Module 5C: A Non-Parametric Test

Odds Ratio, RR, GWAS

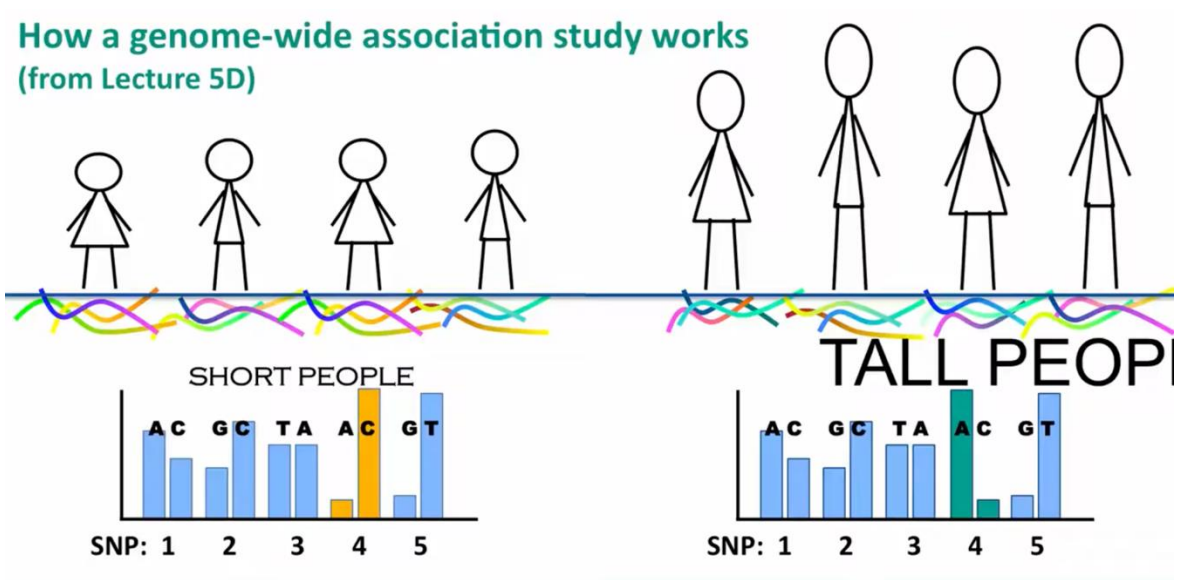
Agenda:

- Odds ratio
- Relative Risk
- **Genome-Wide Association Studies**

Genome-**W**ide **A**ssociation **S**tudies (GWAS)

- Each SNP test is just another 2×2 comparison, but repeated hundreds of thousands of times.
- Associations are tested by comparing the frequency of each allele in cases and controls
  - You can extend this analysis. For instance, the frequency of each of 3 possible genotypes can also be compare. We'll see an example of 2 (allele counting) and 3 (genotype counting).
  - *Always remember: GWAS doesn't assume that the SNP itself is causing the disease, but that it is located close enough to the causative allele that they have tight linkage (they are not usually separated by recombination).*
  - *Here is a short YouTube video from an 'old' series called "Useful Genetics" that has a fantastic introduction to GWAS*

https://www.youtube.com/watch?v=5sgPkRXR6pE



← From YouTube series by Rosie Redfield called "**Useful Genetics**"

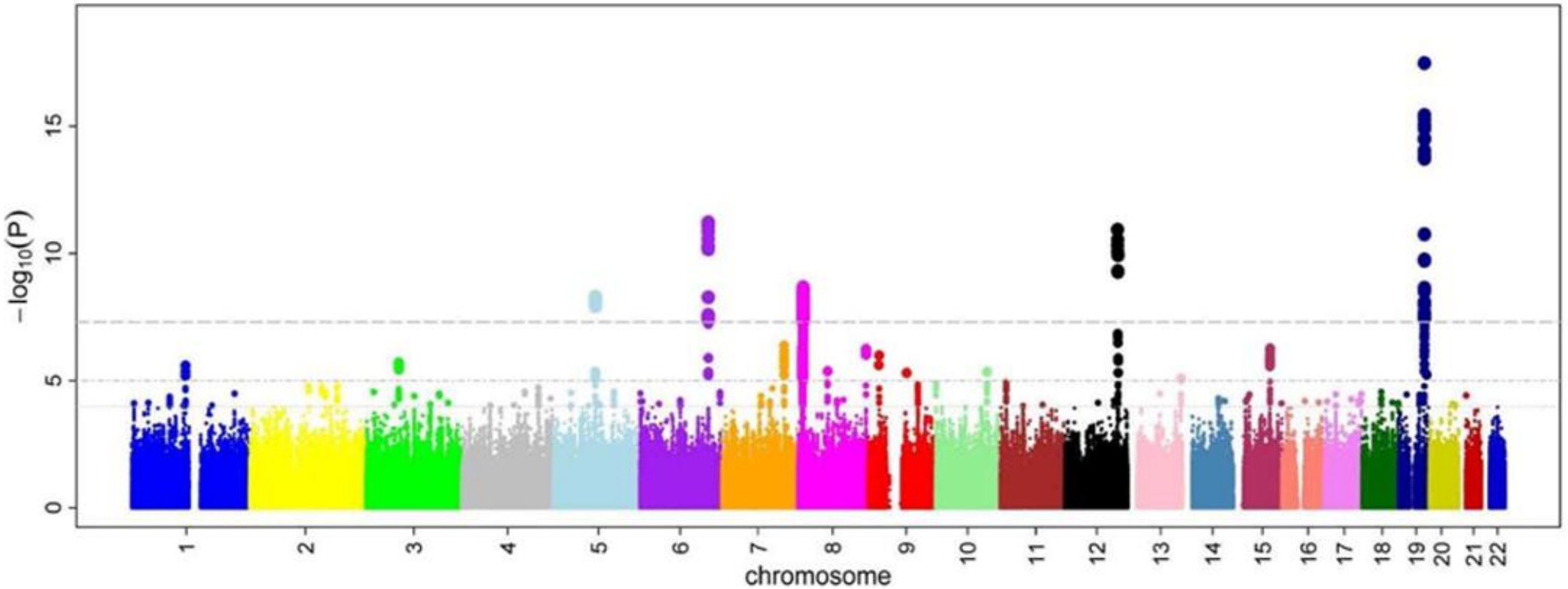


Image from Wikipedia: M. Kamran Ikram et al - Ikram MK et al (2010)

GWAS as odds ratio

Odds ratio = odds(event | exposure)
odds(event | lack of exposure)

Example:

P (D | genotype “AT”) = 0.75

P (D | genotype “TT”) = 0.25

GWAS Question:

OR for getting the disease with genotype AT compared to TT?

OR = (0.75 / 0.25) / (0.25 / 0.75) = 9

What’s the OR for AT individuals relative to an average population risk of 10%?

OR = (0.75 / 0.25) / (0.10/ 0.90) = 27

GWAS and Odds Ratio

Association of rs1234567 with some type of cancer

	CC	CT	TT
Cases	250	375	150
Controls	460	940	500

* C is associated with risk allele

$\mathbf{OR_{TT}}$ = odds(disease|TT)/odds(disease|TT) = 1

$\mathbf{OR_{CT}}$ = odds(disease|CT)/odds(disease|TT) = $\frac{375/940}{150/500}$ = 375*500/150*940 = 1.33

$\mathbf{OR_{CC}}$ = odds(disease|CC)/odds(disease|TT) = $\frac{250/460}{150/500}$ = 250*500/460*150 = 1.81

(These are all taken with respect to the lowest risk genotype, TT; they cannot be applied to an individual. To convert this into risk estimate, the prevalence of the disease and the genotypic frequencies must be considered).

GWAS and Odds Ratio

Association of rs1234567 with some type of cancer

	C*	T
Cases	875 (56.5)	675 (43.5)
Controls	1860 (48.9)	1940 (51.1)

* C is associated with risk allele

$$\mathbf{OR_C} = \frac{\text{odds(disease|C)}}{\text{odds(disease|T)}} = \frac{875/1860}{675/1940} = \frac{875*1940}{1860*675} = 1.35$$

Cases: C alleles = 2* 250 (CC) + 1*375 (CT) = 875

T alleles = 2*150 (TT) + 1*375 (CT) = 675

Controls: C alleles = 2* 460 (CC) + 1*940 (CT) = 1860

T alleles = 2*500 (TT) + 1*940 (CT) = 1940

(Still not useful for individual prognostications)

Odds Ratio to Probabilities

P(Disease) = prevalence = P(Dis|CC)P(CC) + P(Dis|CT)P(CT) + P(Dis|TT)P(TT)

$$= \frac{250}{710}\left(\frac{710}{2675}\right) + \frac{375}{1315}\left(\frac{1315}{2675}\right) + \frac{150}{650}\left(\frac{650}{2675}\right) = 0.290$$

Combining Bayes' with population genotype frequency (HapMap or BioBank) and disease prevalence for that population information!

P(Disease|CC) = 0.352

P(Disease|CT)= 0.285

P(Disease|TT) = 0.231

OR_{CC} = odds(disease | CC) / odds(disease | TT) = (0.352/ (1-0.352)) / (0.231/(1-0.231)) =1.81

OR^*_{CC} = odds(disease | CC) / odds(disease in avg pop) = (0.352/ (1-0.352)) / (0.290/(1-0.290)) =1.33

These OR*s are relative to the average population - can be directly applied to an individual

GWAS Another worked example

SNP	Associated with Risk Allele	Cases with Allele	Cases without	Controls with Allele	Controls without
rs101	A	40	60	20	80
rs202	G	70	30	50	50
rs303	T	25	75	25	75
rs404	C	10	90	20	80

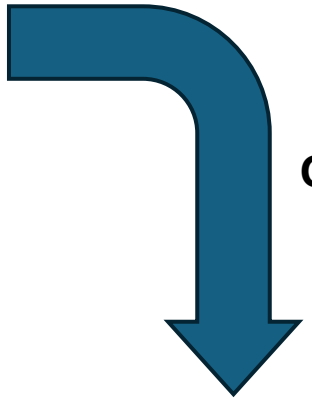
GWAS

SNP	Risk Allele	Cases with Allele	Cases without	Controls with Allele	Controls without
rs101	A	40	60	20	80
rs202	G	70	30	50	50
rs303	T	25	75	25	75
rs404	C	10	90	20	80

OR=(a/b)/(c/d)=(ad)/(bc)

a= cases with allele
b= cases without
c= controls with allele
d= controls without

SNP	Risk Allele	Cases with Allele	Cases without	Controls with Allele	Controls without
rs101	A	40	60	20	80
rs202	G	70	30	50	50
rs303	T	25	75	25	75
rs404	C	10	90	20	80



OR=(a/b)/(c/d)=(ad)/(bc)

SNP	Cases (a,b)	Controls (c,d)	OR	Interpretation
rs101	40, 60	20, 80	(40×80)/(60×20)=**2.67**	Risk allele A roughly doubles odds of disease
rs202	70, 30	50, 50	(70×50)/(30×50)=**2.33**	G allele modestly increases disease risk
rs303	25, 75	25, 75	(25×75)/(75×25)=**1.00**	No association, frequencies equal
rs404	10, 90	20, 80	(10×80)/(90×20)=**0.44**	C allele may be protective (less common in cases)

You could compute 95% Confidence Interval.....

As always: correlation ≠ causation