

# GCB/CIS 535 Microarray Topics

John Tobias  
November 8th, 2004



## Overview

- Normalization
- Sample Quality Control
- Finding Differentially Expressed Genes
- Gene List Annotation

## Overview

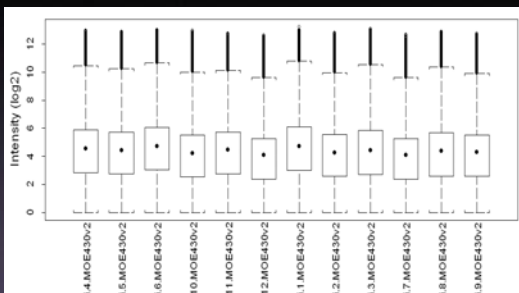
- Normalization**
- Sample Quality Control
- Finding Differentially Expressed Genes
- Gene List Annotation

## Normalization

Adjustment of gene expression values across an experiment

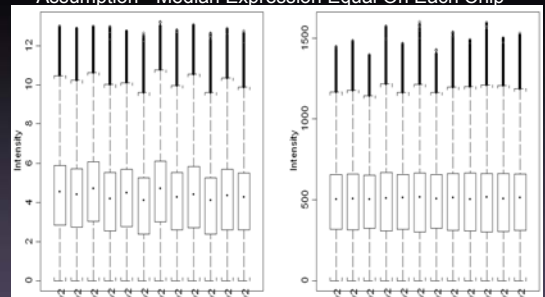
- Necessary step - corrects systematic biases, facilitates biological comparison
- Assumptions about sample (and condition) similarities can help focus on true differences

## The Box Plot



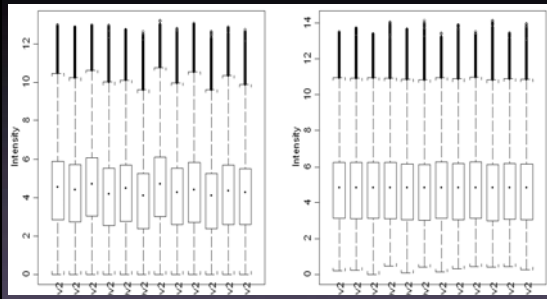
## Linear Scaling

Assumption - Median Expression Equal On Each Chip



# MedianIQR

Assumption - Specific Percentiles Equal On Each Chip



# Other Normalizations

"Housekeeping" Genes - Assume that expression of some specific genes is unchanging  
LOESS (LOWESS) - 2-color - Intensity Dependent

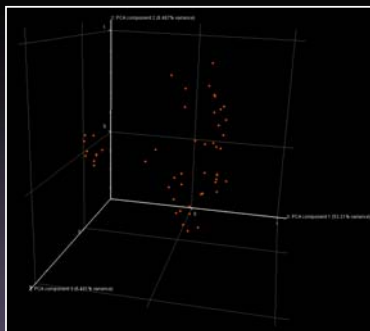
# Overview

Normalization  
Sample Quality Control  
Finding Differentially Expressed Genes  
Gene List Annotation

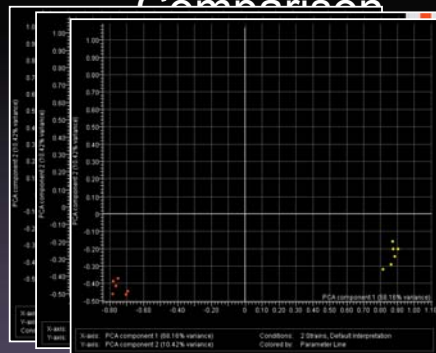
# Sample Quality Control

Affymetrix "report" values  
% Present, 3'/5' ratios, scaling factor  
Clustering or PCA for sample similarity  
Looking for absence of technical grouping – biological condition grouping may be present or absent depending on magnitude of expression changes

# PCA for Sample QC



# PCA for Sample Comparison



# Overview

Normalization  
Sample Quality Control  
Finding Differentially Expressed Genes  
Gene List Annotation

# Finding Differentially Expressed Genes

Choice of algorithm determined by experimental design and goals  
Multiple testing problem (why p-values can be misleading)

# Gene List Characteristics

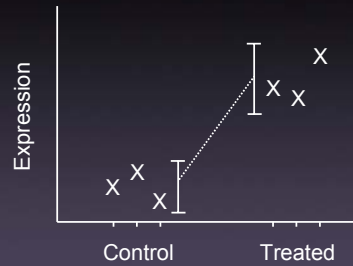
Ranked by statistical significance for differential expression

Gene 384 (high significance)  
Gene 983  
Gene 324  
Gene 002  
Gene 334  
Gene 623  
Gene 356  
Gene 342  
Gene 078  
Gene 645  
Gene 453  
Gene 235  
Gene 193  
Gene 903  
Gene 411  
Gene 212 (low significance)

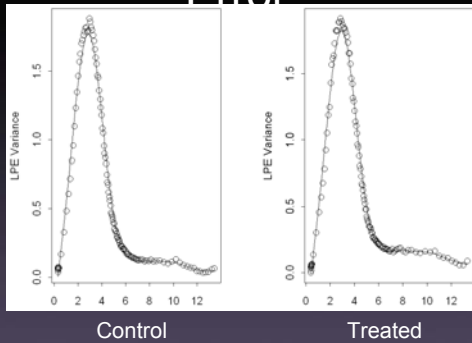
True Positives  
False Positives

True Negatives  
False Negatives

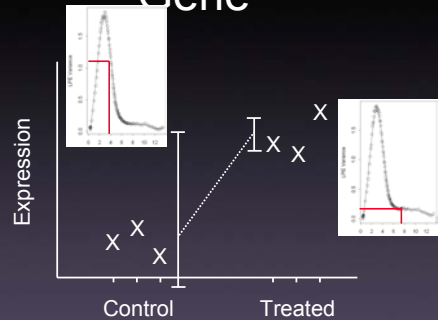
# Data for a Single Gene



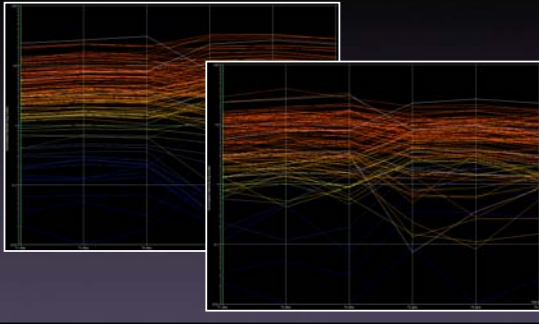
# Pool Genes to Estimate Error



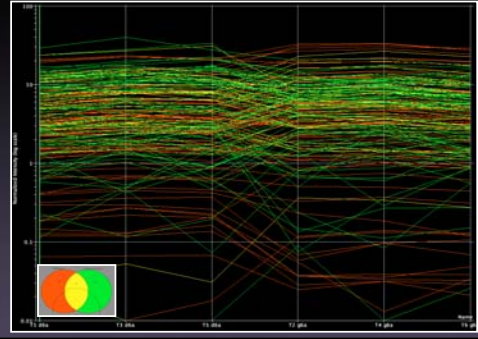
# Data for a Single Gene



## Comparing Gene Lists



## Comparing Gene Lists



## Differential Expression Algorithms

Affymetrix Microarray Suite v5  
 Significance Analysis of Microarrays (SAM)  
<http://www-stat.stanford.edu/~tibs/SAM/>  
 Local Pooled Error (LPE)  
<http://bioinformatics.oupjournals.org/cgi/reprint/19/15/1945>  
 NIA Website (choice of variance models)  
<http://lgsun.grc.nia.nih.gov/ANOVA/>  
 Mixed Model Multi-way ANOVA

## Overview

Normalization  
 Sample Quality Control  
 Finding Differentially Expressed Genes  
 Gene List Annotation

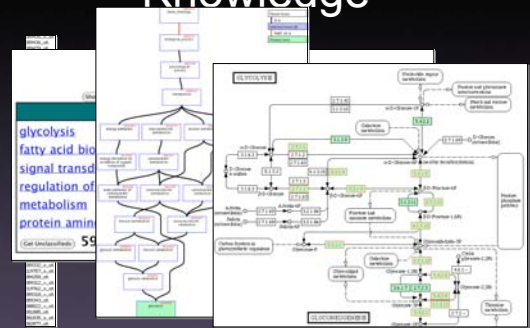
## Gene List Annotation Pathways Functional Groups

Affymetrix  
 GeneSpring  
 DAVID  
 GoMiner  
 GenMapp



<http://apps1.niaid.nih.gov/david/>

## Identifiers to Knowledge



# Ingenuity Pathway Analysis



<http://www.ingenuity.com>  
Curated Interaction and Pathway  
Database  
Mine literature as it relates to  
gene list

# Contact Information

Penn Bioinformatics Core - 13th Floor Blockley Hall  
John Tobias - 1314 - [jtobias@pcbi.upenn.edu](mailto:jtobias@pcbi.upenn.edu)  
Reserve Computers  
<http://core.pcbi.upenn.edu/>

