



# Quantitative Biology Lecture 5 (Hypothesis Testing)

Gurinder Singh "Mickey" Atwal  
Center for Quantitative Biology

# Summary

- Classification Errors
- Statistical significance
- T-tests
- Q-values

# (Traditional) Hypothesis Testing

- Formulate the null hypothesis
- Choose an appropriate statistic
- Decide on a significance level (usually 0.05)
- Compute a p-value for your test and determine if significant
- “Reject” or “Fail to Reject” null hypothesis
- Note ***caveats*** with p-value interpretations.

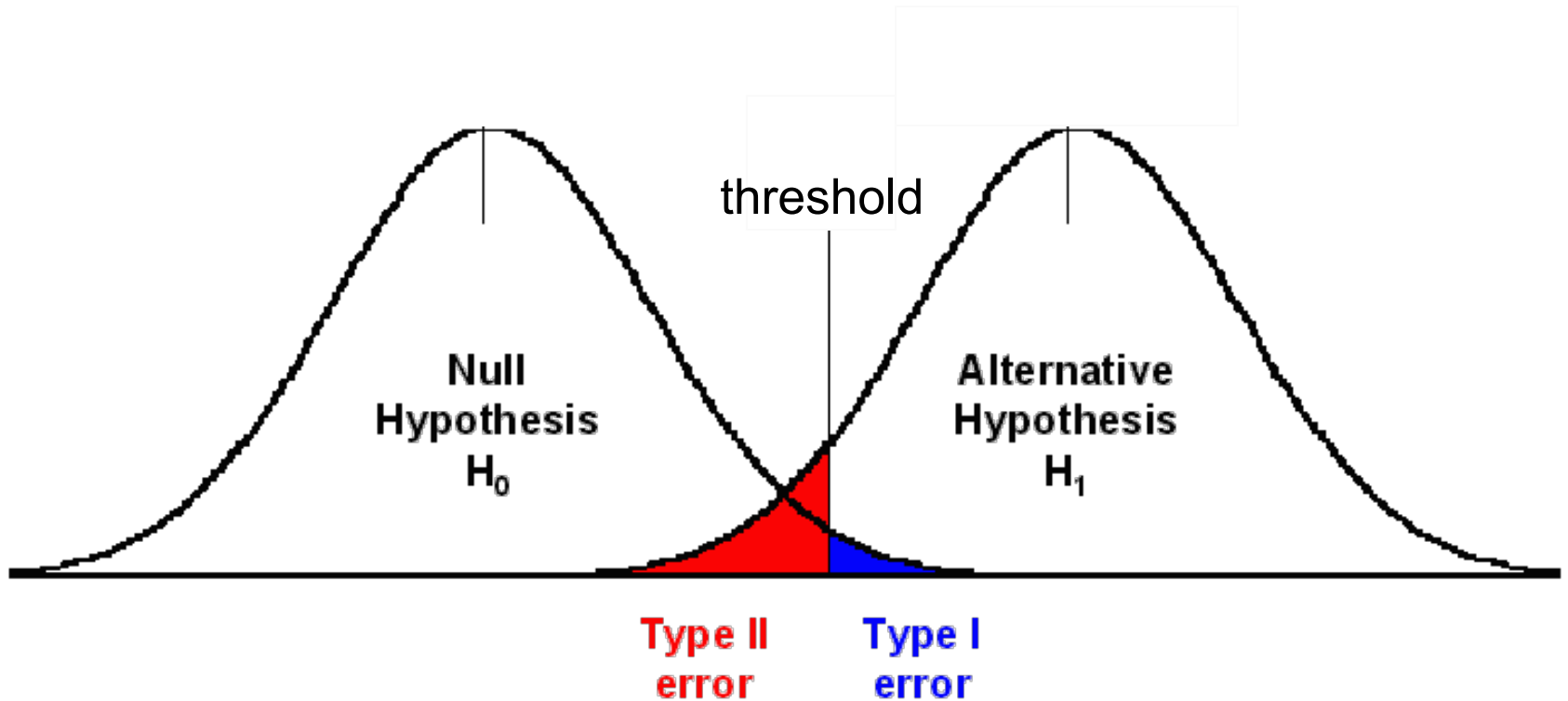
# Binary Classifier

- Take a particular DNA sequence. For example,  
**AGTACAGTATAGCAGATGGAAAATTAC**
- Is it more likely to come from either a Homo sapiens or Plasmodium falciparum sample? (HYPOTHESIS TESTING)
- What is the probability that you will be right/wrong? (STATISTICAL ERRORS)
- How long should the DNA sequence be before I am reasonably accurate? (POWER CALCULATIONS)

# Errors in Hypothesis Testing

		Actual Situation	
		$H_0$ True	$H_0$ False
Test Results	$H_0$ Accepted	<b>Correct</b>	<b>Type II error (<math>\beta</math>) False Negative</b>
	$H_0$ Rejected	<b>Type I error (<math>\alpha</math>) False Positive</b>	<b>Correct</b>

# Errors in Hypothesis Testing



POWER=1 - Type II error

# Sensitivity and Specificity

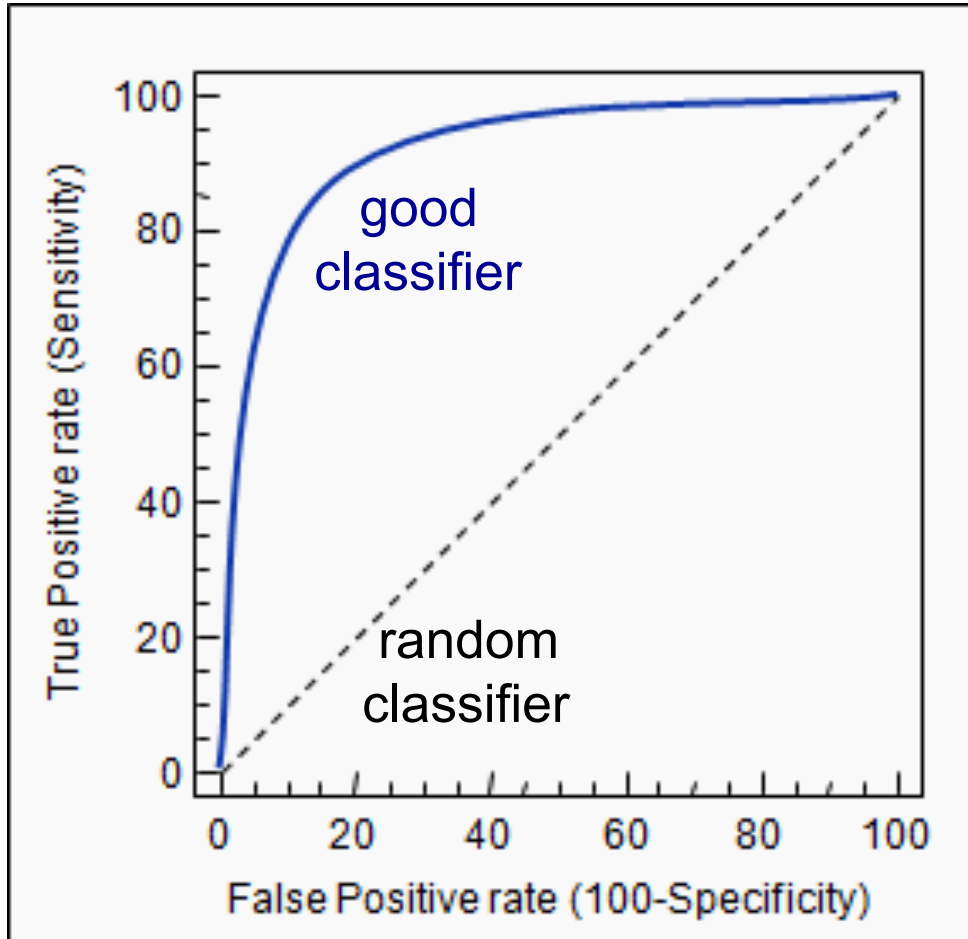
## True positive rate

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

## True negative rate

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

# ROC curve summarizes the sensitivity and specificity of test



Area under curve (AUC)  
often used to judge the  
quality of the classifier



# Statistical Power

- = True positive rate (sensitivity)
- Calculating power aids in design of experiments
- Tells you how many samples you need to detect an effect.

# T-tests

- Used to calculate p-values and reject/accept null hypotheses.
- Null hypotheses commonly assumed to be a Gaussian (normal) distribution.

# One-sample t-test

- Tests if the mean of a population,  $\mu$ , is equal to a particular value  $m$ .
- Null hypothesis:  $H_0: \mu=m$
- Alternate hypothesis:  $H_A: \mu \neq m$
- Test statistic

$$t = \frac{\langle x \rangle - m}{s/\sqrt{n}} \quad (\text{degrees of freedom} = n-1)$$

- $\langle x \rangle$ =sample mean (i.e. estimate of population mean);  
 $s$ =standard deviations;  $n$ =number of samples
- Data assumed to be normally distributed

# Two sample t-test

- Tests whether the population means,  $\mu_1$  and  $\mu_2$ , are equal or not.
- $H_0: \mu_1 = \mu_2$  vs  $H_A: \mu_1 \neq \mu_2$
- Test statistic:
$$t = \frac{\langle x_1 \rangle - \langle x_2 \rangle}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
- Assumes two independent populations
- Does not assume equal variances

# Two sample t-test (equal variance)

- Tests whether the population means,  $\mu_1$  and  $\mu_2$ , are equal or not.
- $H_0: \mu_1 = \mu_2$  vs  $H_A: \mu_1 \neq \mu_2$

- Test statistic:

$$t = \frac{\langle x_1 \rangle - \langle x_2 \rangle}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{df} = n_1 + n_2 - 2$$

- Assumes two independent populations
- Assume equal variances

# Paired t-test

- Tests if means in paired observations are different
- $d = x_1 - x_2$  difference in paired observation
- $H_0: d = 0$     $H_A: d \neq 0$
- Test statistic

$$t = \frac{\langle d \rangle}{s_d / \sqrt{n}} \quad (\text{df} = n - 1)$$

# Independence Tests

- If sample sizes are small, use Fisher Exact (remember the hypergeometric function)
- If sample sizes are large, with more than 5 samples in each cell, use Mutual Information or  $\chi^2$  test

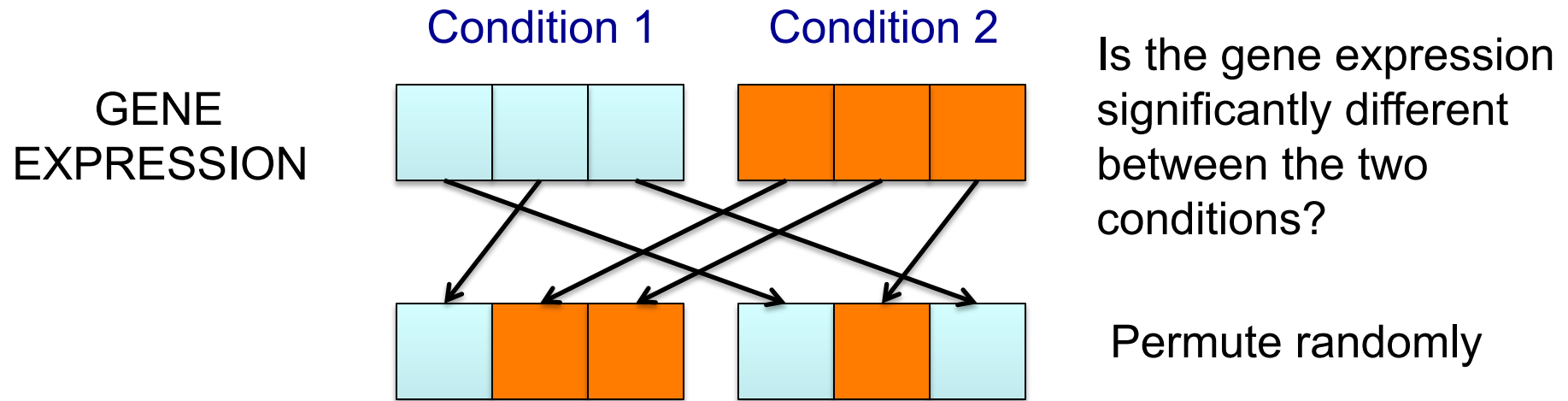
$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \quad \text{df}=(r-1)(c-1)$$

# Permutation Tests

- Sometimes the distribution of the test statistic under the null hypothesis is unknown
- We can generate the null distribution by permuting the data
- If data too large, we can approximate the null distribution by drawing random samples (Monte Carlo method).
- See Question 6 from homework assignment <http://nbviewer.ipython.org/url/atwallab.cshl.edu/teaching/QBHW1.ipynb>



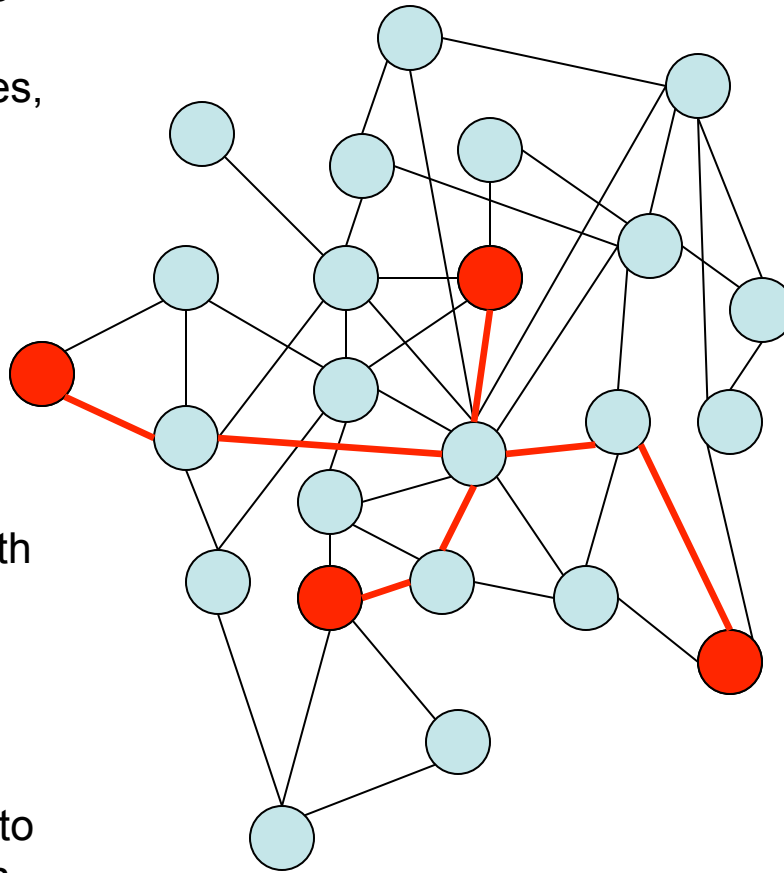
# Monte Carlo Permutation Test



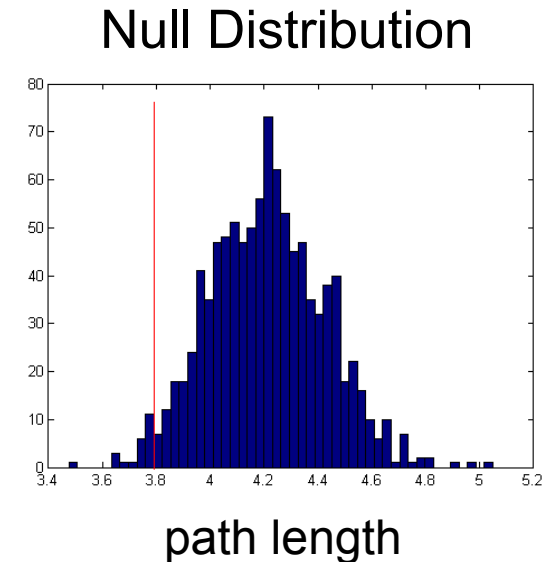
1. Permute gene expression values
2. Calculate relevant statistic quantifying difference between conditions (e.g. t-test)
3. Repeat many times
4. Null distribution = distribution of permuted statistic values

# Is the path length between two nodes significantly short ?

- ❖ E.g. Human Protein Reference Database (HPRD) contains 15,231 protein entries, and 34,624 directly protein-protein interactions.



- ❖ Find the shortest path length between any pair of random proteins in HPRD.
- ❖ Repeat many times to get a null distribution of path lengths



# Multiple Hypothesis Testing

- Testing many hypotheses can give a small p-value by chance
- Need to correct significance threshold for multiple hypothesis testing

$$x' = 1 - (1 - x)^N$$

(Bonferroni Correction)

$$\approx Nx$$

- Bonferroni Correction: multiply observed p-value with total number of hypotheses

# False Discovery Rate

- Fraction of False Positives out of all results we call significant
- Determined by looking at all the p-values generated by the multiple tests

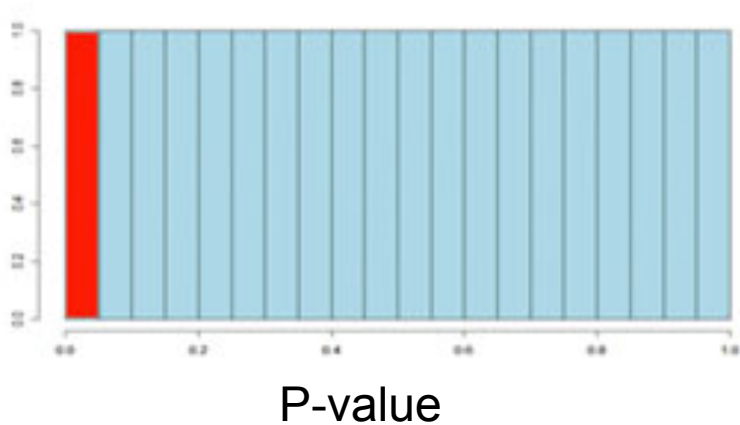
# Q-value

- p-value of 0.05 implies that 5% of all tests will result in false positives
- FDR adjusted p-value (or q-value) of 0.05 implies that 5% of significant tests will result in false positives.

# Q-value

Histogram of p-values can reveal what fraction of the multiple tests are not consistent with the null hypothesis

All tests from from null hypothesis



Fraction of tests from null hypothesis

