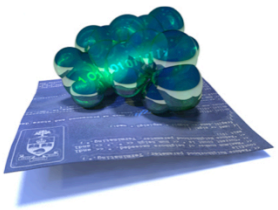
A
Bioinformatics
Course

Multiple Testing



Boris Steipe

department of biochemistry – department of molecular genetics
University of Toronto

Multiple Testing

I might bet you a beer that I can flip a fair coin and get three heads in a row. You might take on the bet, figuring that there is only a one in eight chance of me pulling this off. Then I flip the coin, get heads, flip again, get heads, flip a third time, get tails...

You think you won?

Then I flip again, and again and again ... and limiting the number of flips was never a part of the bet.

If we allow a large number of trials, the improbable becomes virtually certain.

This is the problem of multiple testing.

Translated into expression testing this means: I am looking for a differentially expressed gene and my samples are noisy. I have replicate samples to be able to average the individual measurements, and I can apply a standard statistical test (a *t*-test) to tell me if my samples are significantly different.

That's all fine for one gene, but then I look at another, and another ...

In fact I look at 20,000 genes overall. If there is **any** noise in the data, I will certainly get genes that are not differentially expressed, but show very big differences in their observed expression values.

How do we correct for this?

FWER

The **FamilyWise Error Rate** is the probability of having at least one False Positive (making at least one type I error) in a "family" of observations.

Example: Bonferroni multiple adjustment.

$$\tilde{p}_g = N \times p_g$$

If $\tilde{p}_g \leq \alpha$ then FWER $\leq \alpha$

This is simple, but very conservative

There is nothing wrong with the Bonferroni correction, it is just the most conservative approach. If we take our level of statistical significance to be 0.05, then it becomes 0.025 for two observations, 0.005 for 10 observations, 0.000005 for the 10,000 observations in a high-throughput experiment. Only the most dramatic changes will ever achieve significnce in this way.

There are other procedures that are statistically more sophisticated, still rigorously correct but less conservative and thus more powerful in not having a high percentage of false negatives. (Cf. the multtest Bioconductor package.)

FDR

The False Discovery Rate (FDR) is the proportion of False Positives among the genes called Differentially Expressed (DE).

Order the p-values for each of N observations:

$$p_{(1)} \leq \ldots \leq p_{(i)} \leq \ldots \leq p_{(N)}$$

Let k be the largest i such that $p_{(i)} \leq i / N \times \alpha$

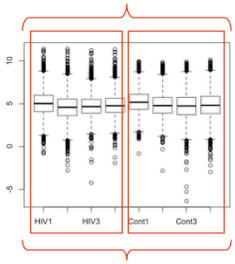
... then the FDR for genes $1 \ldots k$ is controlled at α.

Hypotheses need to be independent!

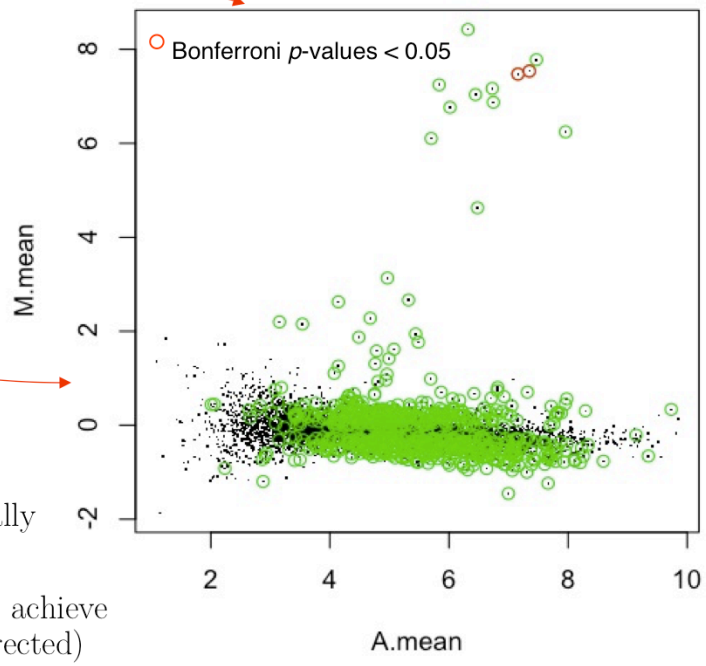
FDR: Benjamini and Hochberg (1995)

A relatively new approch to the problem is the "False Discovery Rate" – we control, not with a cutoff in probability, but by accepting a certain number of false positives. This gives us a much reduced number of false negatives.

(Hiv.mean + Cont.mean) / 2 = A.mean



Hiv.mean – Cont.mean = M.mean



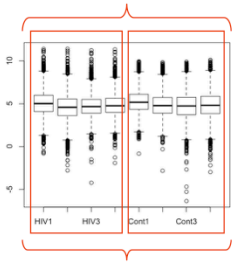
Which of the genes are "differentially expressed"?

Only two of the expression profiles achieve significance at the (Bonferroni-corrected) $p < 5\%$ level.

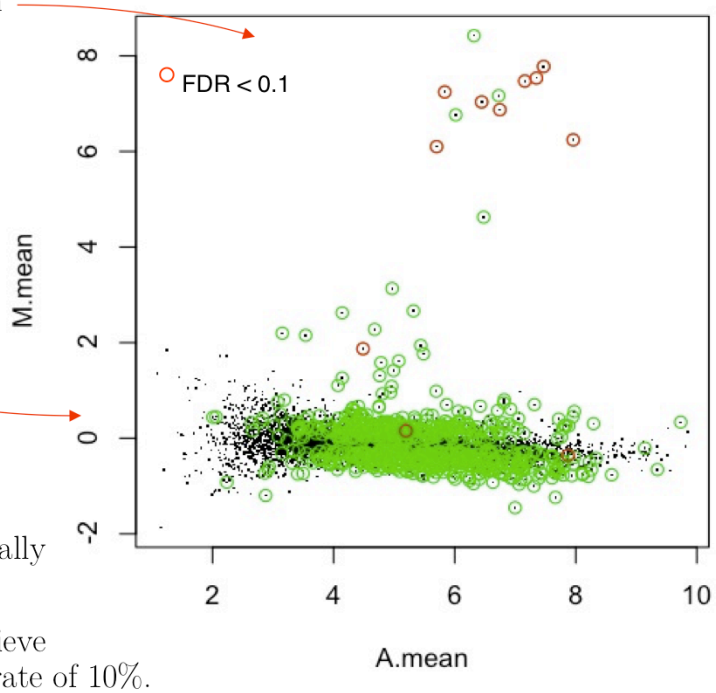
This M/A plot shows data from 7680 genes in CD4-T-cell lines at time t = 24 hours after infection with HIV type 1 virus. 4 replicates were performed for infected cells, and non-infected controls. (Data courtesy of Sohrab Shah).

Comparing *t*-tests, all green circles correspond to genes taht are significantly different at p=0.05. However, applying the Bonferroni correction, only two genes (red) achieve significance!

(Hiv.mean + Cont.mean) / 2 = A.mean



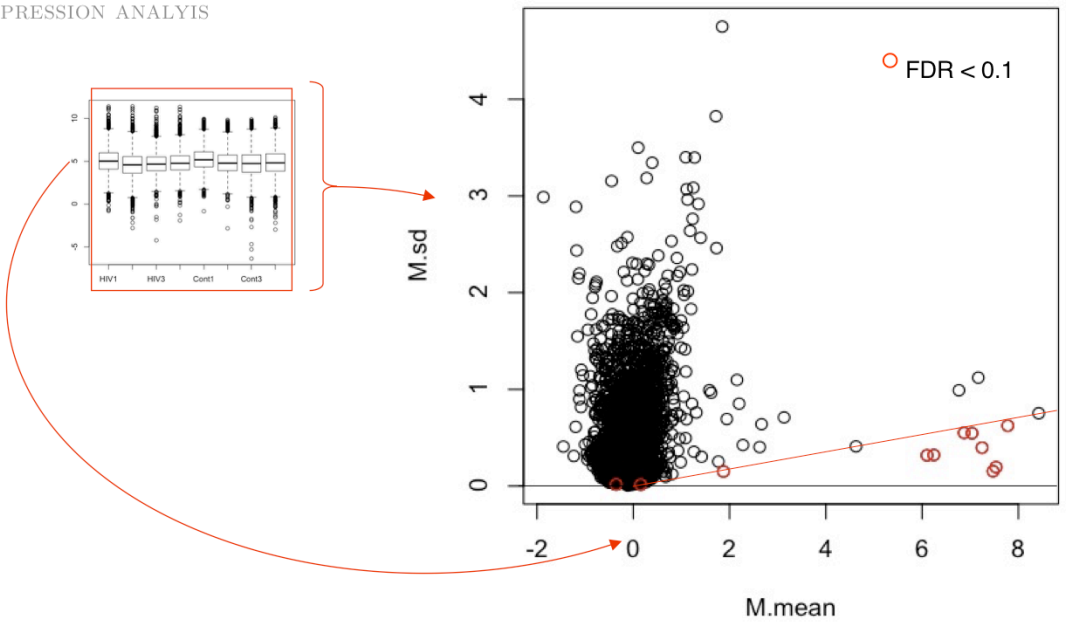
Hiv.mean – Cont.mean = M.mean



Which of the genes are "differentially expressed"?

Ten of the expression profiles achieve significance at a False-Discovery rate of 10%.

Basic Expression analyis



Considering absolute values, and standard deviations, we see that having a small standard deviation plays a critical role in determining whether a gene is considered differentially expressed.

Is this biologically meaningful?

The FDR may overemphasize genes that have small standard-deviations of their expression values, over genes that have consistently strong differences, but it is the standard procedure in the field.

http://steipe.biochemistry.utoronto.ca/abc

B O R I S . S T E I P E @ U T O R O N T O . C A

DEPARTMENT OF BIOCHEMISTRY & DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO, CANADA