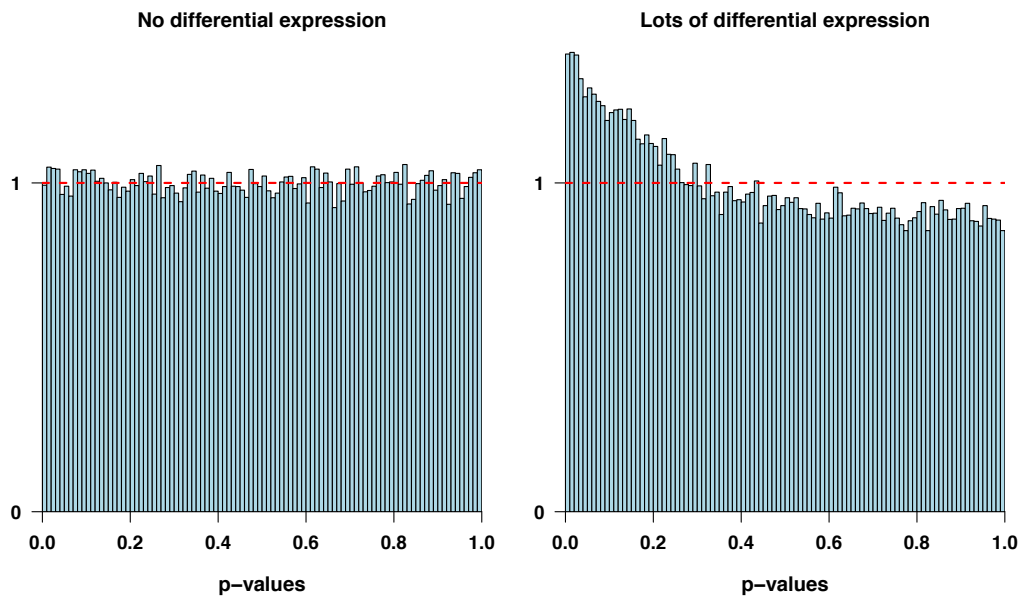


# Multiple Comparisons



Hypothetical example (no differential expression):

- ▶ Microarray with 10,000 genes.
- ▶ Calculate 10,000 p-values.
- ▶ Call genes “significant” if p-value < 0.05.
- ▶ Expected Number of False Positives:  
 $10,000 \times 0.05 = 500.$

- ▶ Many procedures have been developed to control the **Family Wise Error Rate** (the probability of at least one type I error).
- ▶ Two general types of FWER corrections:
  - ▶ Single step: equivalent adjustments made to each p-value.
  - ▶ Sequential: adaptive adjustment made to each p-value.

### Simple single step approach: Bonferroni.

- ▶ Very simple method for ensuring that the overall type I error rate of  $\alpha$  is maintained when performing  $m$  hypothesis tests.
- ▶ Rejects any hypothesis with p-value  $\leq \alpha/m$ .
- ▶ The Bonferroni adjusted p-value is

$$p_j^{Bonf} = \min \{ m \times p_j, 1 \}$$

- ▶ For example, if we want to have an experiment wide type I error rate of 0.05 when we perform 10,000 hypothesis tests, we needed a p-value of  $0.05 / 10,000 = 5 \times 10^{-6}$  to declare significance.

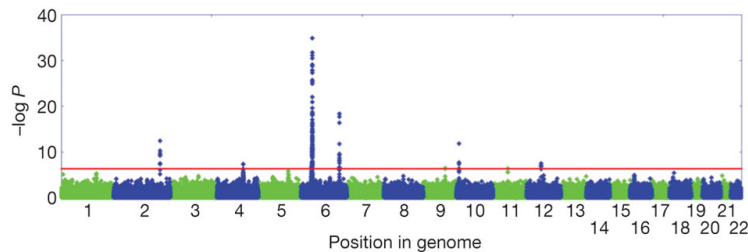
### Simple sequential method: Holm-Bonferroni.

- ▶ Order the unadjusted p-values such that  $p_1 \leq p_2 \leq \dots \leq p_m$ .
- ▶ Holm-Bonferroni uniformly delivers more power than the Bonferroni correction by testing only the most extreme p value against the strictest criterion, and the others against progressively less strict criteria.
- ▶ The Holm adjusted p-value is

$$p_j^{Holm} = \min \{ m - j + 1 \times p_j, 1 \}$$

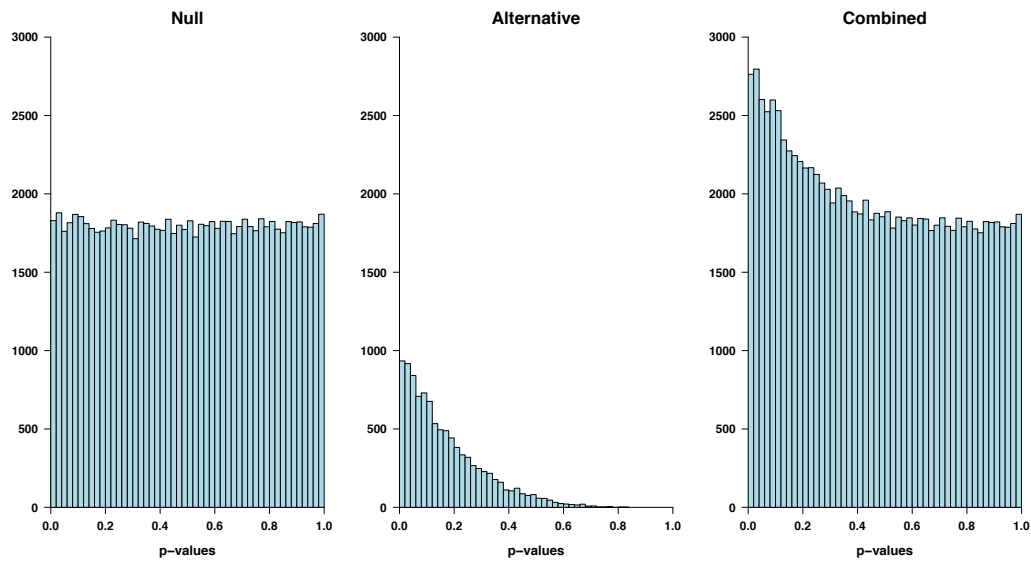
- ▶ The point here is that we do not multiply every  $p_j$  by the same factor  $m$ .

- ▶ The FWER is appropriate when you want to guard against **any** false positives.
- ▶ For example, this is usually done in genome-wide association studies.



- ▶ The general null hypothesis (that all the null hypotheses are true) is rarely of interest.
- ▶ There is a high probability of type 2 errors, i.e. of not rejecting the general null hypothesis when important effects exist.

- ▶ In many cases (particularly in genomics) we can live with a certain number of false positives.
- ▶ This is for example the case in gene expression studies, when we suspect a fair number of genes to be differentially expressed.
- ▶ In these cases, the more relevant quantity to control is the **False Discovery Rate (FDR)**.
- ▶ The FDR is designed to control the proportion of false positives *among the set of rejected hypotheses*.



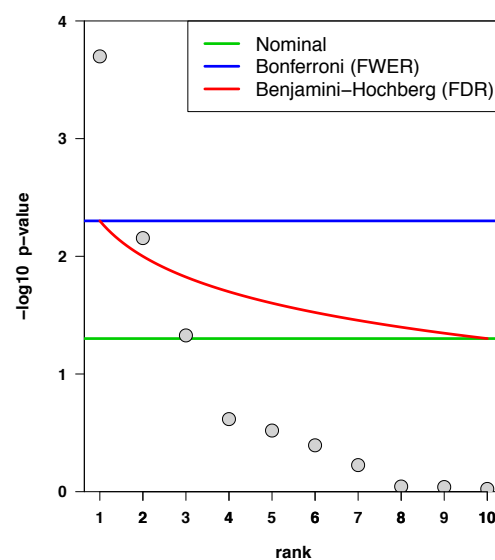
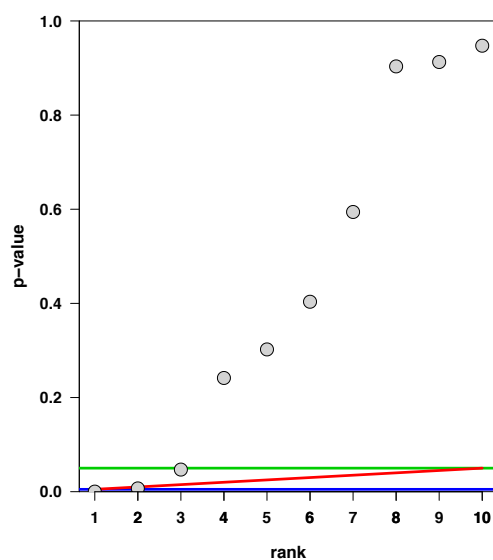
Declared ↓	$H_0$ is true	$H_a$ is true	Total
Significant	V	S	R
Non-significant	U	T	$m - R$
Total	$m_0$	$m - m_0$	$m$

- ▶ Bonferroni and such control the family-wise error rate.  
→  $V/(V+U)$ .
- ▶ The FDR controls the false positive rate.  
→  $V/(V+S)$ .

## Benjamini and Hochberg FDR.

To control FDR at level  $\delta$ :

- ▶ Order the unadjusted p-values:  $p_1 \leq p_2 \leq \dots \leq p_m$ .
- ▶ Find the test with the highest rank  $j$  for which the p-value  $p_j$  is less than or equal to  $(j / m) \times \delta$ .
- ▶ Declare the tests of rank 1, 2, ...,  $j$  as significant.



Difference in interpretation:

Suppose 550 out of 10,000 genes are significant at the 0.05 level.

- ▶ False Discovery Rate  $< 0.05$ :  
Expect  $0.05 \times 550 = 27.5$  false positives.
- ▶ Family Wise Error Rate  $< 0.05$ :  
The probability of at least 1 false positive  $< 0.05$ .
- ▶ In most settings, the latter is extremely unlikely, unless the sample size is huge!

John Storey's positive FDR (pFDR):

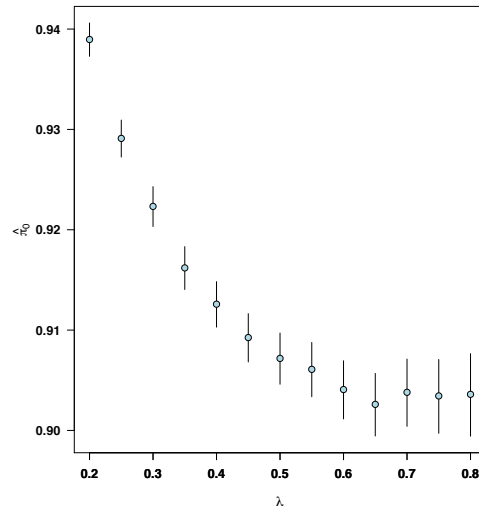
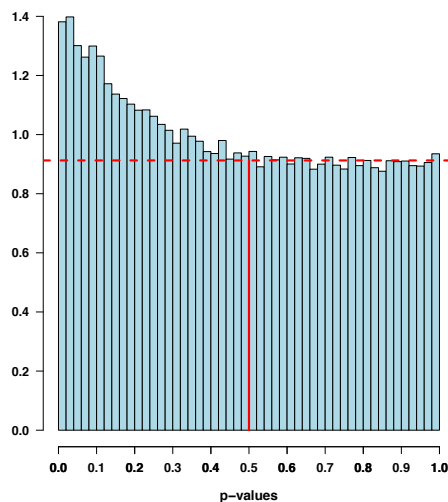
$$\text{FDR} = E \left[ \frac{V}{R} \mid R > 0 \right] \times P(R > 0)$$

$$\text{pFDR} = E \left[ \frac{V}{R} \mid R > 0 \right]$$

- ▶ Since  $P(R > 0)$  is  $\sim 1$  in most genomics experiments, the FDR and the pFDR are very similar.
- ▶ Omitting  $P(R > 0)$  facilitated the development of a measure of significance in terms of the FDR for each hypothesis.

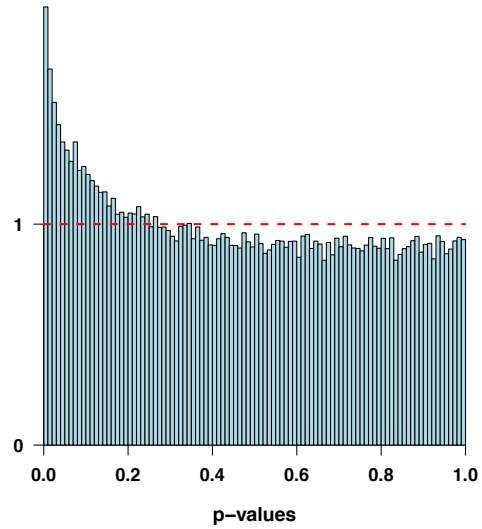
## Q-values:

- ▶ The q-value is defined as the minimum FDR that can be attained when calling that gene significant (i.e., expected proportion of false positives incurred when calling that gene significant).
- ▶ The estimated q-value is a function of the p-value for that test and the distribution of the entire set of p-values from the family of tests being considered.
- ▶ In testing for differential expression, if a gene has a q-value of 0.10 it means that we can expect 10% of genes that show p-values at least as small as this gene to be false positives.





	p	q	p.bonf
1	1.92e-06	0.11814	0.19188
2	3.90e-06	0.11814	0.38955
3	8.43e-06	0.11814	0.84317
4	9.35e-06	0.11814	0.93497
5	1.11e-05	0.11814	1.00000
6	1.14e-05	0.11814	1.00000
7	1.14e-05	0.11814	1.00000
8	1.22e-05	0.11814	1.00000
9	1.24e-05	0.11814	1.00000
10	1.44e-05	0.11814	1.00000
11	1.44e-05	0.11814	1.00000
12	1.64e-05	0.11997	1.00000
13	1.76e-05	0.11997	1.00000
14	1.97e-05	0.11997	1.00000
15	1.99e-05	0.11997	1.00000
16	2.53e-05	0.14035	1.00000
17	2.64e-05	0.14035	1.00000
18	3.03e-05	0.14795	1.00000
19	3.11e-05	0.14795	1.00000
20	3.67e-05	0.15919	1.00000



	p	q	p.bonf
1	2.27e-31	2.05e-26	2.27e-26
2	9.05e-13	4.09e-08	9.05e-08
3	3.37e-12	1.02e-07	3.37e-07
4	2.57e-11	5.80e-07	2.57e-06
5	2.20e-09	3.97e-05	2.20e-04
6	1.92e-06	2.89e-02	1.92e-01
7	3.90e-06	5.03e-02	3.90e-01
8	5.26e-06	5.94e-02	5.26e-01
9	8.43e-06	6.99e-02	8.43e-01
10	9.35e-06	6.99e-02	9.35e-01
11	9.55e-06	6.99e-02	9.55e-01
12	1.11e-05	6.99e-02	1.00e+00
13	1.14e-05	6.99e-02	1.00e+00
14	1.14e-05	6.99e-02	1.00e+00
15	1.22e-05	6.99e-02	1.00e+00
16	1.24e-05	6.99e-02	1.00e+00
17	1.44e-05	7.22e-02	1.00e+00
18	1.44e-05	7.22e-02	1.00e+00
19	1.64e-05	7.77e-02	1.00e+00
20	1.76e-05	7.95e-02	1.00e+00

