Analysis of (cDNA) Microarray Data:

Part III. False Discoveries

False Discoveries

Setting the scene:

1. Suppose we have an instrument that will provide a quantitative measure of the expression of a certain gene with no measurement error.
2. We have developed a drug that we believe will alter the expression of the gene when the drug is injected into a frog.
3. We randomly divide a group of eight frogs into two groups of four.
4. Each rat in one group is injected with the drug. Each frog in the other group is injected with a control substance.

False Discoveries

Setting the scene:

We use out instrument to measure the expression of the gene in each frog after treatment and obtain the following results:

	Control	Drug
Expression	9 12 14 17	18 21 23 26
Average	13	22

The difference in averages is: 22 – 13 = 9.

We wish to claim that this difference was caused by the drug.

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

Setting the scene:

	Control	Drug
Expression	9 12 14 17	18 21 23 26
Average	13	22

1. Clearly there is some natural variation in expression (not due to treatment) because the expression measures differ among frogs within each treatment group.
2. Maybe the observed difference (9) showed up simply because we happened to choose the frogs with larger gene expression to be injected with the drug.

Q: What is the chance of seeing such a large difference in treatment means if the drug has no effect?

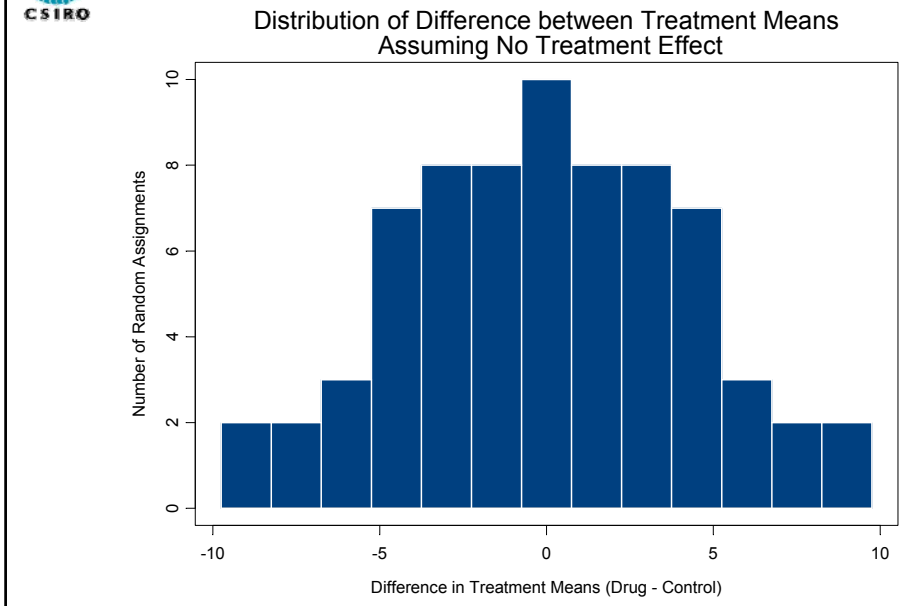
Armidale Animal Breeding Summer Course, UNE, Feb. 2006

False Discoveries

Random Assignment	Control				Drug				Difference in Averages
1	9	12	14	17	18	21	23	26	9.0
2	9	12	14	18	17	21	23	26	8.5
3	9	12	14	21	17	18	23	26	7.0
4	9	12	14	23	17	18	21	26	6.0
5	9	12	14	26	17	18	21	23	4.5
6	9	12	17	18	14	21	23	26	7.0
7	9	12	17	21	14	18	23	26	5.5
8	9	12	17	23	14	18	21	26	4.5
9	9	12	17	26	14	18	21	23	3.0
10	9	12	18	21	14	17	23	26	5.0
11	9	12	18	23	14	17	21	26	4.0
12	9	12	18	26	14	17	21	23	2.5
13	9	12	21	23	14	17	18	26	2.5
14	9	12	21	26	14	17	18	23	1.0
15	9	12	23	26	14	17	18	21	0.0
etc.....									
69	17	21	23	26	9	12	14	18	-8.5
70	18	21	23	26	9	12	14	17	-9.0

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries



Armidale Animal Breeding Summer Course, UNE, Feb. 2006

False Discoveries

P-Values

1. Only 2 of the 70 possible random assignments would have led to a difference between treatment means as large as 9.
2. Thus, under the assumption of no drug effect, the chance of seeing a difference as large as the one observed was 2/70 = 0.0286.
3. Because 0.0286 is a small probability, we have reason to attribute the observed difference to the effect of the drug rather than a coincidence due to the way we assigned our experimental units to treatment groups.
4. This is an example of a randomization test. Sir R.A. Fisher described such tests in the first half of the 20th century.
5. 2/70 = 0.0286 is a p-value which tells us about the probability of seeing a result as extreme as the one observed under the assumption that the null hypothesis (H_0) is true.
6. When p-values are small we have reason to doubt H_0
7. In our example, H_0 was that the drug had no effect on the expression of the gene.

False Discoveries

P-Values

Q: What if instead of the original data, we had observed

	Control	Drug
Expression	9 12 14 17	118 121 123 126
Average	13	122

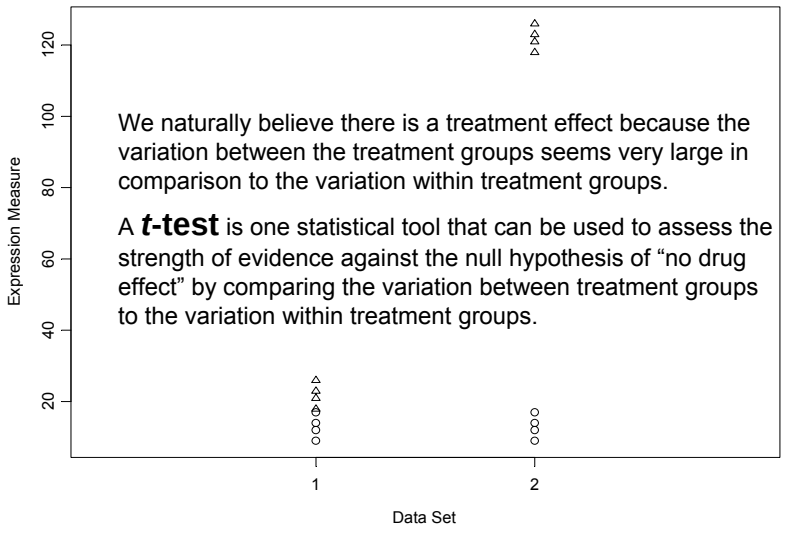
A: Our randomization test *p*-value would still be **2/70 = 0.0286**.

This seems a bit odd because most people would agree that there should be more evidence against H_0 in this new data than there was in the original data.

The reason for this belief is that people assume (perhaps without realizing it) that there should be no big gaps in the data without a drug effect.

False Discoveries

P-Values and t-test



We naturally believe there is a treatment effect because the variation between the treatment groups seems very large in comparison to the variation within treatment groups.

A ***t*-test** is one statistical tool that can be used to assess the strength of evidence against the null hypothesis of "no drug effect" by comparing the variation between treatment groups to the variation within treatment groups.

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

P-Values and t-test

➡ The case of equal variances

Clone i

Sample 1	Sample 2
y_{11}	y_{21}
y_{12}	y_{22}
...	...
y_{1n1}	y_{2n2}
$\bar{y}_1 \pm s_1$	$\bar{y}_2 \pm s_2$

Settings: $y_{1j} \sim N(\mu_1, \sigma_1^2)$ and $y_{2j} \sim N(\mu_2, \sigma_2^2)$

$H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$

$\sigma_1^2 = \sigma_2^2$

Test statistic: $t = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$

$s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

'weighted' average of s_1^2 and s_2^2

COMMENT: If $\sigma_1^2 \neq \sigma_2^2$, the *t*-test becomes:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{(\varphi)}, \quad \text{where:} \quad \varphi = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2+1}}$$

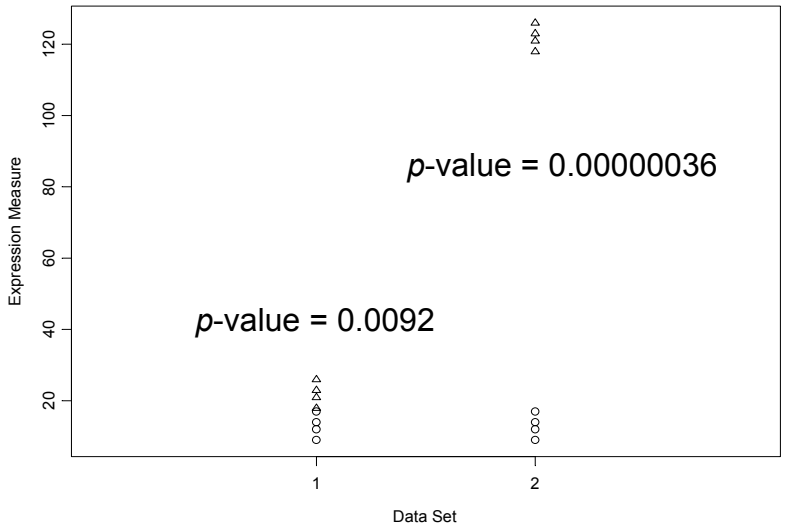
(Satterthwaite, 1946)

Source: G Rosa 2003.

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

False Discoveries

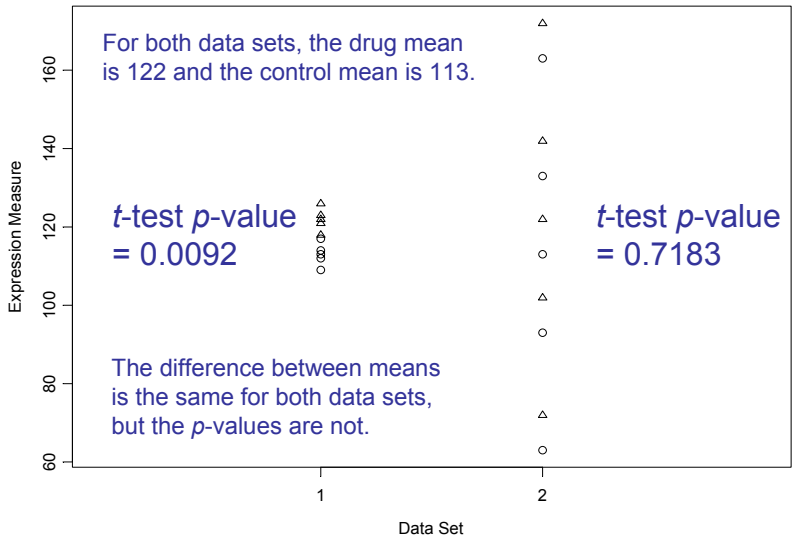
P-Values and t-test



A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

P-Values and t-test

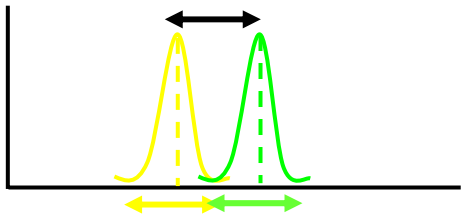


Armidale Animal Breeding Summer Course, UNE, Feb. 2006

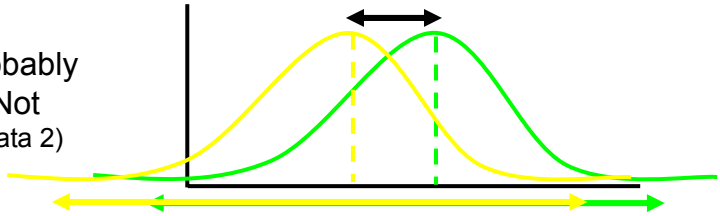
False Discoveries

P-Values and t-test

A significant Difference
(Data 1)



Probably Not
(Data 2)



False Discoveries

Biological vs Technical Replication

1. Regardless of the statistical method used, if there had been only one frog per treatment, there would have been no way to refute the idea that natural variation in expression (rather than a drug effect) was responsible for the observed difference between the drug and control.
2. Thus using more than one experimental unit per treatment is essential. This is type of replication is known in the microarray literature as ***biological replication***.
3. Although we began by assuming that we had a device that could provide a quantitative measure of a gene's expression without error, that assumption was not necessary.
4. The main point is that if biological replication is needed when there is no measurement error, it is certainly needed when there is measurement error.
5. If our measurement device measures with error, we may want to obtain multiple measures of the expression in each of our experimental units. This type of replication is know in the microarray literature as ***technical replication***.
6. Technical replication is helpful but not essential

False Discoveries

The Multiple Testing Problem

1. Suppose one test of interest has been conducted for each of m genes in a microarray experiment.
2. Let $p_1, p_2, \ldots, p_m$ denote the p-values corresponding to the m tests.
3. Let $H_{01}, H_{02}, \ldots, H_{0m}$ denote the null hypotheses corresponding to the m tests.
4. Suppose m_0 of the null hypotheses are true and m_1 of the null hypotheses are false.
5. Let c denote a value between 0 and 1 that will serve as a cutoff for significance:

   - Reject H_{0i} if $p_i \leq c$ (declare significant)

   - Fail to reject (or accept) H_{0i} if $p_i > c$ (declare non-significant)

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

U = Number of true negatives
= Power (1 – β)

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

V = Number of false positives
= Number of false discoveries
= Number of type I errors (α)

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

T = Number of False Negatives
= Number of type II errors (β)

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

S = Number of true positives
= Number of true discoveries
= Confidence $(1 - \alpha)$

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

W = Number of non-rejections
Number of H_0 accepted

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

R = Number of rejections
(of null hypotheses)

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

The Multiple Testing Problem

"Power (1 – β) plays the same role in hypothesis testing that Standard Error plays in parameter estimation"

"The practice in designing studies is to hold β at 0.20 and α at 0.05 simply because those are conventional values. The idea is that a false positive is four times as bas as a false negative"

Mood, Graybill, Boes
Introduction to the Theory of Statistics

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

False Discoveries

The Multiple Testing Problem

Table of Outcomes

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

Random Variables

Constants

False Discoveries

The Multiple Testing Problem

Table of Outcomes

Unobservable

	Accept Null Declare Non-Sig. No Discovery Negative Result	Reject Null Declare Sig. Declare Discovery Positive Result	
True Nulls	U	V	m_0
False Nulls	T	S	m_1
Total	W	R	m

Observable

False Discoveries

The Multiple Testing Problem

1. FDR was introduced by Benjamini and Hochberg (1995) and is formally defined as

   $$FDR = V/R \quad \text{if } R>0$$
   $$\text{and} \quad FDR = 0 \quad \text{otherwise.}$$

2. Controlling FDR amounts to choosing the significance cutoff *c* so that FDR is less than or equal to some desired level α.
3. Suppose a scientist conducts many independent microarray experiments in his or her lifetime.
4. For each experiment, the scientist declares a list of genes to be differentially expressed using some method.
5. For each list consider the ratio of the number of false positive results to the total number of genes on the list (set this ratio to 0 if the list contains no genes).
6. The FDR for the method used by the scientist is approximated by the average of the ratios described above.

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

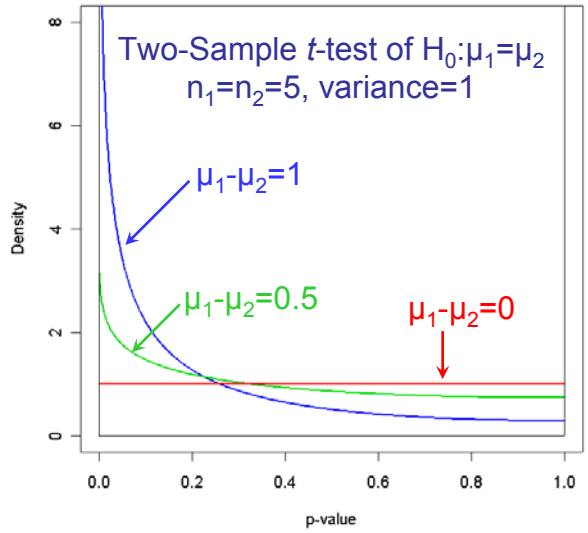
The Multiple Testing Problem

- Note that some of the gene lists may contain a high proportion of false positive results and yet the method used by the scientist may still control FDR at a given level because it is the average performance across repeated experiments that matters.
- There is no useful method that will guarantee a small proportion of false positive results in a single experiment.
- The distribution of the *p*-value is uniform on the interval (0,1) whenever the null hypothesis is true.
- The above statement is correct irrespective of the statistical test used (as long as the test is valid).
- The distribution of the *p*-value is stochastically smaller than uniform whenever the null hypothesis is false.

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

False Discoveries

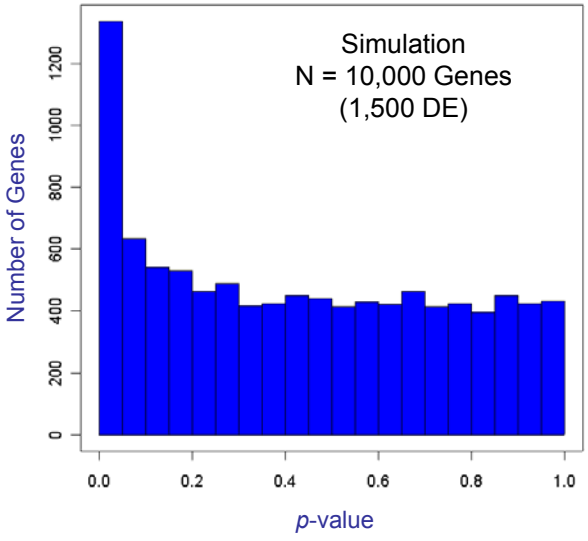
Distribution of P-Values



A Quantitative Overview to Gene Expression Profiling in Animal Genetics

False Discoveries

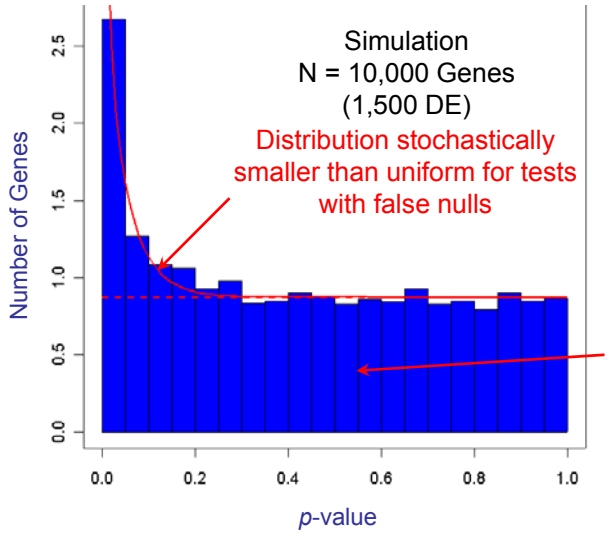
Histogram of *p*-values for a Test of Interest



Armidale Animal Breeding Summer Course, UNE, Feb. 2006

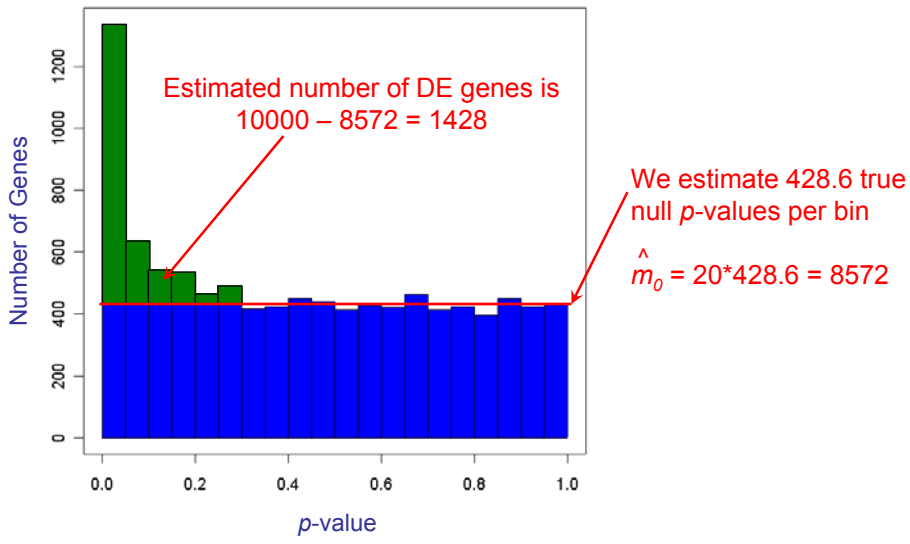
False Discoveries

Mixture of a Uniform Distribution and a Distribution Stochastically Smaller than Uniform



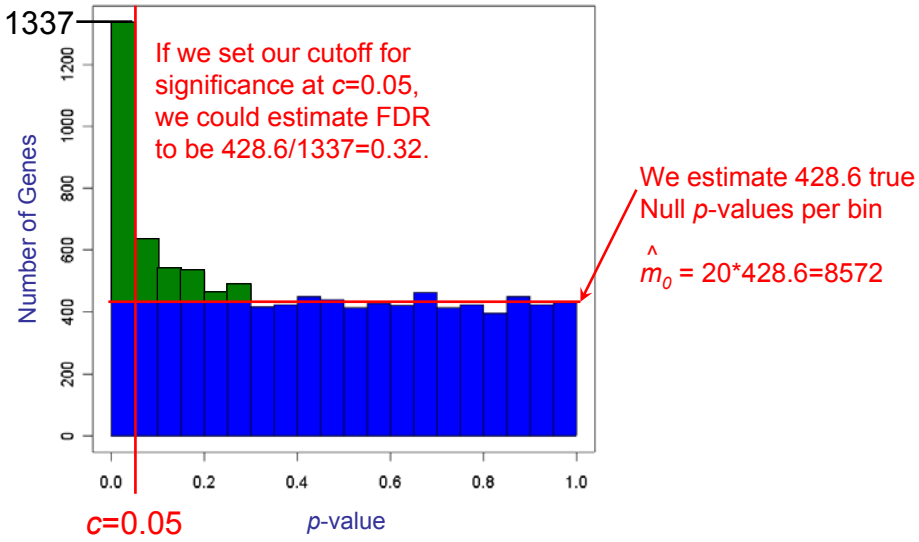
False Discoveries

Histogram of *p*-values for a Test of Interest



False Discoveries

Histogram of *p*-values for a Test of Interest



False Discoveries

Concluding Remarks

1. In many cases, it will be difficult to separate the many of the DE genes from the non-DE genes (→ Validation)
2. Genes with a small expression change relative to their variation will have a *p*-value distribution that is not far from uniform if the number of experimental units (animals) per treatment is low.
3. To do a better job of separating the DE genes from the non-DE genes we need to use good experimental designs with more replications per treatment.
4. Don't get to hung up on p-values. They only help evaluating the strength of the evidence.
5. Ultimately what matters is Biological Relevance.
6. Statistical significance is not necessarily the same as biological significance.
7. Give me enough microarrays and I'll call all genes DE.