Basic Genome-wide Association Studies

Robert Brown

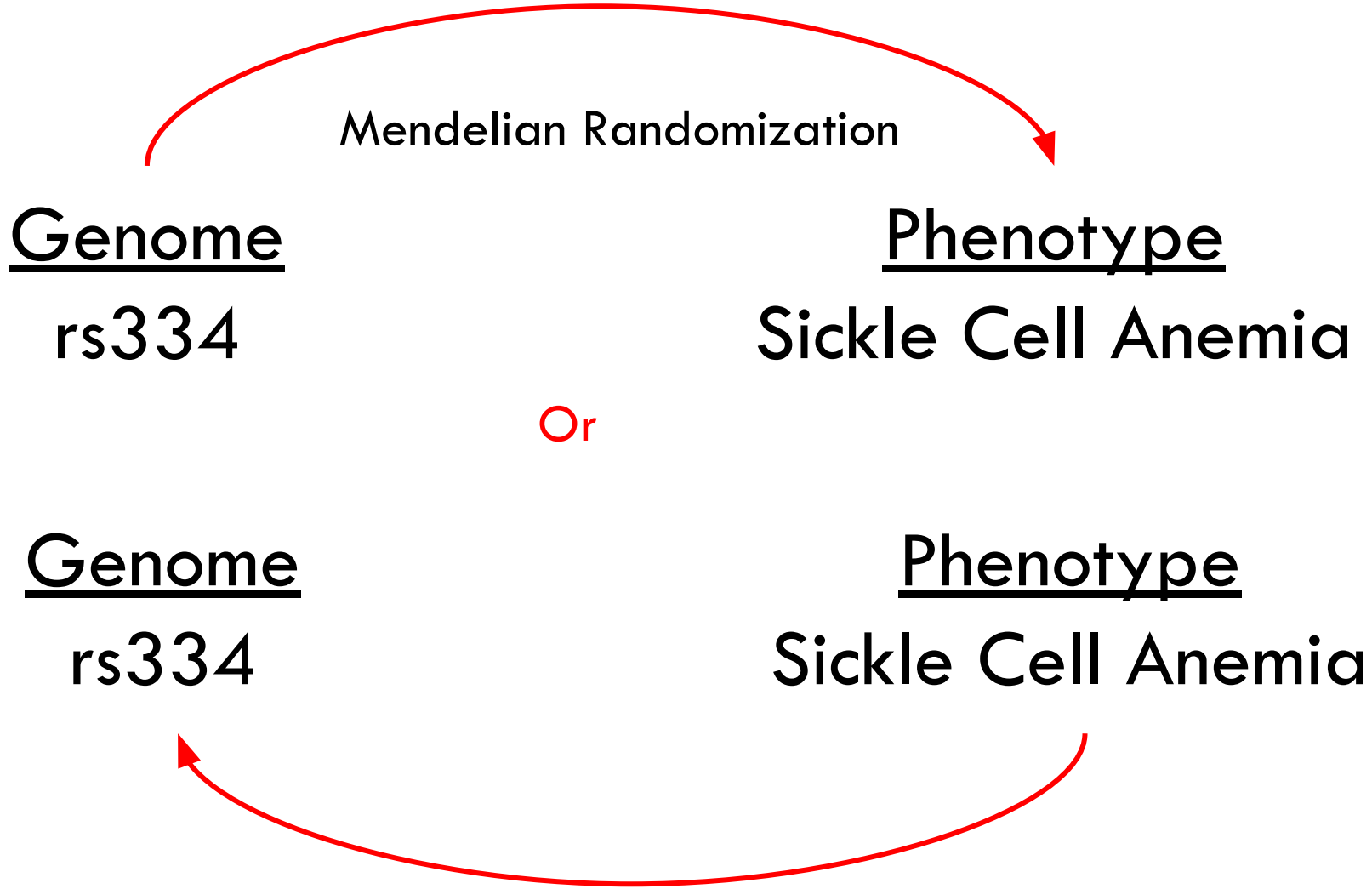
Thank you!

UCLA Computational Medicine

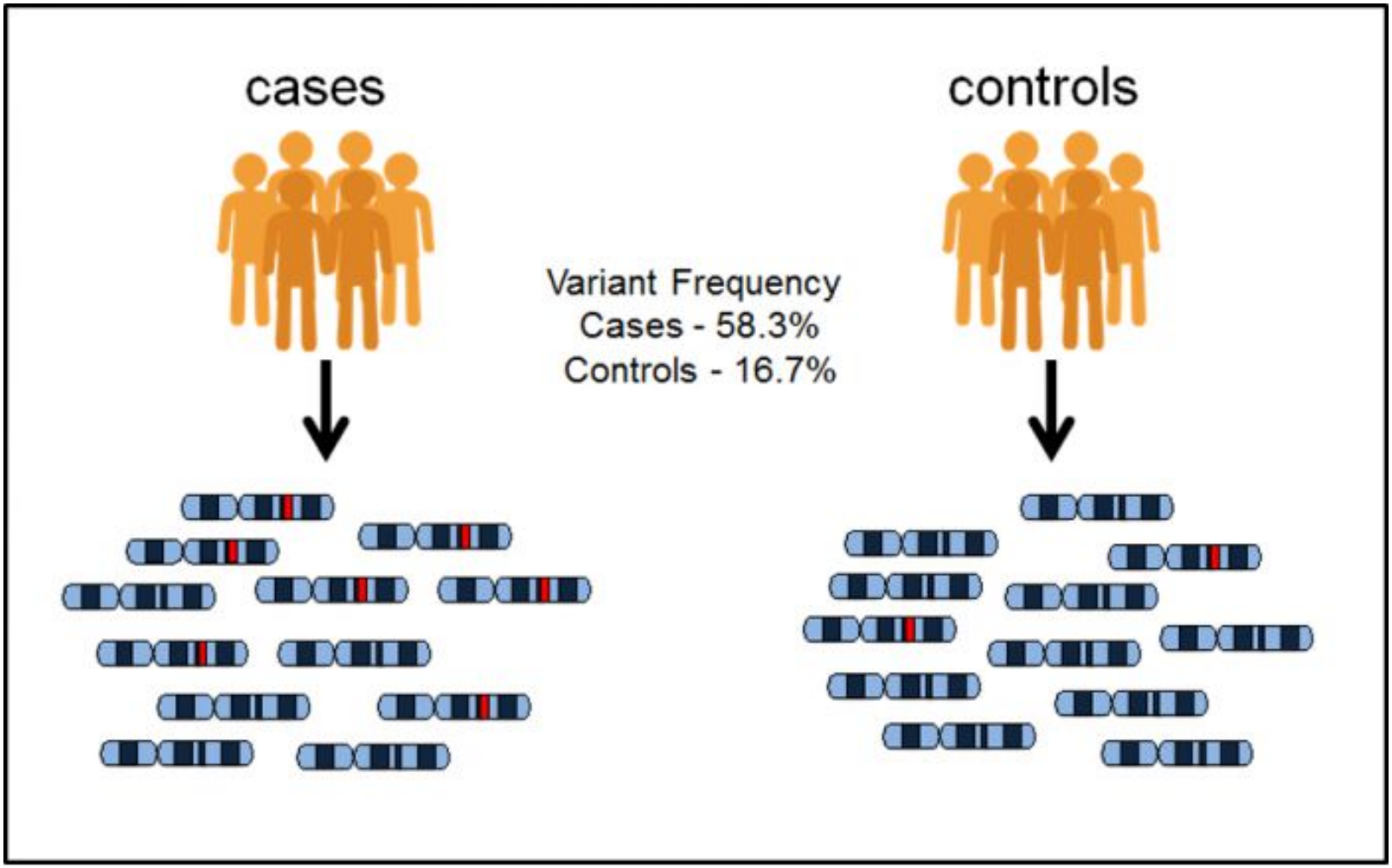
What is a phenotype

- Height
- BMI
- hair color
- diabetes
- autism
- annual salary
- favorite color
- blood sugar level
- RELN protein level

Sickle Cell Anemia: Causal Direction



Genome-wide association study (GWAS) Case-Control



https://www.ebi.ac.uk/training/online/course/gwas-catalog-exploring-snp-trait-associations/why-do-we-need-gwas-catalog/what-are-genome

Genome-wide association study (GWAS) Continuous Phenotype

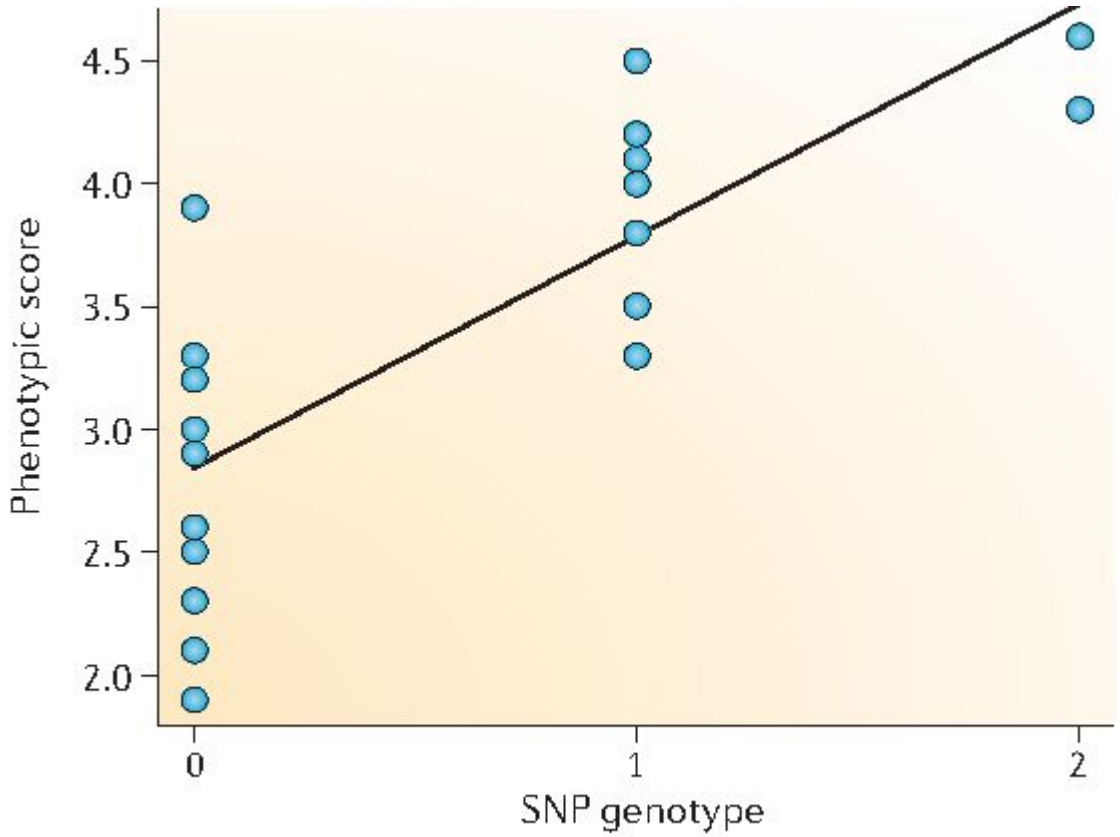


Figure 3 | **Linear regression test of single-SNP**

Balding, NRG 2006

GWAS Model: linear algebra style

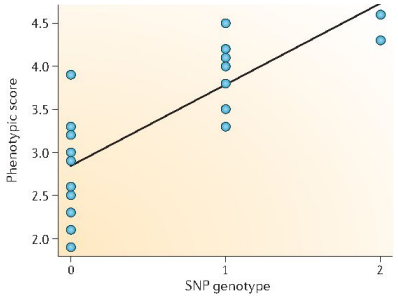


Figure 3 | **Linear regression test of single-SNP**

Full Model

$$Y = G\beta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1M} \\ g_{21} & g_{22} & \cdots & g_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ g_{N1} & g_{1N} & \cdots & g_{NM} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Not enough power to "fit" this model

Marginal Model

$$Y = G_i\beta_i + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} g_{1i} \\ g_{2i} \\ \vdots \\ g_{Ni} \end{bmatrix} \beta_i + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

GWAS Model

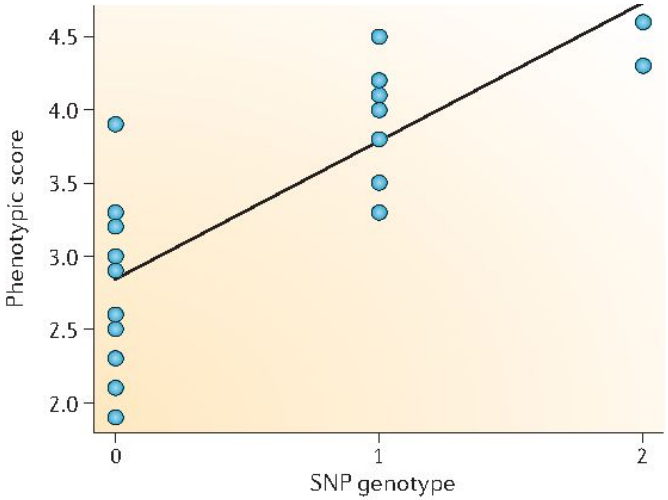


Figure 3 | **Linear regression test of single-SNP**

$$y_j = \beta_i g_{ij} + \varepsilon_j$$

individual: j
SNP: i
phenotype: y
genotype: g
Effect of SNP i: **β**
all other effects: **ε**

$$y_j = \beta_{Sex} S_j + \beta_{Age} A_j + \beta_{PC1} PC_{1j} + \beta_{PC2} PC_{2j} + \beta_i g_{ij} + \varepsilon_j$$

β_i: true effect

$\hat{\beta}_i$: estimated effect

GWAS Model

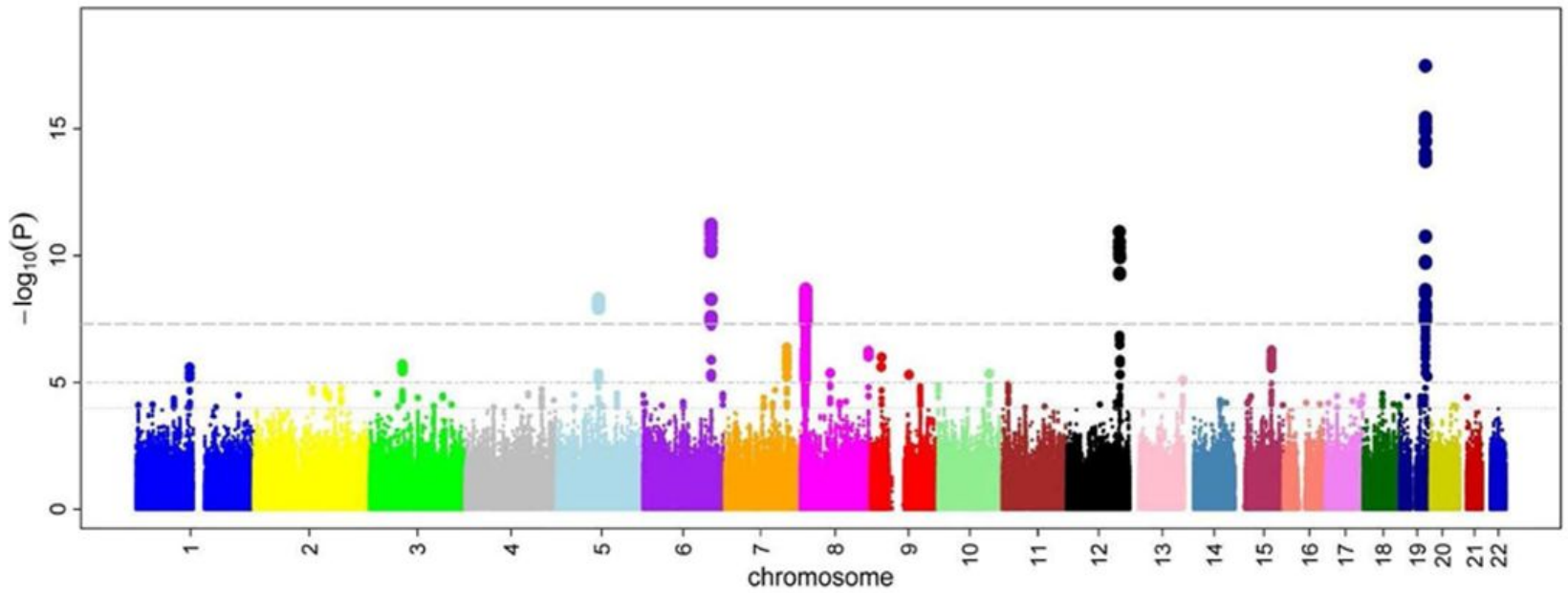
Marginal Model

$$Y = G_i\beta_i + \varepsilon$$

Null: genotype has no estimated effect

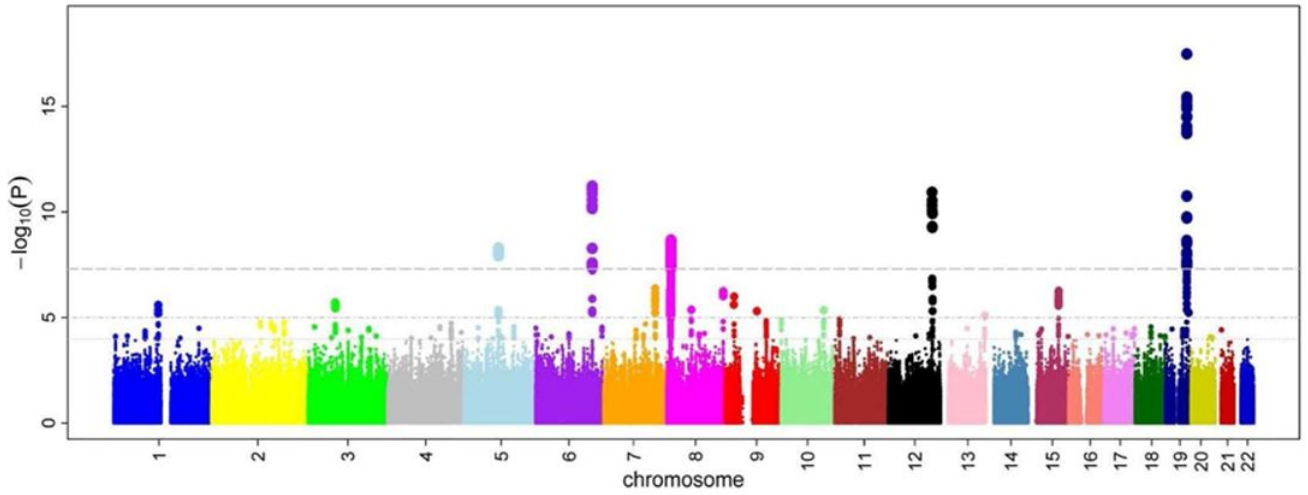
Alternate: genotype has non-zero estimated affect

$\hat{\beta}_i$: estimated effect



M. Kamran Ikram et al., PLoS Genet. 2010

GWAS Model: P-values and FWER



$$-log_{10}(0.05) = 1.3$$

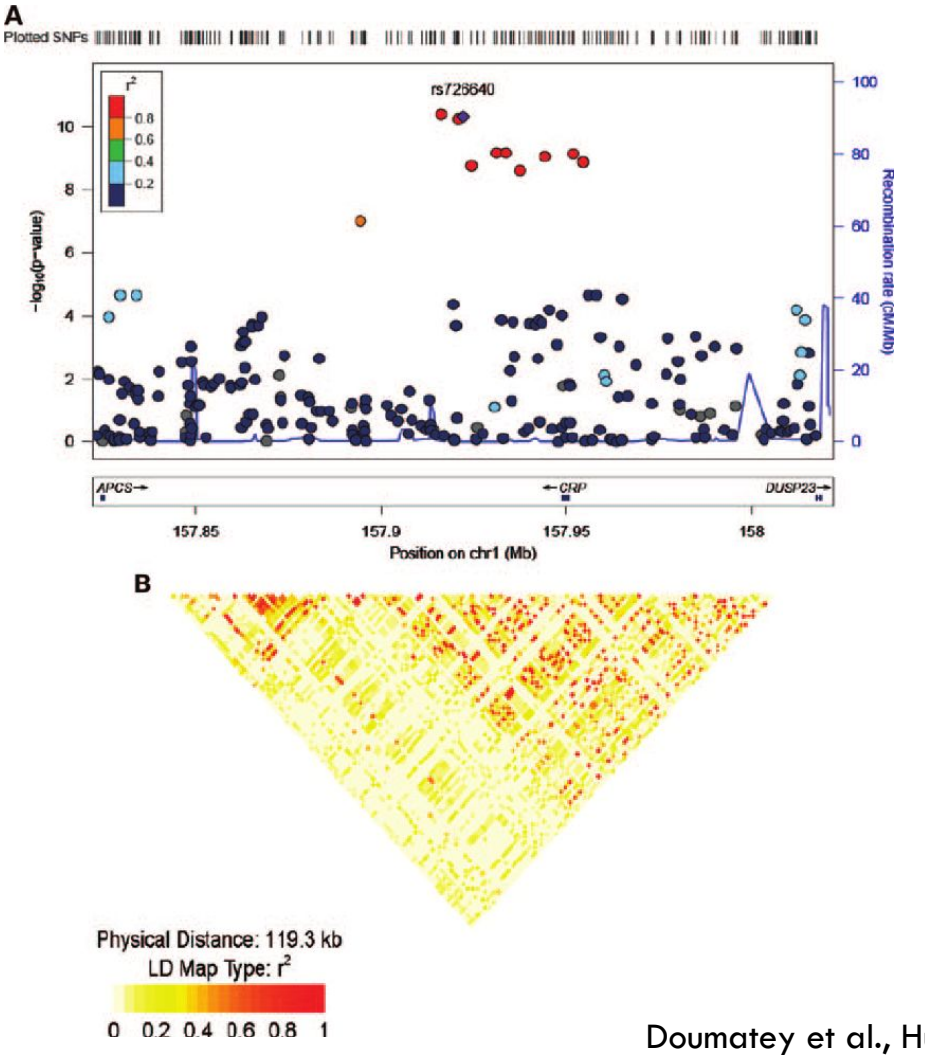
Family-wise error rate (FWER): The probability of have one or more false discoveries.

$$-log_{10}(5 \times 10^{-8}) = 7.3$$

The above significance threshold controls FWER at a 0.05 rate in GWAS

M. Kamran Ikram et al., PLoS Genet. 2010

GWAS locus vs GWAS gene



Doumatey et al., Human Molecular Genetics 2012

GWAS: Summary

- Mendelian Randomization: phenotypes cannot affect genotypes
- GWAS "top hits" tag causal/risk variation, but it is not clear what is causal/risk variation
- GWAS finds a locus of the genome associated with a phenotype
- GWAS output: Estimated effects for each variant and the corresponding p-value
- FWER is controlled using a very small significance threshold