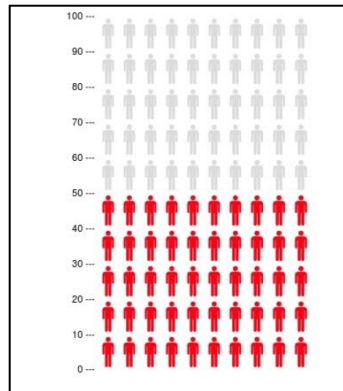


MODULE 3: Genome-wide association study — analysis

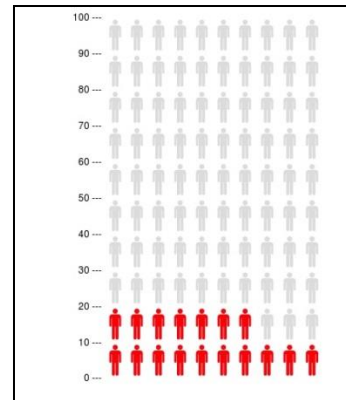
For a given SNP, how many people carry the variant allele?

With disease



50% carry variant

Without disease

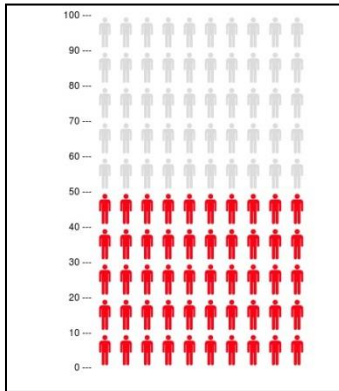


17% carry variant

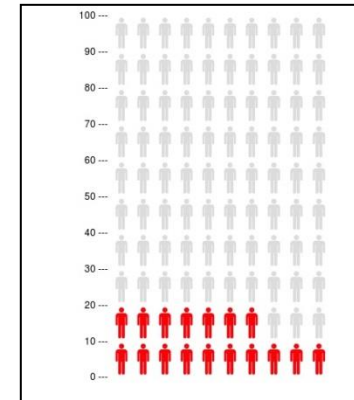
- Statistical test to compare the proportion of diseased and non-diseased individuals with the variant allele

Need to account for fact that humans have 2 copies of each gene

With disease



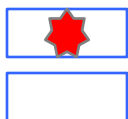
Without disease



50% carry variant



25% have 2 copies

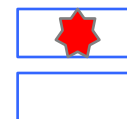


25% have 1 copy

17% carry variant



3% have 2 copies

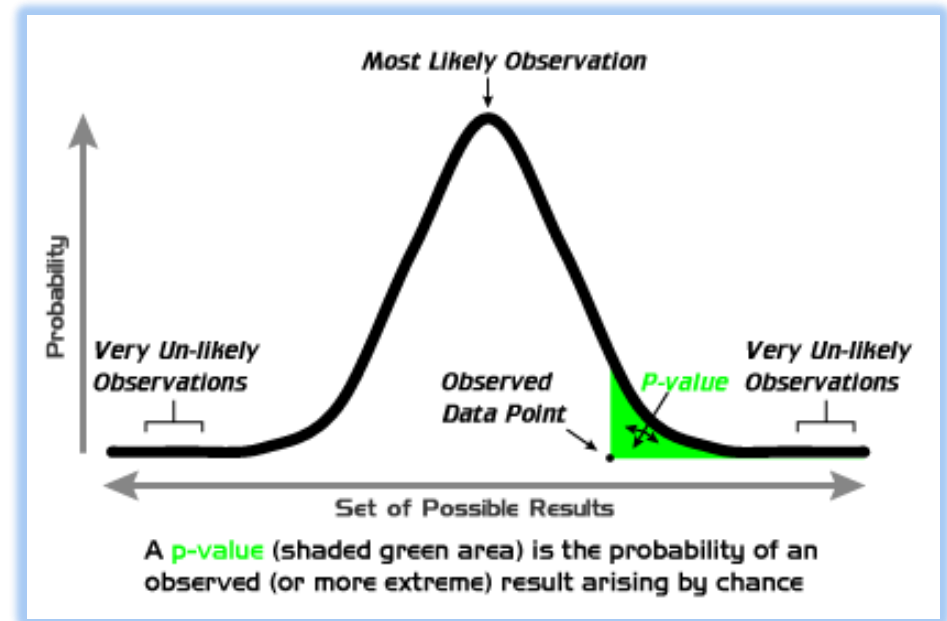


14% have 1 copy

A statistical test tells us how likely the results are true

- Compare proportion of diseased/non-diseased with zero, one or two copies of variant allele

	With disease	Without disease
2 copies	25 (25%)	26 (3%)
1 copy	25 (25%)	124 (14%)
0 copies	50 (50%)	750 (83%)



- Statistical test: Armitage trend test (1 d.f.)

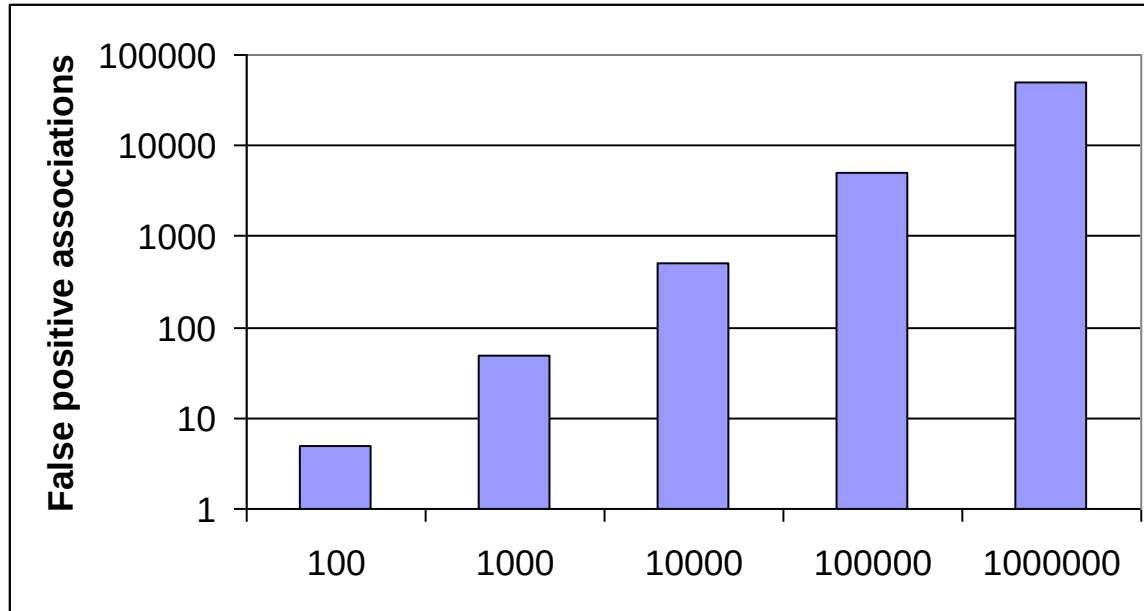
Hypothesis testing and p values

- Statistical test tells you whether the difference in allele distribution between the two groups is likely to be due to chance or not



- What does a $p < 0.05$ mean?
- $< 5\%$ probability that the observation is due to chance (i.e. a false positive)
- This association is 'statistically significant'

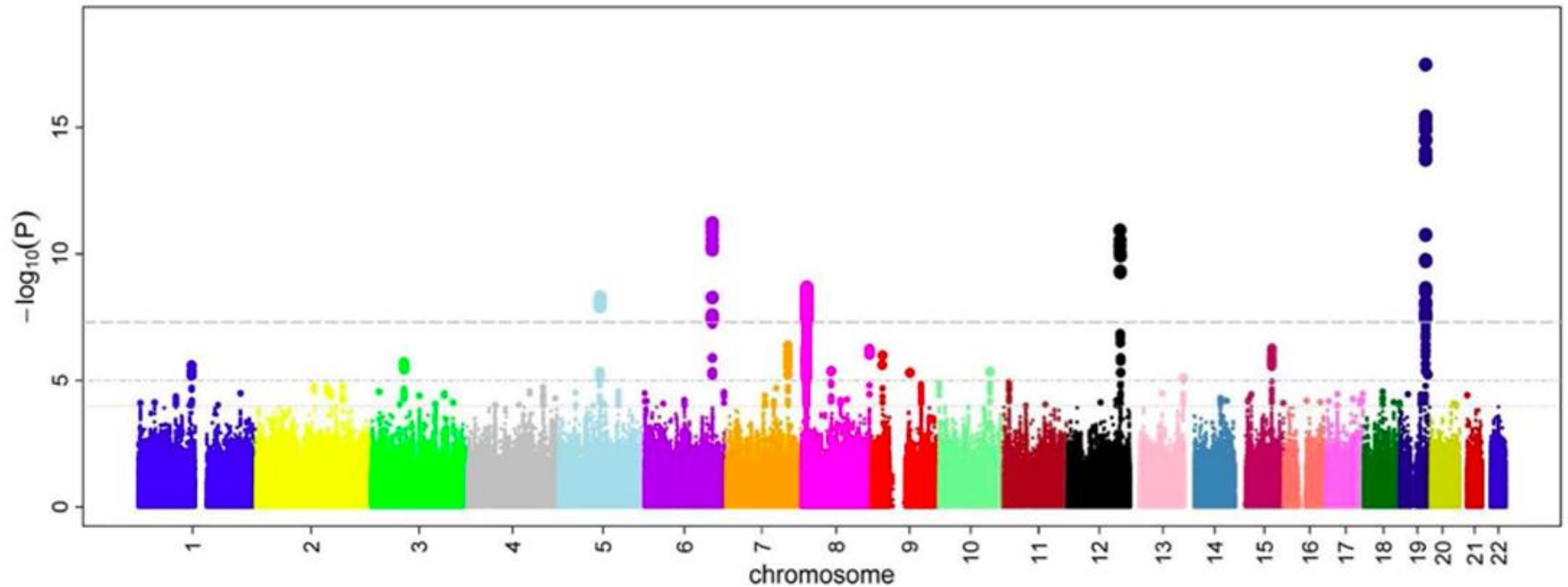
Correction for multiple testing



- For 1M tests, by chance alone we expect to see 50,000 'significant' associations at $p < 0.05$

- $p < .05$ not stringent enough in this situation
- Genome-wide significance $\approx p < 0.00000005$ (5×10^{-8})

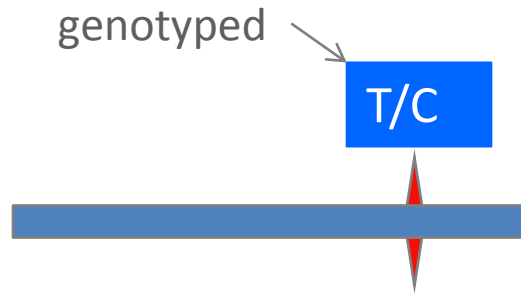
Manhattan plot showing genome-association with early microvascular disease



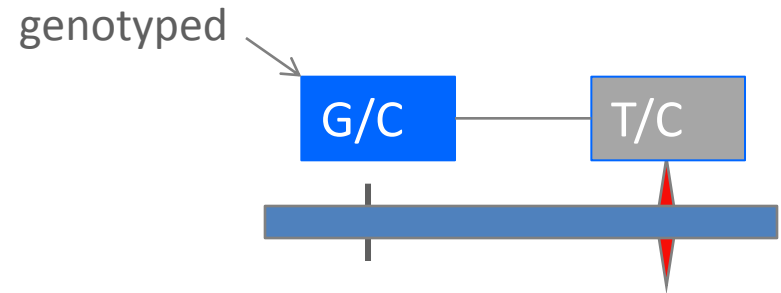
Reasons for association

○ True association (true positive)

- Causal variant (direct)

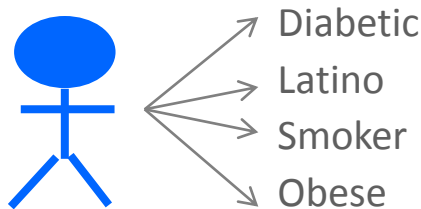


- Linked to causal variant (indirect)



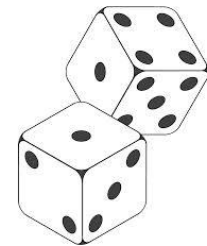
○ False association (false positive)

- Spurious (confounding/bias)



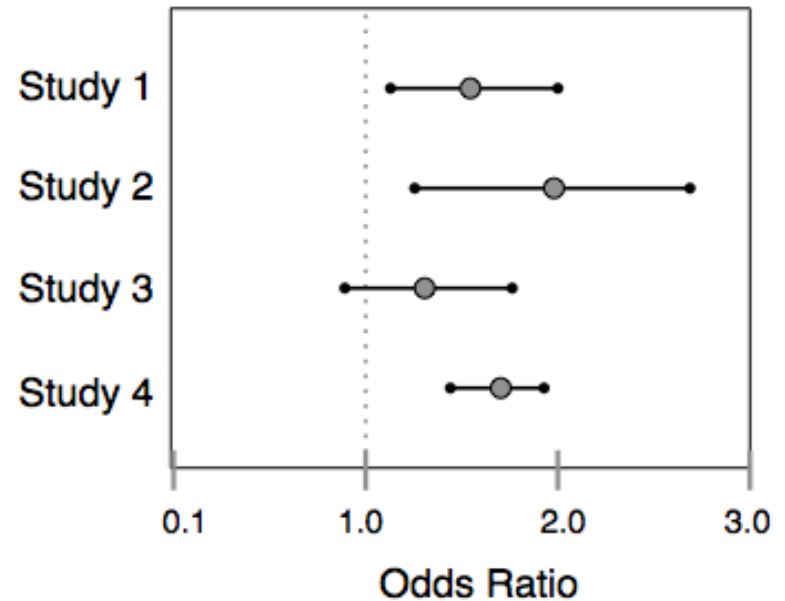
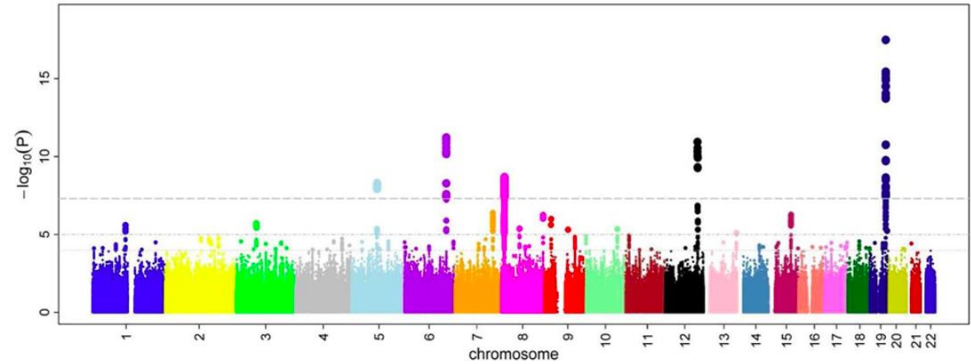
Association with
wrong trait

Chance



Properties of a valid association

- ✓ Not due to chance
- ✓ Free of bias
- ✓ Reproducible



Question

The role of P values in GWAS is to:

- A. Guard against confounding and bias
- B. Guard against chance associations
- C. Both

Answer

B. Guard against chance associations.

You can have a statistically significant result that is still confounded or otherwise biased.

Epidemiologically-sound study design is the best guard against bias and confounding.