P-values, false discovery rates and fold-change cutoffs

Gordon Smyth

Bioinformatics Division
Walter and Eliza Hall Institute of Medical Research

1

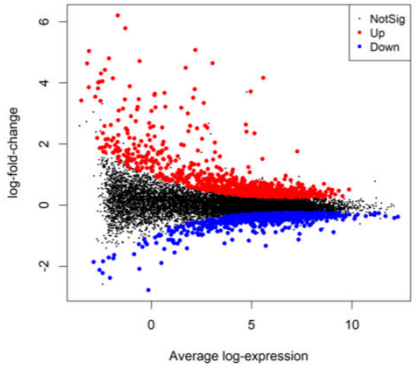
FDR and fold-changes

- Multiple testing and FDR
- A Bayesian interpretation for Benjamini-Hochberg FDR
- Why Benjamini-Hochberg FDRs can't be combined with fold-change cutoffs
- Why empirical Bayes statistical tests make fold-change cutoffs unnecessary

2

2

Genome-wide differential expression analyses



3

3

Adjusting p-values for multiple testing

In genomic research, we report on tens of thousands of statistical tests (at very least) in each paper.

Doing thousands of tests at 5% significance level would lead to an unacceptable number of false positives.

Hence we need to adjust the p-values to account for the number of tests done.

4

4

Family-wise error rate (FWER)

We could control the probability of any false positive amongst *n* tests. OK for a modest number of tests.

To control the FWER, sort p-values from smallest to largest and adjust:

	p_1	p_2	$\cdots$	p_n	
Bonferroni:	$\times n$	$\times n$	$\cdots$	$\times n$	Strong
Holm:	$\times n$	$\times (n-1)$	$\cdots$	$\times 1$	Strong
Simes:	$\times n$	$\times \frac{n}{2}$	$\cdots$	$\times 1$	Weak

5

5

False discovery rate

J. R. Statist. Soc. B (1995)
57, *No.* 1, *pp.* 289–300

Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing

By YOAV BENJAMINI† and YOSEF HOCHBERG
Tel Aviv University, Israel

Unlike family-wise error methods, FDR tolerates a few false discoveries but controls the proportion of them

6

6

Interpreting FDR as adjusted p-values

- No p-values in BH's paper, they took an entirely hypothesis testing approach.
- So I re-interpreted BH's method in terms of adjusted p-values and contributed code for the p.adjust() function in R.
- BH FDR adjusted p-values turn out to almost the same as Simes adjusted p-values but with an additional monotonicity step.

7

7

FDR can be controlled over the long term

- The FWER is not scalable – one can't control it over a career, or even over a large experiment.
- FDR is scalable: controlling FDR for each individual experiment also gives longer term assurance

8

8

askville.amazon.com

Question:

Can you explain the concept of "p-value" to me in simple English?

Answer:

The p-value is the probability that your null hypothesis is actually correct.

9

9

askville.amazon.com

Question:

Can you explain the concept of "p-value" to me in simple English?

Answer:

The p-value is the probability that your null hypothesis is actually correct.

No!

10

10

Probability that null hypothesis is true

p-value

prior

$$P(H_0 \mid \text{data}) = \frac{P(\text{data} \mid H_0)P(H_0)}{P(\text{data})}$$

marginal distribution

11

11

Testing for differential expression

Let's suppose we are testing n genes for differential expression in an RNA-seq or microarray experiment.

The genes are ranked by p-value and we are assessing whether the i^{th} gene is DE.

H_0 = gene i is non-DE

p_i = p-value

data = top i p-values are <= p_i

12

12

From definition of p-value

$$P(\text{data} \mid H_0) = p_i$$

In a well-designed experiment, most genes should be non-DE, so

$$P(H_0) \approx 1$$

We are considering the top *i* genes out of *n* genes, so

$$P(\text{data}) = i / n$$

13

13

Testing for differential expression

$$P(H_0 \mid \text{data}) = \frac{P(\text{data} \mid H_0)P(H_0)}{P(\text{data})} \leq \frac{p_i}{i/n}$$

which is the Benjamini-Hochberg adjusted p-value

The BH FDR is (an upper bound for) the posterior probability that the gene is not DE!

14

14

FDR and gene ranking

- BH FDR requires genes to be ranked by p-value
- If the genes are reordered or filtered (post BH) then the FDR calculations no longer hold

Reordering the genes by fold-change, or applying a fold-change cutoff, may invalidate the FDRs

15

15

Example of applying FDR and fold-change cut-offs

Suppose we apply both FDR and fold-change (FC) criteria simultaneously for an RNA-seq analysis,
e.g., FDR < 0.05 and |logFC| > 3

Gene	logFC	p-value	FDR
Agr3	-2.9	8.4e-06	0.0478
Pthlh	-4.1	2.2e-05	0.0478
Tslp	5.1	2.3e-05	0.0478
Smc2	-3.5	2.9e-05	0.0478
Wwc1	2.7	4.8e-05	0.0478
Six2	3.8	7.8e-05	0.0478
Ddah1	-4.4	3.1e-04	0.0755

Gene	logFC	p-value	FDR
Pthlh	-4.1	2.2e-05	0.0478
Tslp	5.1	2.3e-05	0.0478
Smc2	-3.5	2.9e-05	0.0478
Six2	3.8	7.8e-05	0.0478

Remaining p-values would not have FDR < 0.05 if BH was re-applied

Andy Chen

16

16

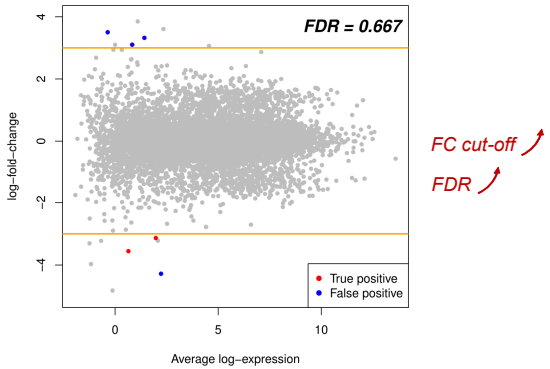
Simulate RNA-seq data

- Similar simulation setup to voom paper
- Negative binomial counts
- Two groups, n=3 vs n=3
- 10,000 genes
- 1000 DE with fold-change = 3
- NB dispersions inverse-chisquare with df=5.
- Use BH to control FDR < 0.05

17

17

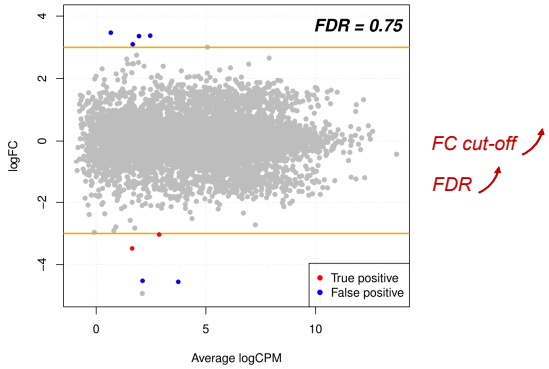
limma-voom



Andy Chen

18

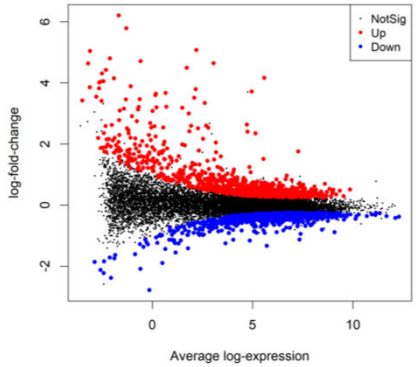
edgeR likelihood ratio test



Andy Chen

19

Fold-change cutoffs prioritize low-expressed genes



20

20

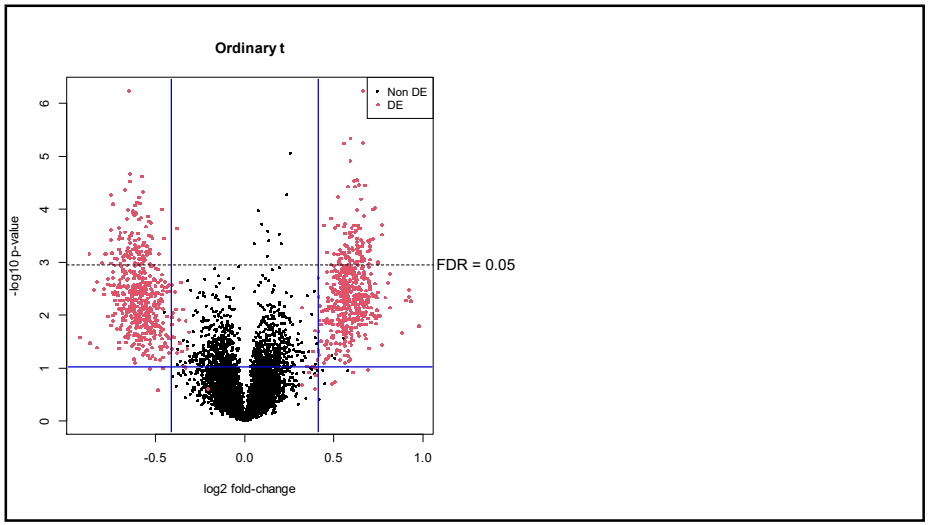
Empirical Bayes makes fold-change cutoffs largely unnecessary

To see this, conduct a simple normal simulation:

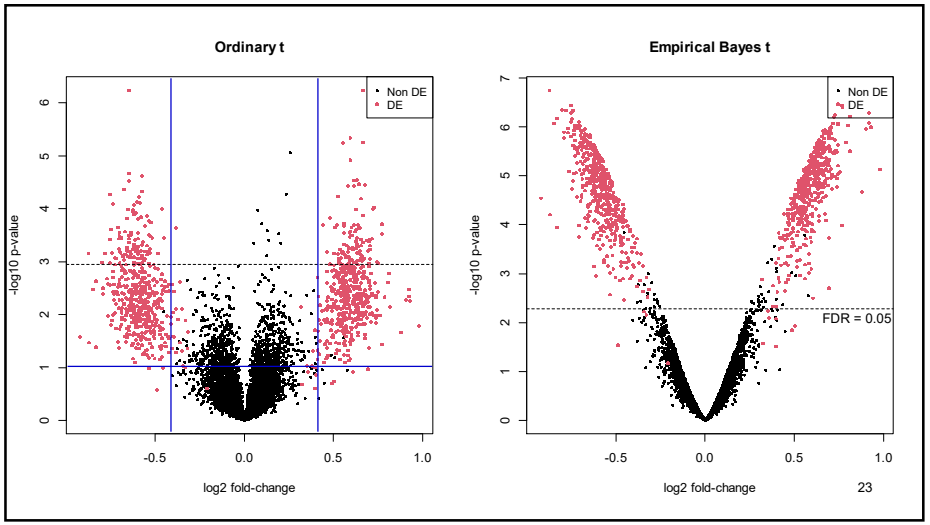
- Two groups, n=3 vs n=3
- 10,000 genes
- 1000 DE with fold-change = 1.5
- Variances inverse-chisquare with df=8

21

21



22



23

23

FDR can be integrated with a fold-change threshold by using an interval null

- We created the TREAT methods in limma and edgeR to integrate fold-change thresholding and FDR control
- Works relative to true fold-changes rather than estimated
- P-values are redefined relative to the threshold so that BH is applied to properly ordered p-values
- limma::treat() and edgeR::glmTreat()

24

24

Summary

- BH FDR is roughly interpretable as the probability that the null hypothesis is true
- BH FDR gives requires the tests to be ordered by p-values
- Applying fold-change cut-offs or reordering may invalidate the FDR calculation
- If empirical Bayes statistical tests are used, simple fold-change cut-offs tend to prioritize low-expressed genes instead of biological meaningful genes.

25

25

Acknowledgements

- Yunshun (Andy) Chen
- Davis McCarthy

26

26

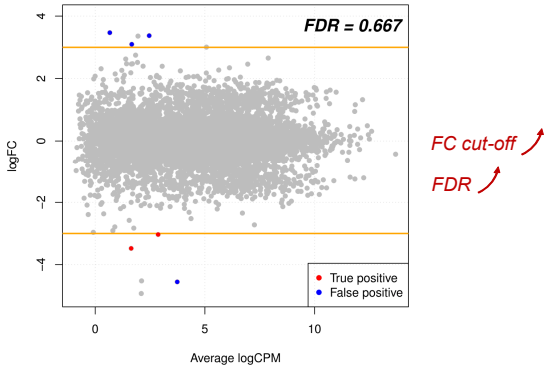
What is a fold-change anyway?

- Limma, edgeR and DESeq2 all report shrunk log-fold-changes rather than raw logFCs.
- Genes at low expression are shrunk more.
- The amount of shrinkage is tunable and changes the order of the genes.
- People apply fold-changes rules without taking into account how the fold-changes were defined.

27

27

edgeR quasi-likelihood F-test



Andy Chen

28