Multiple Testing
GWAS

Christian Werner

(Computational and quantitative geneticist) **EiB CIMMYT**, Texcoco (Mexico)

Filippo Biscarini

HerrFalloppio

(Biostatistician, bioinformatician and quantitative geneticist) **CNR-IBBA**, Milan (Italy)

Oscar González-Recio

OscarGenomics

(Computational biologist and quantitative geneticist) **INIA-UPM**, Madrid (Spain)

simple testing

- **inference** → is there a **difference** between groups?
  - e.g. AA vs AB vs BB
- **significance** is related to the **size** and **variance** of this difference
- ***p*-value**: likelihood of the data under H_0 (no difference)
  - small p-value → small likelihood of the data under H_0 → significant difference
  - large p-value → there is a high chance of observing these data if there is no difference between groups
- $\alpha = 0.05$ → threshold: 5% of rejecting H_0 when it is true (Type I error).
  - **false positive**: significant result when there is no difference (H_0 is true)

multiple testing

- **many tests → many false positives**
  - e.g. 2000 (independent) tests, α=0.05 → How many expected false positives?

    100 false positives by chance alone
- multiple testing problem

- GWAS:
  - many SNPs, many statistical tests, many p-values

How to cope with the problem

- Increase the sample size
  (e.g. Bio Banks)
- Reduce the number of tests
  - Based on LD
  - Choose relevant regions (functional analysis)

- Decrease the significance threshold
  - **Bonferroni correction**
  - **False discovery rate**
  - ***q* values**
  - **Go Bayesian...**

Bonferroni correction

- Bonferroni, mathematician (1892 - 1960)
- **adjust** the significance threshold:
    - **New significance threshold $\leqslant \alpha/m$**

      [m: number of tests (markers)]

- Bonferroni correction tends to be too conservative
    - few false positives
    - many false negatives

False discovery rate (FDR)

- Decrease the significance threshold

0.010
0.025
0.026
0.031
0.042
0.049
0.050
0.065
0.078
0.101
0.125
0.128
...

List of ordered p-values

1) If I apply a threshold alpha to decide on significance, how much can I trust the results?
2) Where should I draw a line (threshold) of significance so that at most e.g. 10% of results are false positives?

False discovery rate (FDR)

- **FDR:** how many of the positive results are false positives?

- Benjamini & Hochberg (1995), Storey (2002), Storey & Tibshirani (2003)

- **Significance level = 0.05** → 5% of **all** tests on average will be false positives (assuming independency)
- **FDR = 0.05** → 5% of **significant** tests will on average be false positives

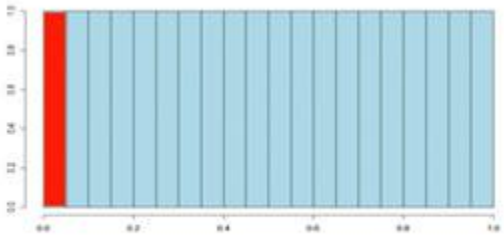
False discovery rate (FDR)

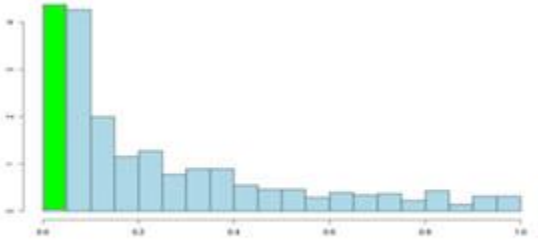
FDR

		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$	F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$		

q-values

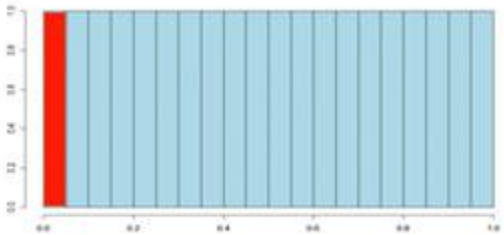
- q-values: proxies for FDR based on the **distribution of p-values**

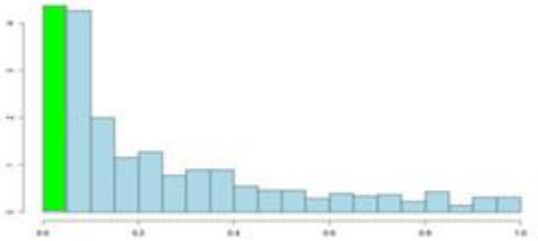




q-values

- q-values: proxies for FDR based on the **distribution of p-values**



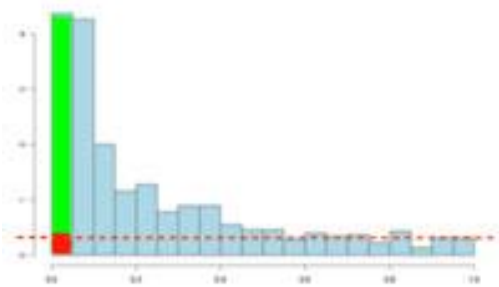


no significant differences

significant differences

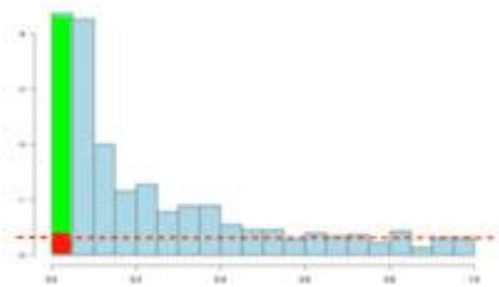
q-values

- the q-value approach tries to find the proportion of significant results which are likely to be false positives
- intuitively, it finds the height (density) at which the distribution of p-values flattens out

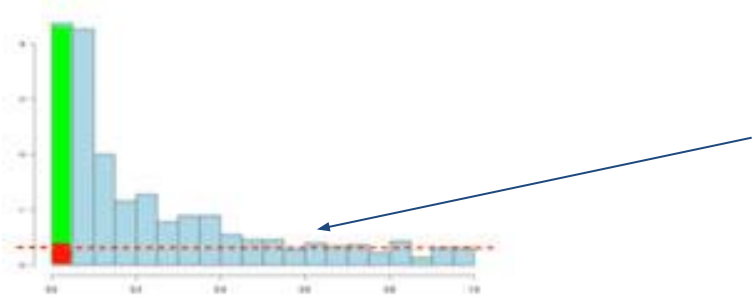


q-values

- the q-value approach tries to find the proportion of significant results which are likely to be false positives
- intuitively, it finds the height (density) at which the distribution of p-values flattens out



q-values



here the distribution is similar to the case where there is no actual difference

- this proportion of false positives is then incorporated in the calculation of adjusted p-values (**q-values**)

interpretation of q-values

- *Significance level* = 0.01 → probability of the p-value under H_0
- q-value = 0.02 → probability of the SNP being a false positive

interpretation of the single SNP

- *Significance level* = 0.01 → 1% chance of false positives (e.g. 7900 SNPs → 79 false positives expected)
- q-value = 0.02 → 2% of positive results may be false positives (e.g. 800 SNPs with q-value ≤ 0.02 → 16 false positives expected)

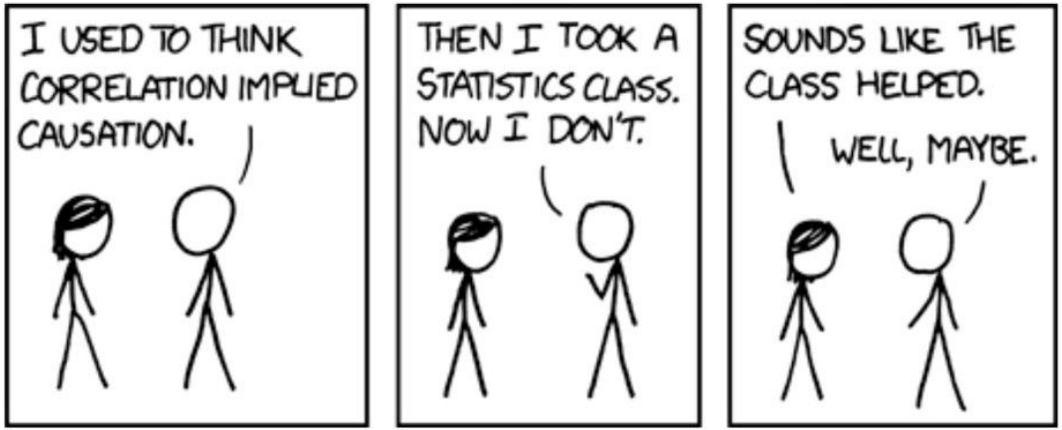
interpretation of the distribution of SNPs

q-values

- What's **wrong** with **q-values**?
  - They assume p-value is the probability of rejecting the null hypothesis when it is true
  - They do not consider that p-values are drawn from a probability distribution, and assume an infinite repetition of the experiment (obtaining different p-values for each experiment).

REMEMBER

- Correlation does not imply causation



https://xkcd.com/552/

Make your rationale choice

NEXT LECTURE

Power of GWAS experiments