height?**

**Quantitative trait loci (QTL)**

→ not the causal variant

→ a variant LINKED to the causal (un-measured variant)



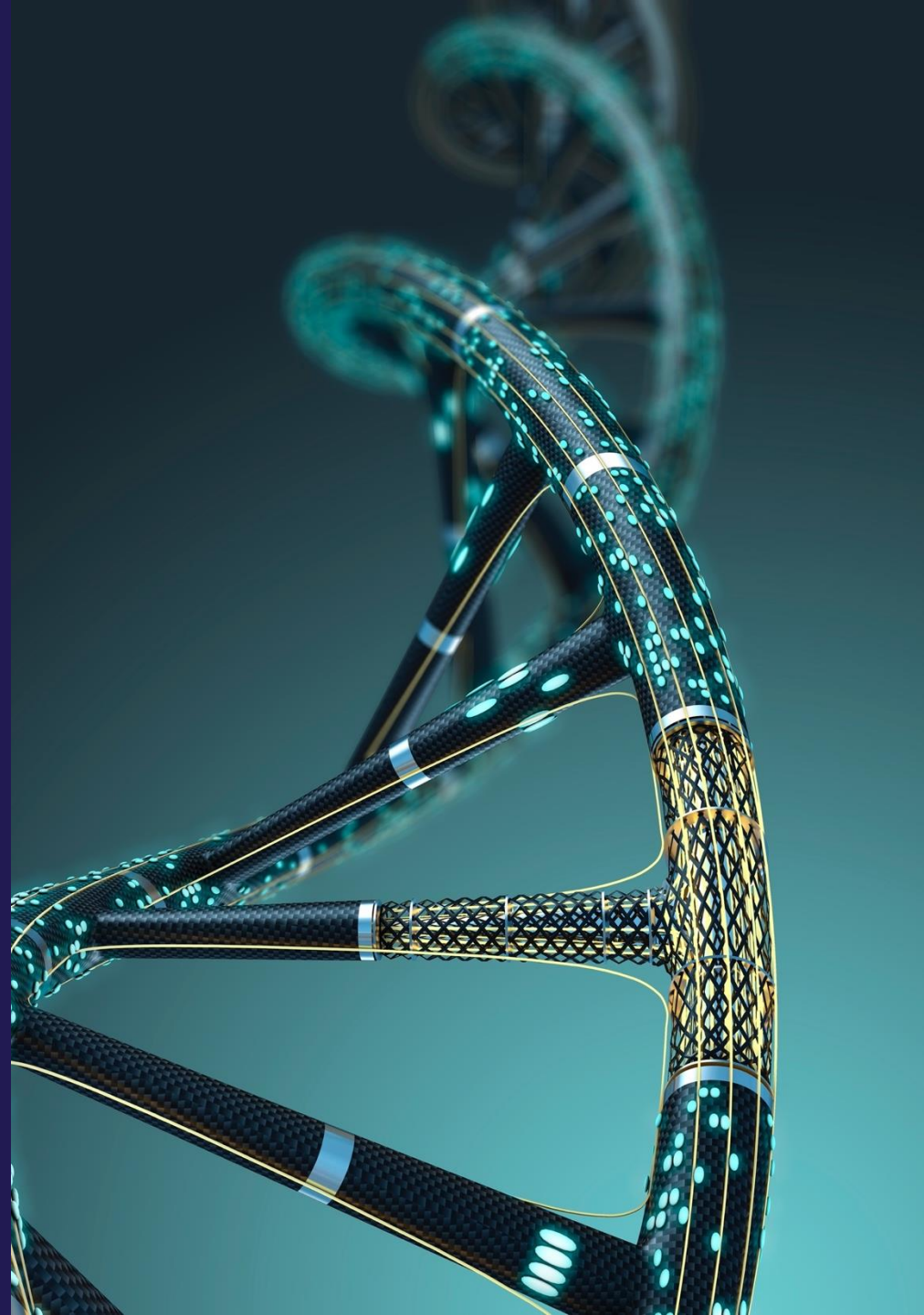

BREAK

# AGENDA

| | |
|---|---|
| **08:15 – 08:45** | Recap [*GWAS + R exercise from last*] |
| **08:45 – 08:50** | Break |
| **08:50 – 09:20** | Lecture [*PGS*] |
| **09:20 – 09:30** | Break |
| **09:30 – 10:30** | Exercise 1 + 2 [*+ joint discussion* ] |
| **10:30 – 10:40** | Break |
| **10:40 – 11:30** | Group work |
| **11:30 – 11:55** | Presentation of group work |
| **11:55 – 12:00** | Evaluation at Moodle |

# PREDICTING DISEASE RISK FROM GENETIC DATA?

A “polygenic score” is one way by which people can learn about their risk of developing a disease, based on the total number of changes (i.e., SNPs) related to the disease (NHI)



# DIFFERENT NAMES

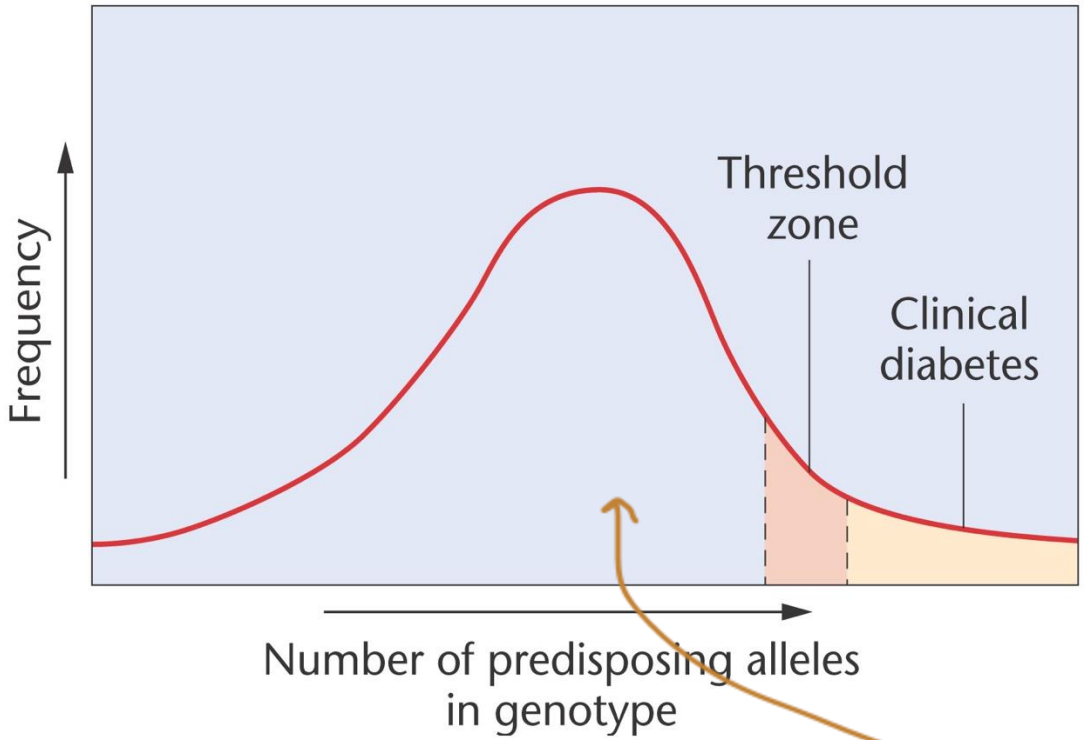
## – BUT THE SAME

- Polygenic risk score (PRS)
- Polygenic score (PGS)
- Genetic score (GS)
- Genetic risk score (GRS)
- Genetic value
- Genetic liability
- …

# DIFFERENT NAMES
## – BUT THE SAME

- Polygenic risk score (PRS)
- **Polygenic score (PGS)**
- Genetic score (GS)
- Genetic risk score (GRS)
- Genetic value
- Genetic liability
- …

# WHAT IS A PGS?



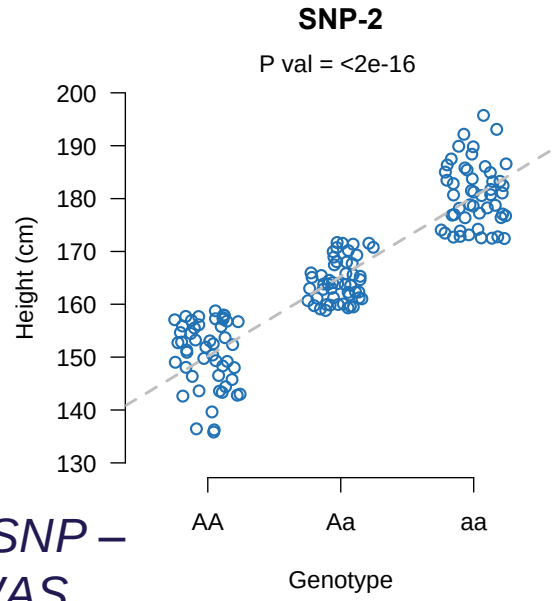

SNP1 A G
SNP2 T T
SNP3 A C
SNP4 T A
SNP5 C C

PGS =7

# WHAT IS A PGS?

"*A PGS combines information from large numbers of markers across the genome (hundreds to millions) to give a single numerical score for an individual's risk for developing a specific disease on the basis of the DNA variants they have inherited.*"

*b* is the slope (effect size) from reggresion




*The effect size of the SNP – obtain from the GWAS*

$$PGS = \sum X_i b_i$$

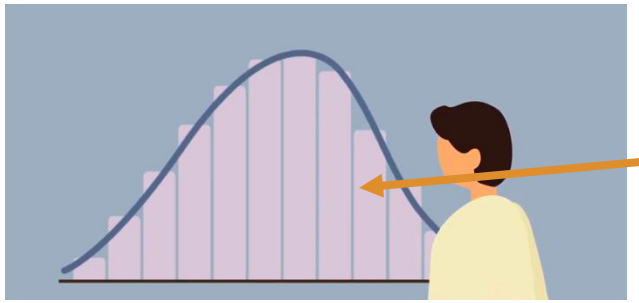
*The genotype of the individual for SNP i (0, 1, 2 – counting the number of the alternative allele)*

AA = 0
Aa = 1
aa = 2

$$PGS = \sum X_i b_i$$

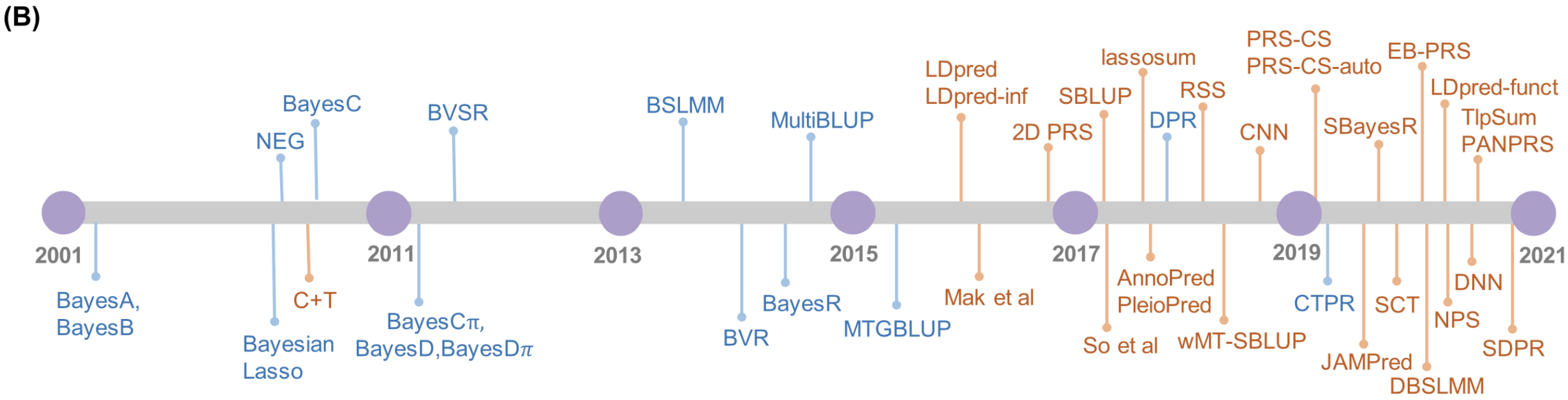
# HOW TO COMPUTE A (simple) PGS?

| SNPs | Adams Genotypes | Ref allele | Alt allele | X | b | Xb |
|---|---|---|---|---|---|---|
| SNP-1 | TC | T | C | 1 | 0.04 | 0.04 |
| SNP-2 | GG | G | T | 0 | 0.02 | 0.00 |
| SNP-3 | CC | A | C | 2 | 0.05 | 0.10 |
| SNP-4 | TG | T | G | 1 | 0.02 | 0.02 |
| SNP-5 | AA | A | G | 0 | 0.06 | 0.00 |



**PGS = 0.16**

# A LARGE PALETTE OF PGS METHODS

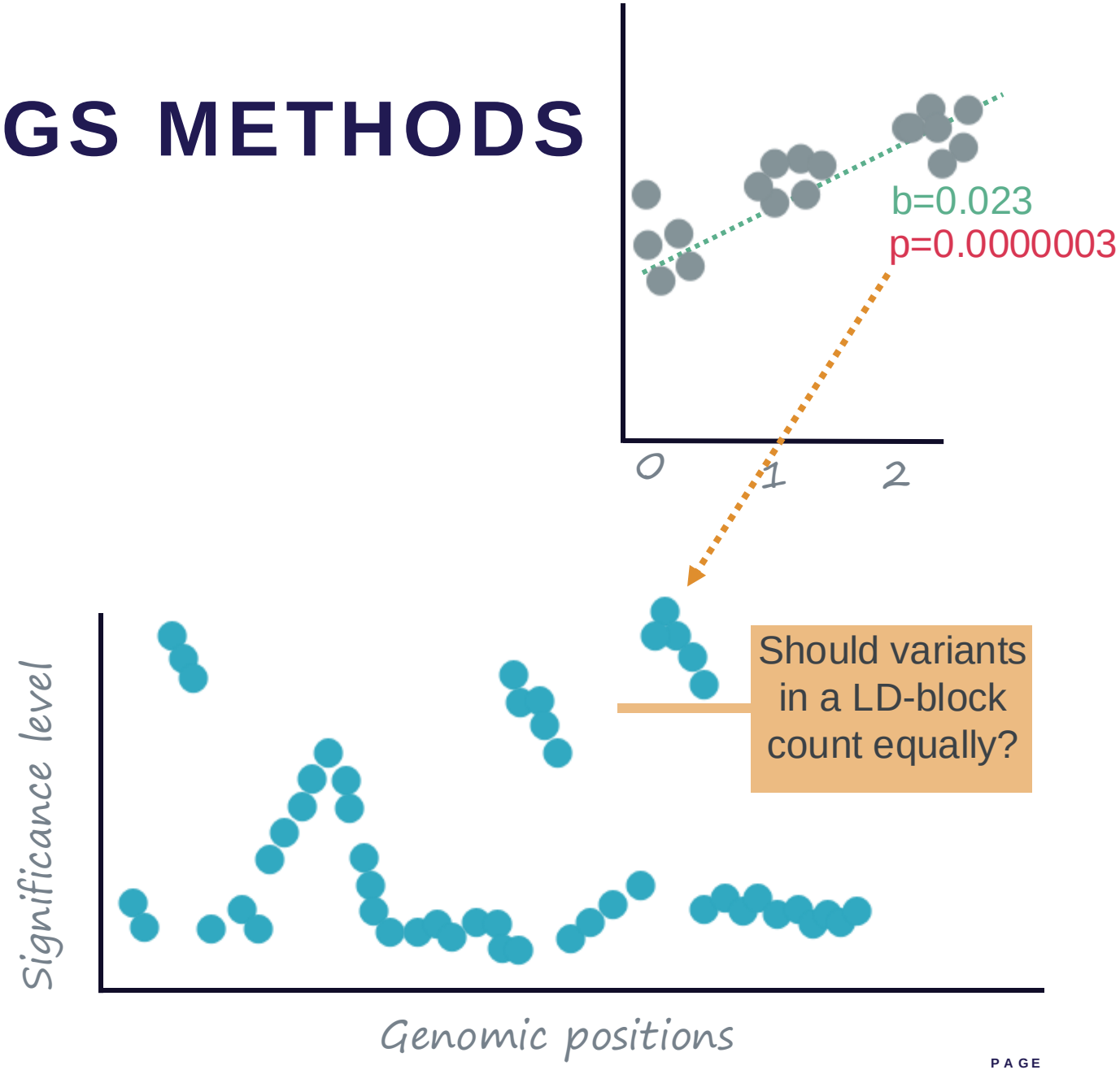





Ma, Y., & Zhou, X. (2021). *Trends in Genetics*, *37*(11), 995–1011.

# WHY DIFFERENT PGS METHODS

$$PGS = \sum X_i b_i$$



# CLUMPING AND THRESHOLDING (C+T)

### 0: Set LD (=0.8) and *P* values (0.01)

| SNP | b | p |
|---|---|---|
| 1 | 0.21 | 0.005 |
| 2 | 0.22 | 0.0048 |
| 3 | 0.25 | 0.0003 |
| 4 | 0.1 | 0.04 |
| 5 | 0.05 | 0.15 |
| 6 | 0.02 | 0.49 |
| 7 | 0.03 | 0.87 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 10 | 0.18 | 0.0004 |
| 11 | 0.21 | 0.00003 |
| 12 | 0.12 | 0.15 |
| 13 | 0.14 | 0.12 |
| 14 | 0.03 | 0.84 |
| 15 | 0.02 | 0.32 |

### 1: Sort by P-value

| SNP | b | p |
|---|---|---|
| 11 | 0.21 | 0.00003 |
| 3 | 0.25 | 0.0003 |
| 10 | 0.18 | 0.0004 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 2 | 0.22 | 0.0048 |
| 1 | 0.21 | 0.005 |
| 4 | 0.1 | 0.04 |
| 13 | 0.14 | 0.12 |
| 5 | 0.05 | 0.15 |
| 12 | 0.12 | 0.15 |
| 15 | 0.02 | 0.32 |
| 6 | 0.02 | 0.49 |
| 14 | 0.03 | 0.84 |
| 7 | 0.03 | 0.87 |

### 2: Compute LD and select variants based of thresholds

| SNP | b | p | $r^2$ |
|---|---|---|---|
| 11 | 0.21 | 0.00003 | |
| 3 | 0.25 | 0.0003 | 0.96 |
| 10 | 0.18 | 0.0004 | 0.93 |
| 8 | 0.12 | 0.003 | 0.88 |
| 9 | 0.14 | 0.0034 | 0.74 |
| 2 | 0.22 | 0.0048 | 0.4 |
| 1 | 0.21 | 0.005 | 0.03 |
| 4 | 0.1 | 0.04 | 0.04 |
| 13 | 0.14 | 0.12 | 0.05 |
| 5 | 0.05 | 0.15 | 0.03 |
| 12 | 0.12 | 0.15 | 0.04 |
| 15 | 0.02 | 0.32 | 0.01 |
| 6 | 0.02 | 0.49 | 0.01 |
| 14 | 0.03 | 0.84 | 0.01 |
| 7 | 0.03 | 0.87 | 0.01 |

1st variant in LD-pair

Have LD>$r^2$ – ignore those

# CLUMPING AND THRESHOLDING (C+T)

### 0: Set LD (=0.8) and *P* values (0.01)

| SNP | b | p |
|---|---|---|
| 1 | 0.21 | 0.005 |
| 2 | 0.22 | 0.0048 |
| 3 | 0.25 | 0.0003 |
| 4 | 0.1 | 0.04 |
| 5 | 0.05 | 0.15 |
| 6 | 0.02 | 0.49 |
| 7 | 0.03 | 0.87 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 10 | 0.18 | 0.0004 |
| 11 | 0.21 | 0.00003 |
| 12 | 0.12 | 0.15 |
| 13 | 0.14 | 0.12 |
| 14 | 0.03 | 0.84 |
| 15 | 0.02 | 0.32 |

### 1: Sort by P-value

| SNP | b | p |
|---|---|---|
| 11 | 0.21 | 0.00003 |
| 3 | 0.25 | 0.0003 |
| 10 | 0.18 | 0.0004 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 2 | 0.22 | 0.0048 |
| 1 | 0.21 | 0.005 |
| 4 | 0.1 | 0.04 |
| 13 | 0.14 | 0.12 |
| 5 | 0.05 | 0.15 |
| 12 | 0.12 | 0.15 |
| 15 | 0.02 | 0.32 |
| 6 | 0.02 | 0.49 |
| 14 | 0.03 | 0.84 |
| 7 | 0.03 | 0.87 |

### 2: Compute LD and select variants based of thresholds

| SNP | b | p | $r^2$ |
|---|---|---|---|
| 11 | 0.21 | 0.00003 | |
| 3 | 0.25 | 0.0003 | |
| 10 | 0.18 | 0.0004 | |
| 8 | 0.12 | 0.003 | |
| 9 | 0.14 | 0.0034 | |
| 2 | 0.22 | 0.0048 | 0.98 |
| 1 | 0.21 | 0.005 | 0.96 |
| 4 | 0.1 | 0.04 | 0.96 |
| 13 | 0.14 | 0.12 | 0.52 |
| 5 | 0.05 | 0.15 | 0.34 |
| 12 | 0.12 | 0.15 | 0.10 |
| 15 | 0.02 | 0.32 | 0.04 |
| 6 | 0.02 | 0.49 | 0.01 |
| 14 | 0.03 | 0.84 | 0.01 |
| 7 | 0.03 | 0.87 | 0.01 |

1st variant in LD-pair

Have LD>$r^2$ – ignore those

# CLUMPING AND THRESHOLDING (C+T)

### 0: Set LD (=0.8) and *P* values (0.01)

| SNP | b | p |
|---|---|---|
| 1 | 0.21 | 0.005 |
| 2 | 0.22 | 0.0048 |
| 3 | 0.25 | 0.0003 |
| 4 | 0.1 | 0.04 |
| 5 | 0.05 | 0.15 |
| 6 | 0.02 | 0.49 |
| 7 | 0.03 | 0.87 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 10 | 0.18 | 0.0004 |
| 11 | 0.21 | 0.00003 |
| 12 | 0.12 | 0.15 |
| 13 | 0.14 | 0.12 |
| 14 | 0.03 | 0.84 |
| 15 | 0.02 | 0.32 |

### 1: Sort by P-value

| SNP | b | p |
|---|---|---|
| 11 | 0.21 | 0.00003 |
| 3 | 0.25 | 0.0003 |
| 10 | 0.18 | 0.0004 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 2 | 0.22 | 0.0048 |
| 1 | 0.21 | 0.005 |
| 4 | 0.1 | 0.04 |
| 13 | 0.14 | 0.12 |
| 5 | 0.05 | 0.15 |
| 12 | 0.12 | 0.15 |
| 15 | 0.02 | 0.32 |
| 6 | 0.02 | 0.49 |
| 14 | 0.03 | 0.84 |
| 7 | 0.03 | 0.87 |

### 2: Compute LD and select variants based of thresholds

| SNP | b | p | $r^2$ |
|---|---|---|---|
| 11 | 0.21 | 0.00003 | |
| 3 | 0.25 | 0.0003 | |
| 10 | 0.18 | 0.0004 | |
| 8 | 0.12 | 0.003 | |
| 9 | 0.14 | 0.0034 | |
| 2 | 0.22 | 0.0048 | |
| 1 | 0.21 | 0.005 | |
| 4 | 0.1 | 0.04 | |
| 13 | 0.14 | 0.12 | |
| 5 | 0.05 | 0.15 | 0.86 |
| 12 | 0.12 | 0.15 | 0.82 |
| 15 | 0.02 | 0.32 | 0.81 |
| 6 | 0.02 | 0.49 | 0.85 |
| 14 | 0.03 | 0.84 | 0.85 |
| 7 | 0.03 | 0.87 | 0.81 |

1st variant in LD-pair

Have LD>$r^2$ – ignore those

# CLUMPING AND THRESHOLDING (C+T)

**0: Set LD (=0.8) and *P* values (0.01)**

| SNP | b | p |
|---|---|---|
| 1 | 0.21 | 0.005 |
| 2 | 0.22 | 0.0048 |
| 3 | 0.25 | 0.0003 |
| 4 | 0.1 | 0.04 |
| 5 | 0.05 | 0.15 |
| 6 | 0.02 | 0.49 |
| 7 | 0.03 | 0.87 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 10 | 0.18 | 0.0004 |
| 11 | 0.21 | 0.00003 |
| 12 | 0.12 | 0.15 |
| 13 | 0.14 | 0.12 |
| 14 | 0.03 | 0.84 |
| 15 | 0.02 | 0.32 |

**1: Sort by P-value**

| SNP | b | p |
|---|---|---|
| 11 | 0.21 | 0.00003 |
| 3 | 0.25 | 0.0003 |
| 10 | 0.18 | 0.0004 |
| 8 | 0.12 | 0.003 |
| 9 | 0.14 | 0.0034 |
| 2 | 0.22 | 0.0048 |
| 1 | 0.21 | 0.005 |
| 4 | 0.1 | 0.04 |
| 13 | 0.14 | 0.12 |
| 5 | 0.05 | 0.15 |
| 12 | 0.12 | 0.15 |
| 15 | 0.02 | 0.32 |
| 6 | 0.02 | 0.49 |
| 14 | 0.03 | 0.84 |
| 7 | 0.03 | 0.87 |

**2: Compute LD and select variants based on LD**

| SNP | b | p | $r^2$ |
|---|---|---|---|
| 11 | 0.21 | 0.00003 | |
| 3 | 0.25 | 0.0003 | |
| 10 | 0.18 | 0.0004 | |
| 8 | 0.12 | 0.003 | |
| 9 | 0.14 | 0.0034 | |
| 2 | 0.22 | 0.0048 | |
| 1 | 0.21 | 0.005 | |
| 4 | 0.1 | 0.04 | |
| 13 | 0.14 | 0.12 | |
| 5 | 0.05 | 0.15 | |
| 12 | 0.12 | 0.15 | |
| 15 | 0.02 | 0.32 | |
| 6 | 0.02 | 0.49 | |
| 14 | 0.03 | 0.84 | |
| 7 | 0.03 | 0.87 | |

**3: Compute PGS based on effect sizes (b) and *P*-values**

$$PGS = \sum X_i b_i$$

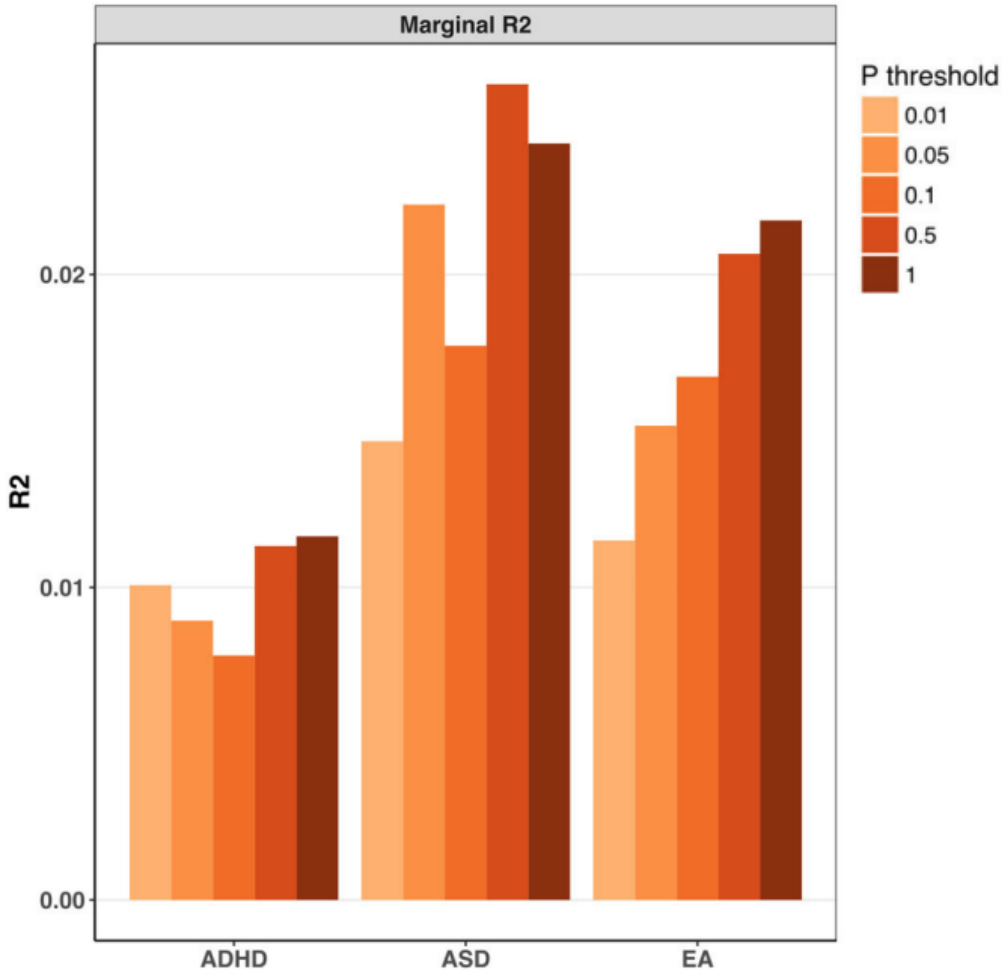
$$= \boldsymbol{X}_{11} \times 0.21 + \boldsymbol{X}_{9} \times 0.14$$

# CLUMPING AND THRESHOLDING (C+T)

**Repeat for other *P*-value cutoffs (and LD values)**

**How does the PGS associate with the disease**

$$y_{disease} = PGS + \varepsilon$$



# SCHRINKAGE METHODS

**Clumping and thresholding** (C+T)

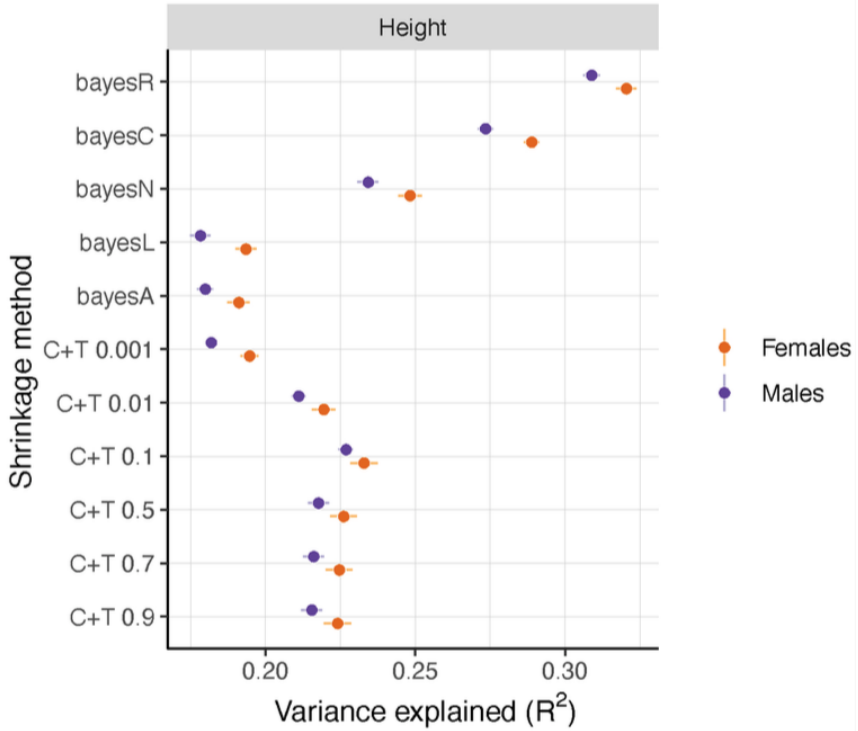
**Bayes-N**; $\beta \sim N(0, \sigma^2_\beta)$

**Bayes-L**; $f(\beta_j|\tau^2_j, \sigma^2_e) \sim N(\beta_j|0, \tau^2_j \times \sigma^2_e)$

**Bayes-A**; $\beta_j \sim N(0, \sigma^2_{\beta_j})$

**Bayes-C**; $\beta_j \sim N(0, \sigma^2_{\beta_j})$ with probability $\pi$, and $\beta_j = 0$ with probability $(1 - \pi)$, where $\pi$ is assumed to follow a beta-distribution.

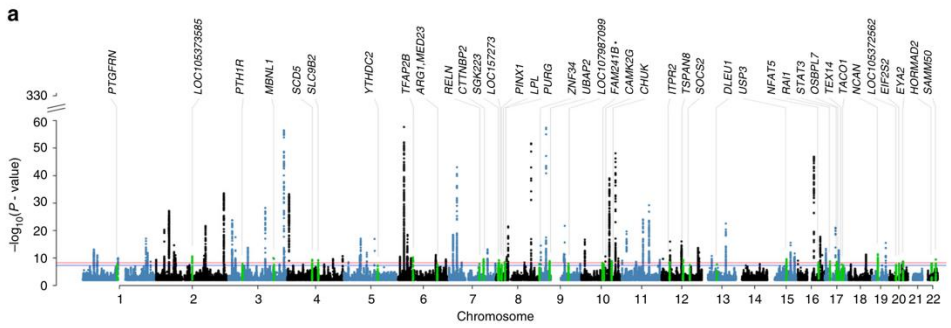
**Bayes-R**; $\beta_j \sim N(0, \gamma_C \sigma^2_{\beta_j})$, where $C$ defines number of classes (e.g., C=4,$\gamma = (0, 0.01, 0.1, 1.0)$)




Rohde et al., 2020
Rohde et al., 2023

# WHAT DO YOU NEED?

**1. A large well-powered GWAS for your trait of interest**




**2. An independent cohort that has been genotyped**

**(3. That some individuals in the cohort has the phenotype)**


BREAK

# AGENDA

| | |
|---|---|
| **08:15 – 08:45** | Recap [*GWAS + R exercise from last*] |
| **08:45 – 08:50** | Break |
| **08:50 – 09:20** | Lecture [*PGS*] |
| **09:20 – 09:30** | Break |
| **09:30 – 10:30** | Exercise 1 + 2 [*+ joint discussion* ] |
| **10:30 – 10:40** | Break |
| **10:40 – 11:30** | Group work |
| **11:30 – 11:55** | Presentation of group work |
| **11:55 – 12:00** | Evaluation at Moodle |


BREAK

# GROUP WORK

1) Make three groups
   - ❑ Within each group discuss the questions you get, and prepare to briefly present your thoughts in plenum
2) Plenum discuss [7 min pr group]

# YOUR OPPINION MATTERS

## MOODLE EVALUATION