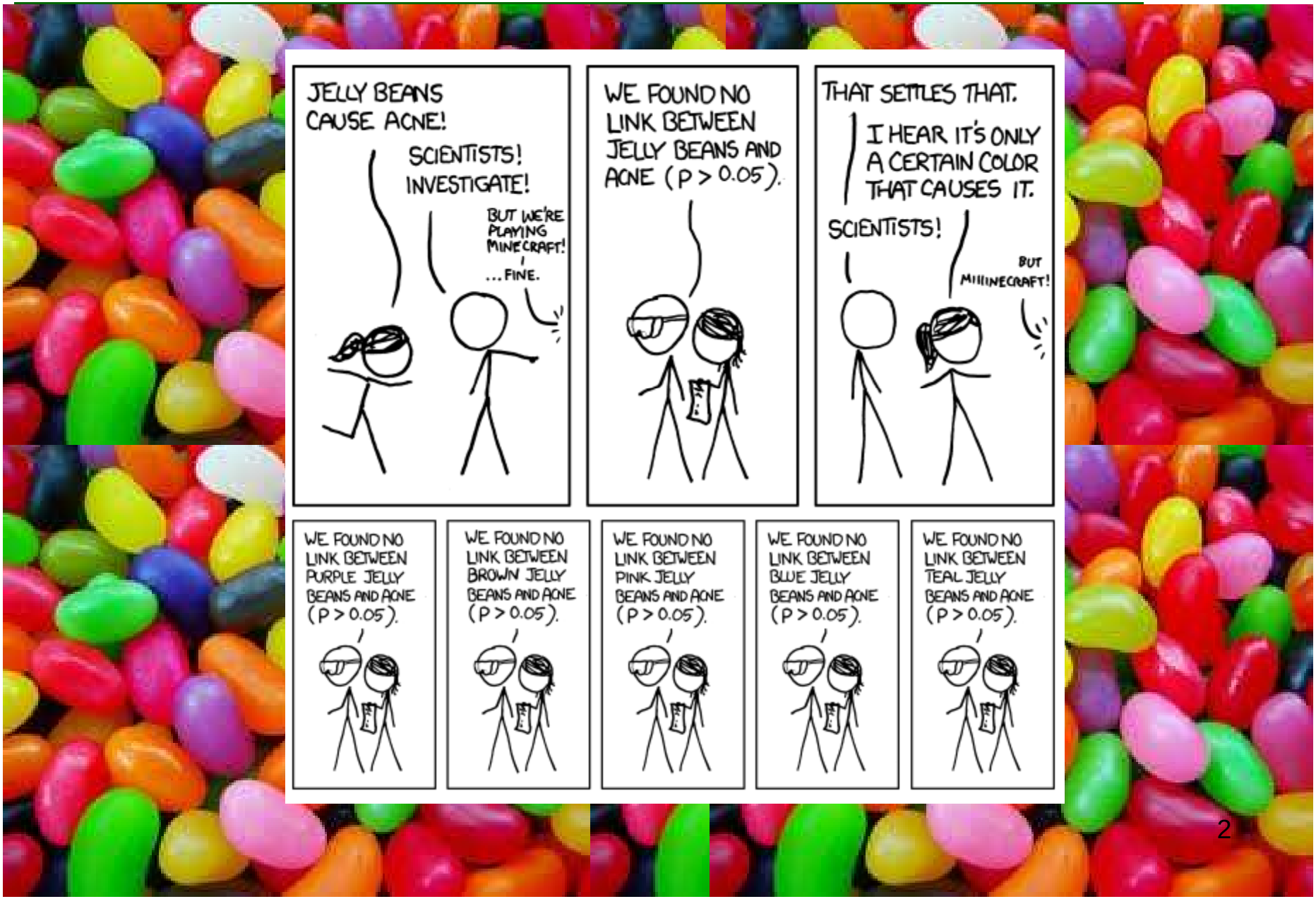
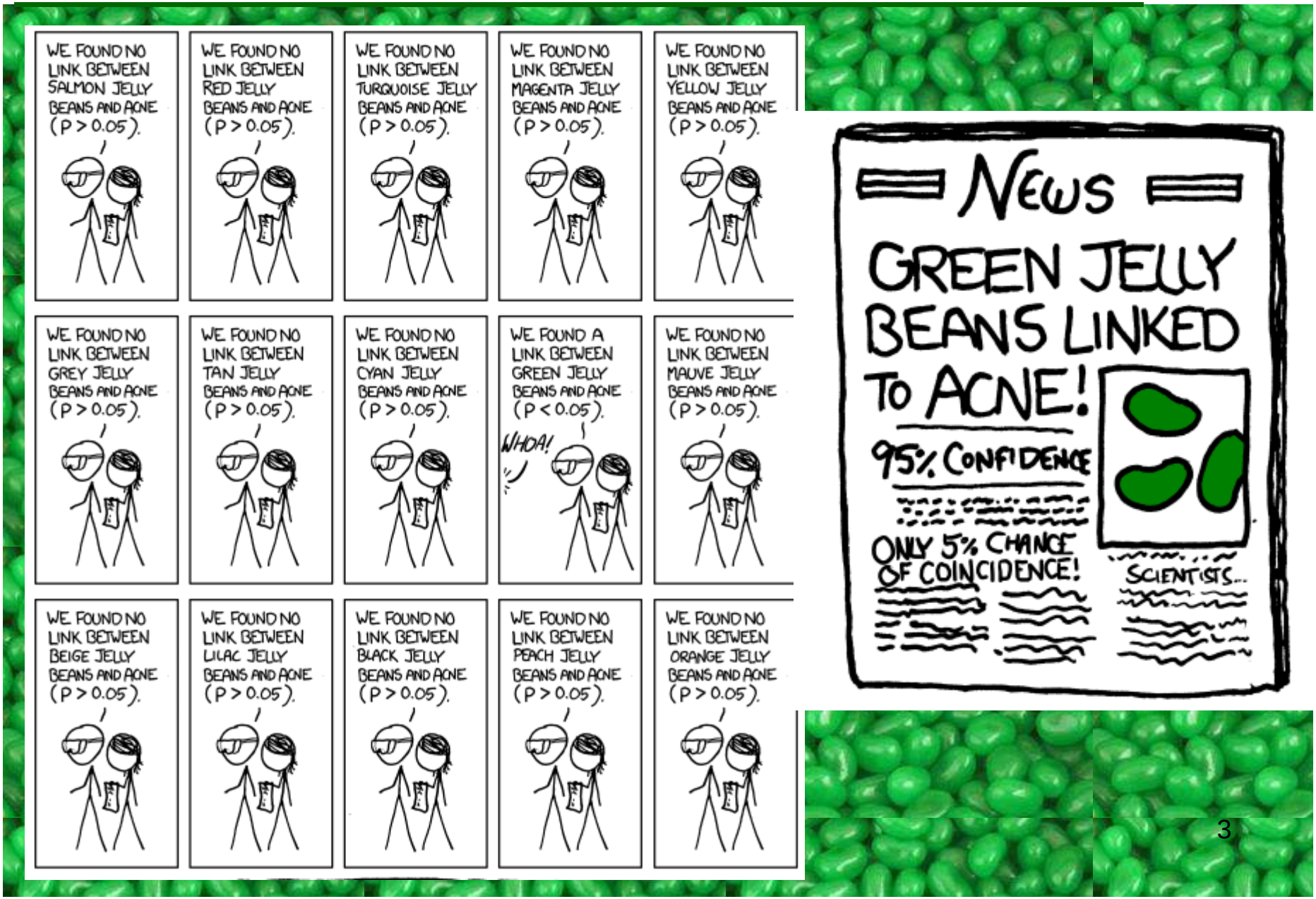
Multiple Testing in QTL Mapping

Lucia Gutierrez lecture notes
Tucson Winter Institute

Multiple Testing



Multiple Testing



Multiple Testing

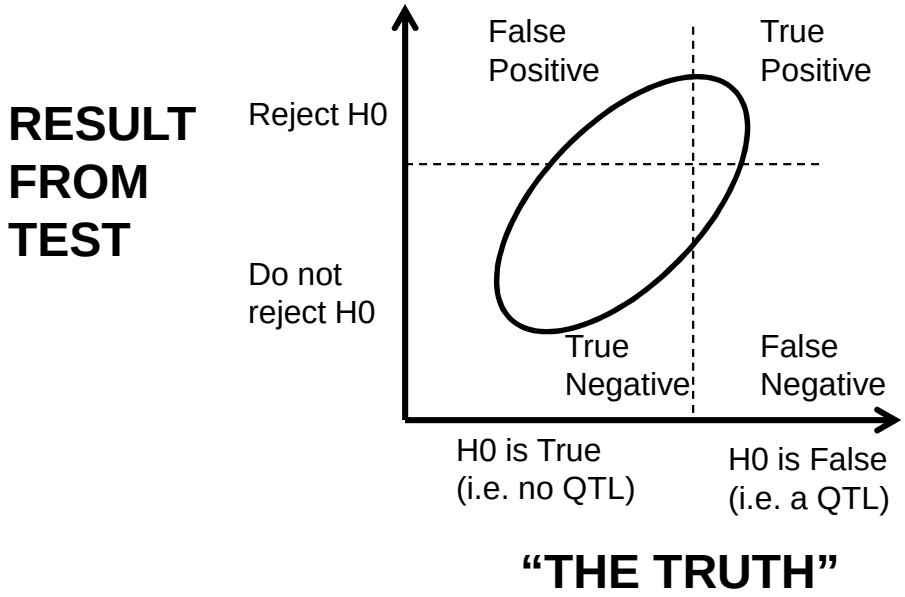
WHAT IS WRONG WITH THE PREVIOUS CARTOON?

A common practice is to use $P<0.05$ to decide about significance of a test. But with large number of tests (as is common in QTL mapping were we perform one test at each marker), the chance of having at least one false positive reaches 1 pretty quickly.

Let's look at some theory to clarify this concept.

Hypothesis Testing (in QTL context)

OUTCOMES OF A STATISTICAL TEST



FALSE POSITIVE: occurs when a QTL is incorrectly declared present

TRUE POSITIVE: occurs when a QTL is correctly declared present.

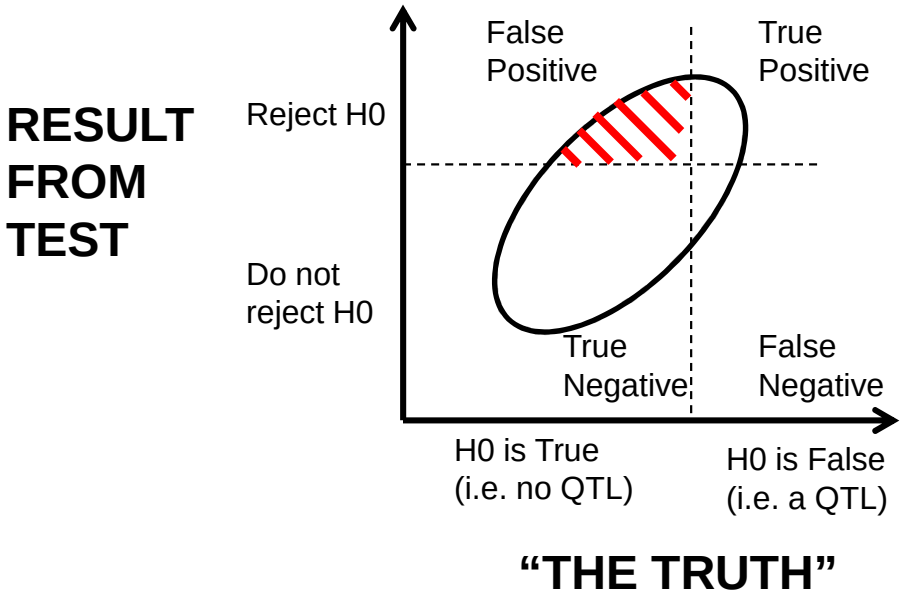
FALSE NEGATIVE: occurs when a QTL is incorrectly declared absent.

TRUE NEGATIVE: occurs when a QTL is correctly declared absent.

Type I Error

Type I error: It is the incorrect rejection of a true null hypothesis (an incorrectly declared QTL) in a given test.

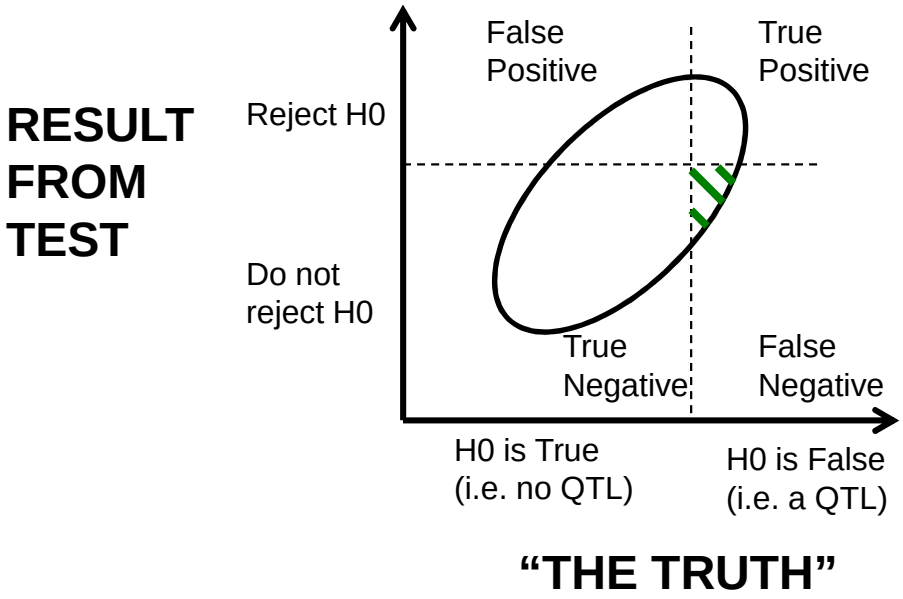
Probability of Type I error (α): It is the probability of finding a false positive (an incorrectly declared QTL) in a given test.



B5

Type II Error

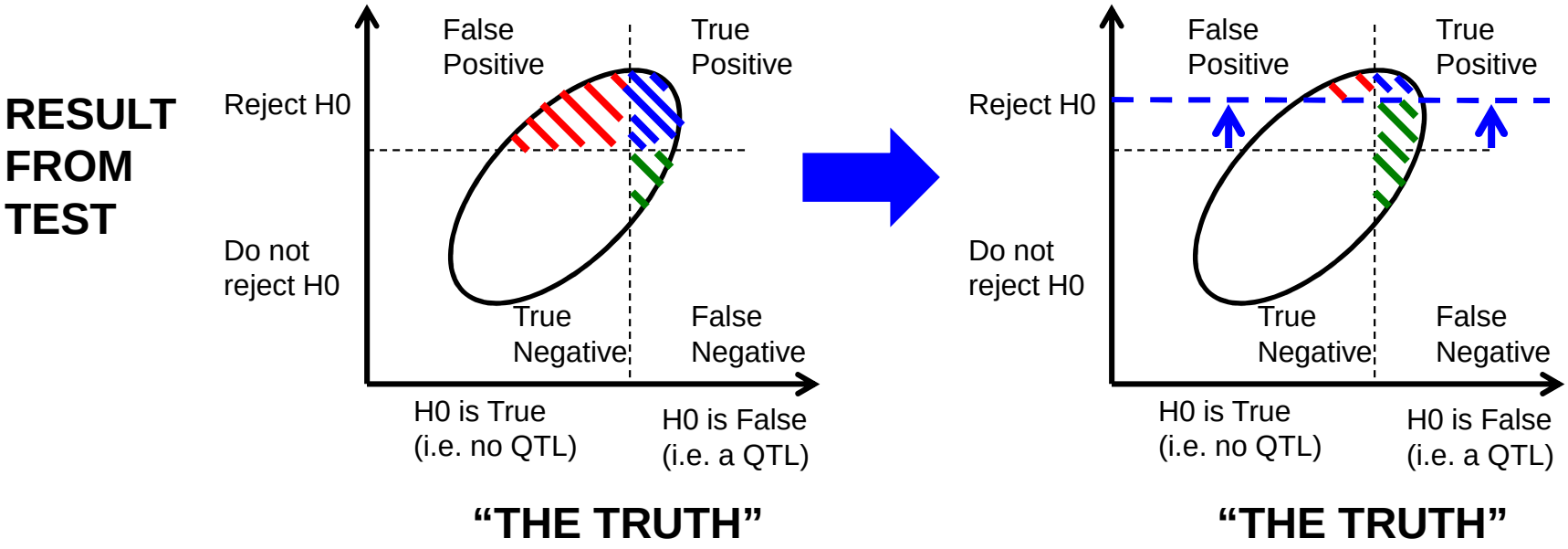
Type II error: It is the failure to reject a true null hypothesis (when there is a QTL and we fail to declare it) in a given test.
Probability of Type II error (β): It is the probability of finding a false negative (failure to declare a true QTL) in a given test.



B5

Type I and II Error

NOTE that for a given test, decreasing the false positive rate means that **the power** (the proportion of true positives, or 1- **β**) is also decreased. The only way to decrease false positives and increase the power is by changing your design (i.e. increasing the population sizes, reducing experimental error, etc.).



Multiple Testing

Let's now assume that we are not interested in a single hypothesis testing, but we are conducting multiple hypothesis testing. For example, we are performing one hypothesis testing on each marker (at each marker we ask the question about whether the marker is associated to a QTL or not). We could summarize the information of the number of hypothesis that follow each category in the following chart:

	Decision		
"Truth"	Do not reject H0	Reject H0	Total
H0 is true	U	V	m_0
H0 is false	Z	S	$m_1=m-m_0$
Total	m-R	R	m

Benjamini and Hochberg, 1995

Multiple Testing

	Decision		
"Truth"	Do not reject H0	Reject H0	Total
H0 is true	U	V	m_0
H0 is false	Z	S	$m_1=m-m_0$
Total	m-R	R	m

Some more definitions:

Per-family error rate (PFER): expected number of false positives: E(V).
Per-comparison error rate (PCER): proportion of false positives in the total: E(V)/m.
Family-wise error rate (FWER): probability of at least one false positive: P(V≥1).
False Discovery Rate (FDR): proportion of false positives among rejected: E(V/R).

We could be interested in controlling the probability of having at least one false positive test (using family control). Or we could be interested in correcting the results by the proportion of false discoveries among the reject tests (using FDR).

Benjamini and Hochberg, 1995

Multiple Testing – Family control

If we aim to control FWER, it is necessary to use a somewhat smaller alpha value. But how small should it be? There are many methods that aim to control FWER. We will discuss two common methods that have been used in QTL mapping.

Bonferroni:

Since the increase in the error is related to the number of independent tests, Bonferroni proposed to use a new alpha value as threshold:

$$\alpha^* = 1 - (1 - \alpha)^{1/m}$$

$$\cong \frac{\alpha}{m}$$

However, many studies have shown that a Bonferroni correction for QTL studies is overly conservative mainly because tests are not independent (i.e. markers are linked and therefore not-independent). Having an unnecessarily stringent threshold reduces power to detect true QTL as has been shown.

Benjamini and Hochberg, 1995; Li and Ji, 2005

Multiple Testing – Family control

Li and Ji (2005):

An alternative is to use a Bonferroni correction but instead of using the total number of tests (which we know are not independent because markers are correlated), is to use the effective number of independent tests. This idea was first proposed by Cheverud (2001) and then modified by Li and Ji (2005). The steps are:

1. Calculate the correlation matrix of markers.
2. Use the number of significantly different from zero eigenvalues (λ_i) of the correlation matrix to determine the effective number of independent tests (M_{eff}) as:

$$M_{eff} = \sum_{i=1}^{M} f\left(|\lambda_i|\right)$$

$$f(x) = I(x \geq 1) + (x - \lfloor x \rfloor), x \geq 0$$

3. Use a Bonferroni-type of correction to determine the threshold but using the effective number of independent tests instead of the total number of tests:

$$\alpha^* = 1 - (1 - \alpha)^{1/Meff} \cong \frac{\alpha}{Meff}$$

Li and Ji, 2005

Multiple Testing – Family control

Li and Ji (2005):

This method have been shown to be better than the Bonferroni correction and the Cheverud (2001) method.

It performs equally good as permutation but is fast and simple to perform.

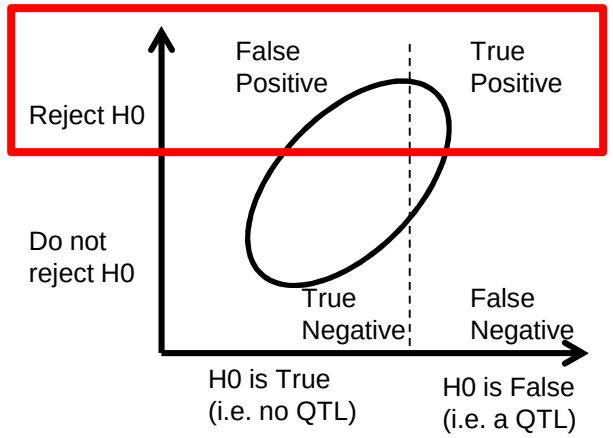
Li and Ji, 2005

Multiple Testing – False Discovery

	Decision		
"Truth"	Do not reject H0	Reject H0	Total
H0 is true	U	V	m_0
H0 is false	Z	S	$m_1=m-m_0$
Total	m-R	R	m

The False Discovery Rate (FDR) is the proportion of false positives amongst the rejected hypothesis: E(V/R).

We are looking at the problem from a different perspective: out of the total tests that we reject, how many are true positives?



Benjamini and Hochberg, 1995

Multiple Testing – False Discovery

False Discovery Rate (FDR): The steps to use the False Discovery Rate are as follow:

1. Order the observed p-values: p(1) ≤ ... ≤ p(m)
2. Calculate an arithmetic sequence as follows: $\frac{i}{m}\alpha$
3. Reject all hypothesis where: $k = \max_{1 \le i \le m} \left\{ i : p_{(i)} \le \frac{i}{m}\alpha \right\}$

FDR is also conservative in most cases. Effective number of markers and a FDR using M_{eff} instead of *m* could be used (Li and Ji, 2005).

Benjamini and Hochberg, 1995

Multiple Testing - Permutation

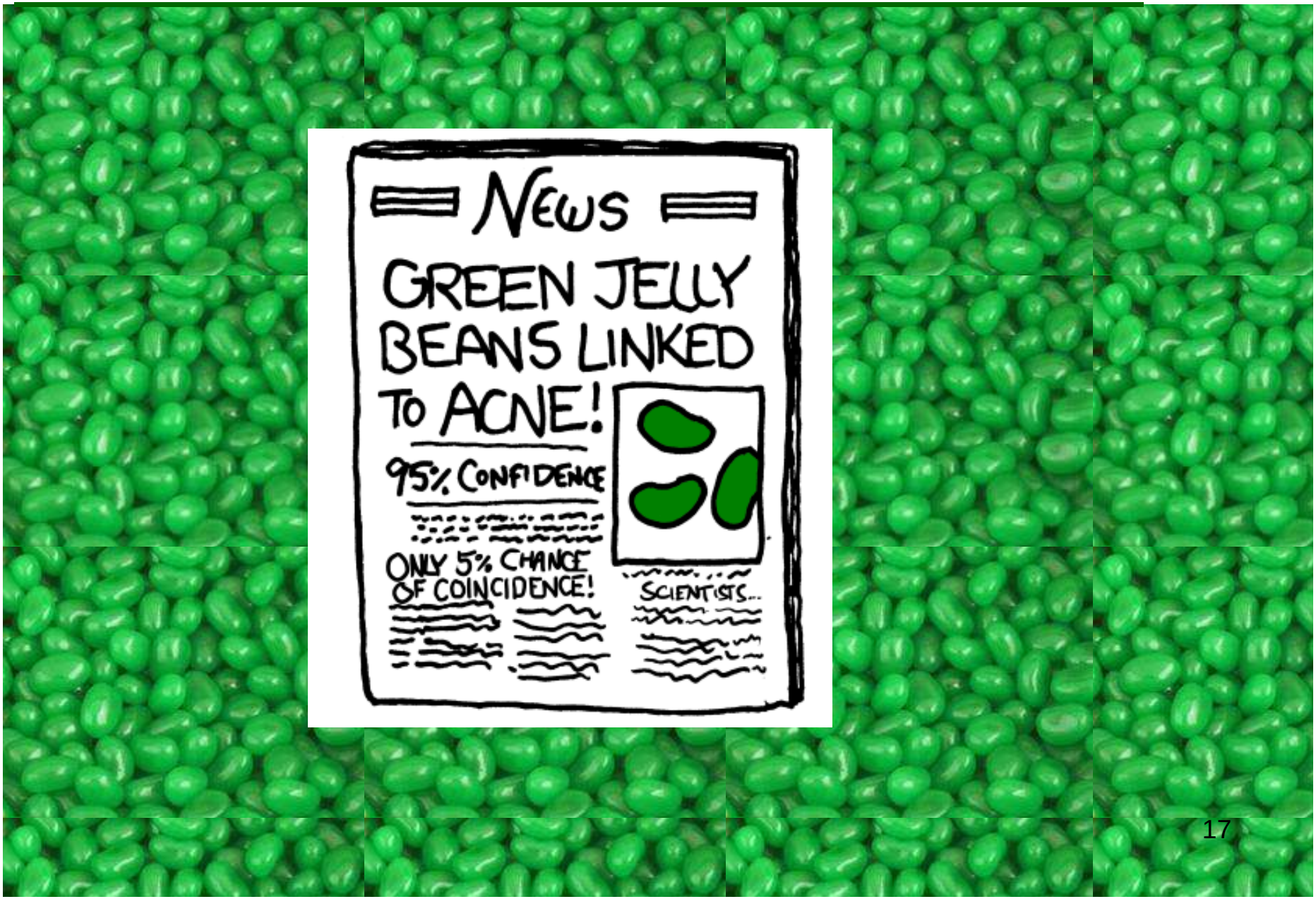
Permutations (Broman and Sen, 2009).

The steps involved are:

1. Randomize (shuffle) the phenotypes relative to the marker data.
2. Perform a QTL mapping and obtain p-values (or LOD scores) and keep the most extreme value (i.e. Maximum LOD or smallest p-value): M_i^*
3. Repeat 1 and 2 several (r) times (i.e. 1,000 or 10,000).
4. Produce the empirical distribution of the extreme values: M_1^*,... M_r^*.
5. Use the 95^{th} percentile of M_i^* values as the threshold.

This are computing-intensive but are precise for all cases.

Multiple Testing



Multiple Testing

SO WHAT WAS WRONG WITH THE PREVIOUS CARTOON?

A common practice is to use P<0.05 to decide about significance of a test. But with large number of tests the chance of having at least one false positive is close to 1.

LET'S LOOK AT OUR OPTIONS IN HYPOTHESIS TESTING

1. **Family Control**: Bonferroni multiple-testing protection
    1. P = 0.05 / number of tests (this is very conservative)
    2. P = 0.05 / effective number of tests, with the effective number of tests estimated from the marker data (Li and Ji, 2005).

2. **False Discovery Rate** (Benjamini and Hochberg, 1995).

3. **Permutations** (Broman and Sen, 2009).