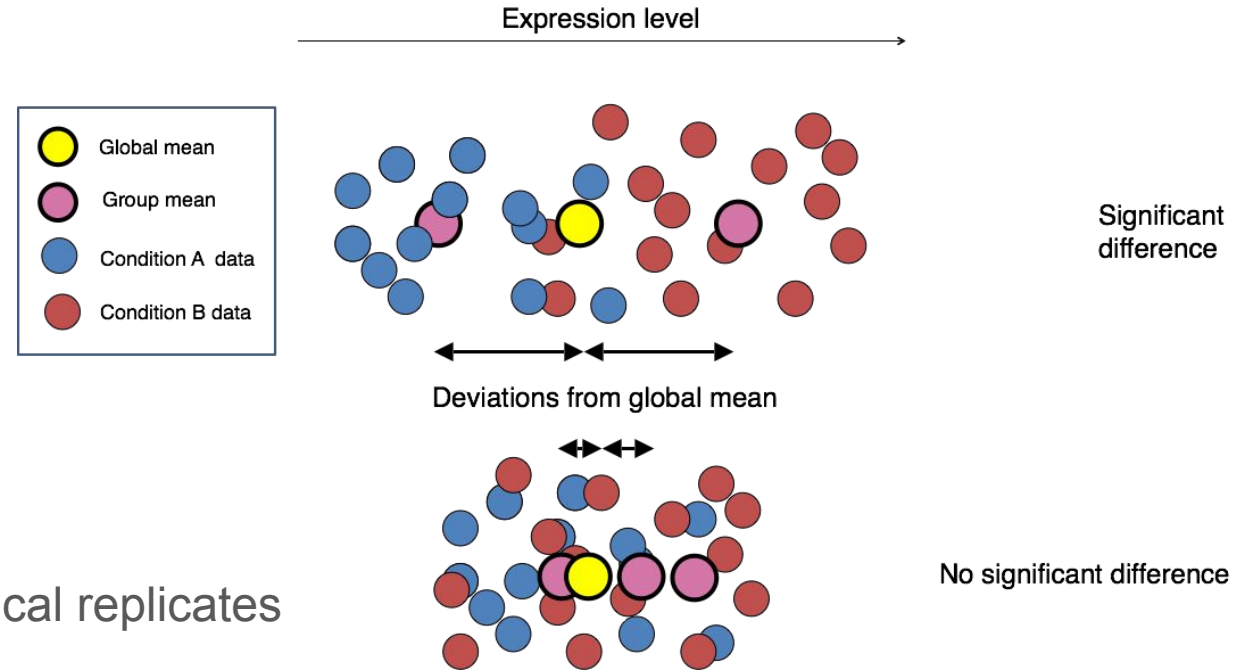
# Introduction to RNA-seq

Differential gene expression and pathway enrichment

# What is the purpose?

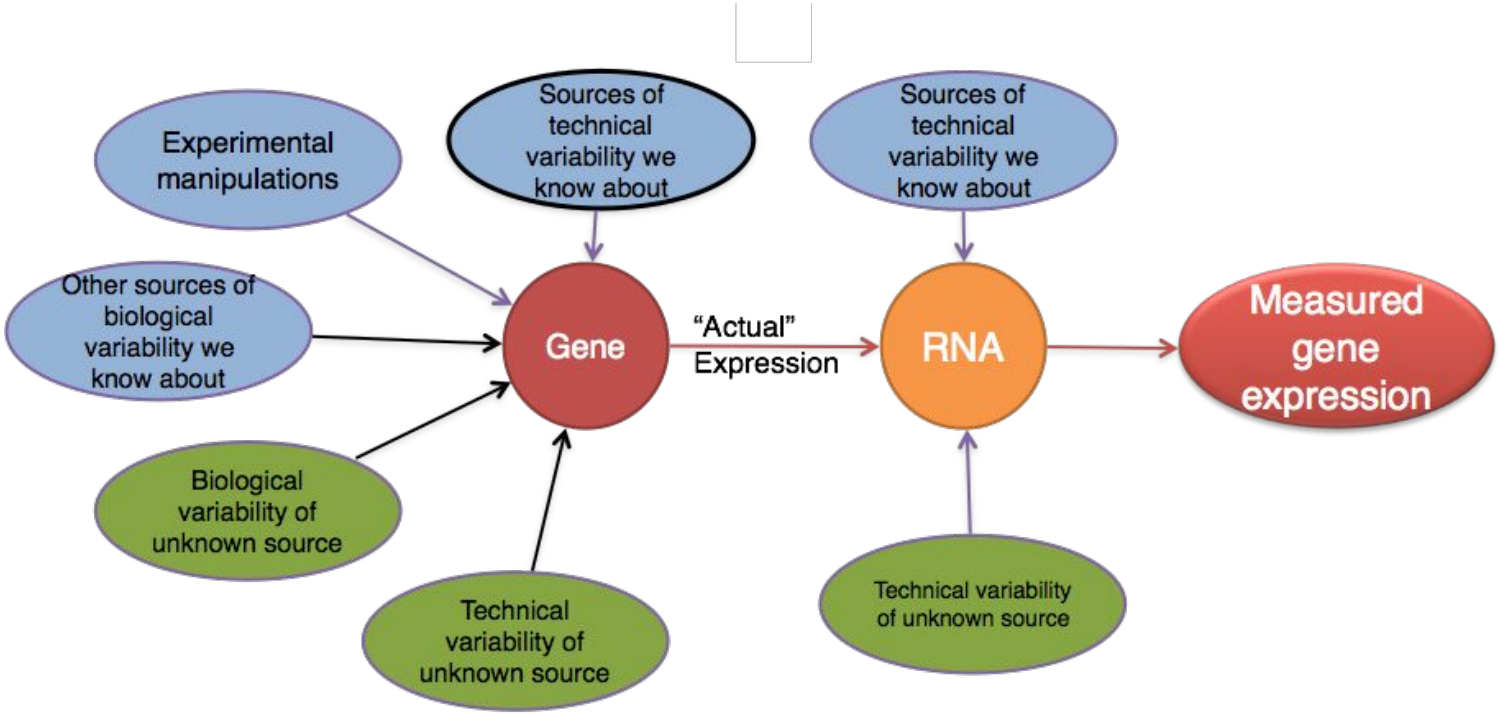
We want to identify why expression levels differ across conditions




Challenges

- Few replicates
- Discrete count data
- Variability from biological replicates

# Sources of variability




Courtesy of Paul Pavlidis, UBC

# Normalization

We want to account for factors that prevent direct comparison of raw count data

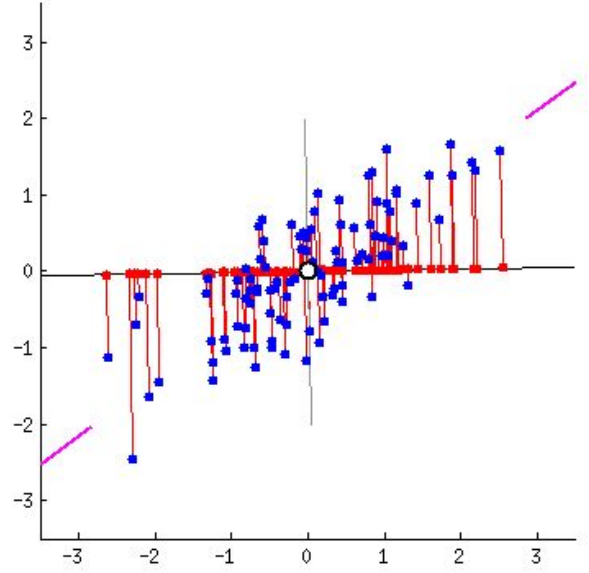
1. Within samples: gene transcript length and GC content
2. Between samples: batch effects and sequencing depth

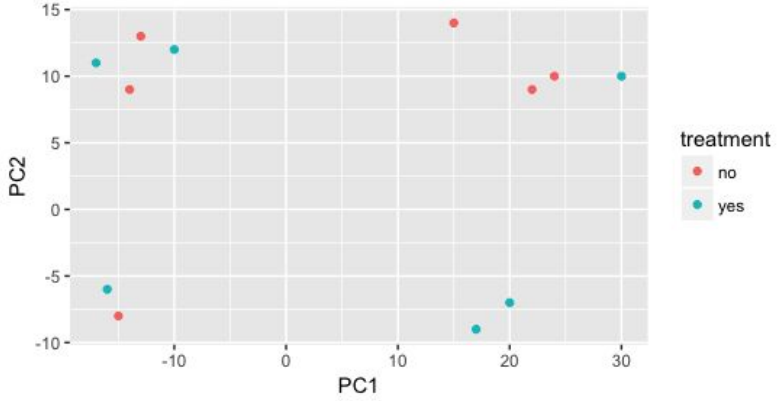
Approach

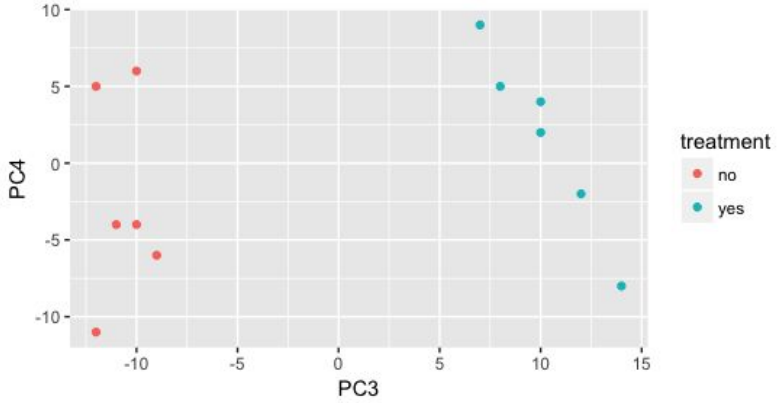
1. RPKM/FPKM: between genes within a sample
2. TPM: within a sample or between samples
3. Median of ratios: between samples (DESeq2)

# Principal components analysis

- DImensionality reduction
- Finds principal components that maximize variance

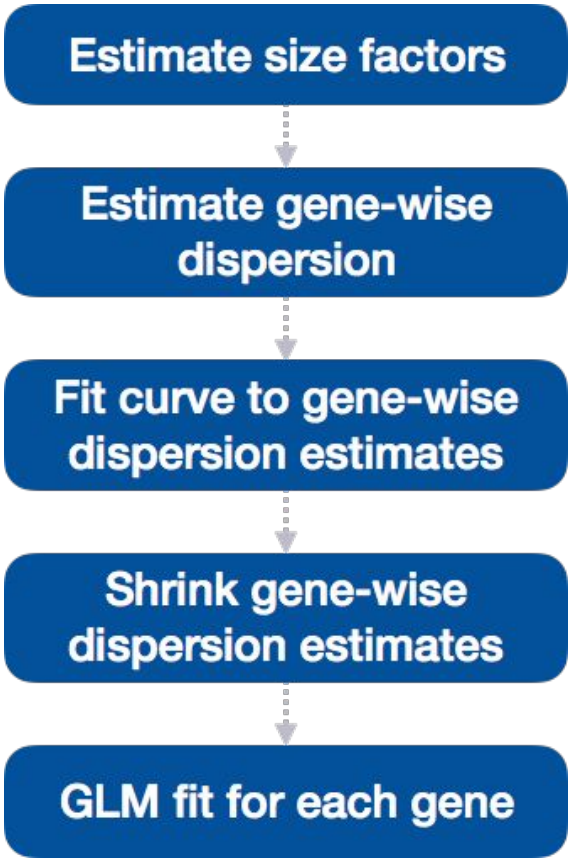


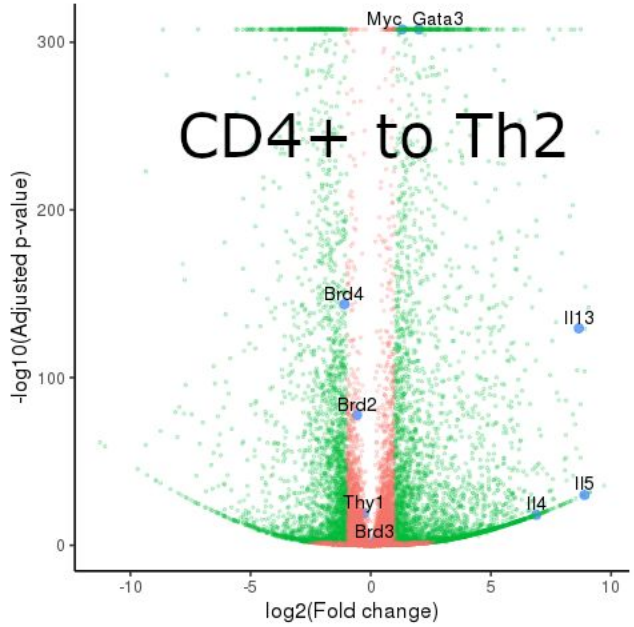



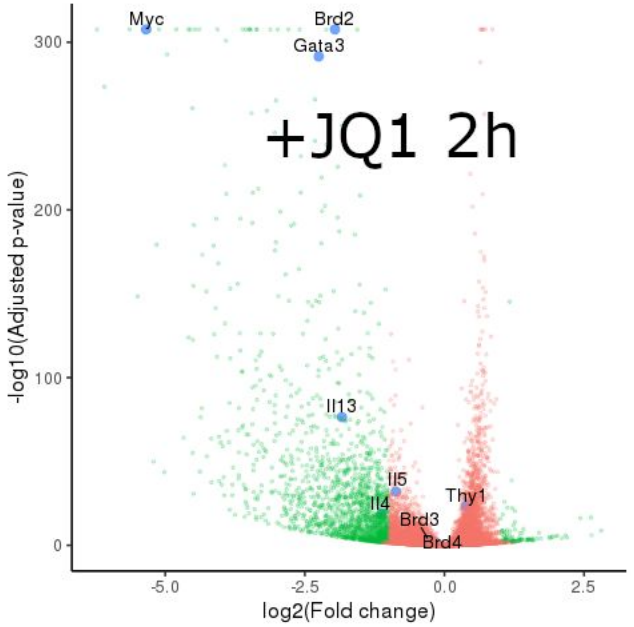
# DESeq2

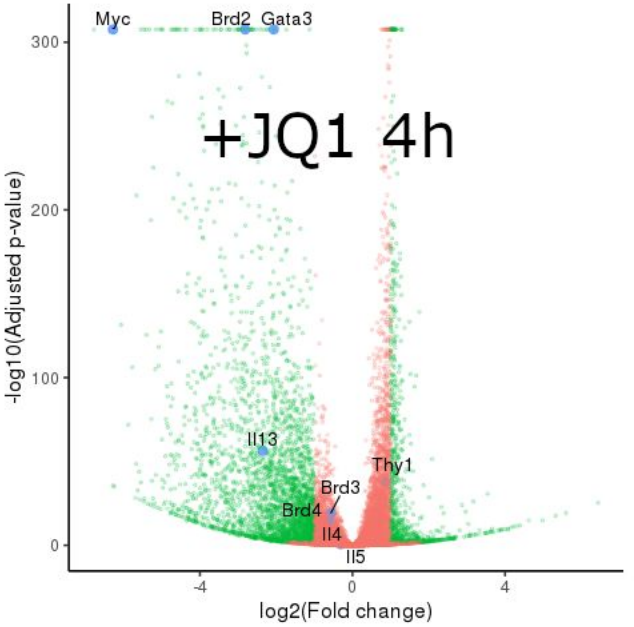
- Estimating the variability of genes given their sequencing depth and behaviour relative to other genes
- Fit a negative binomial distribution to the count data
- Wald test evaluates the impact of explanatory variables
    - Approximation of the likelihood ratio test



# Volcano plots



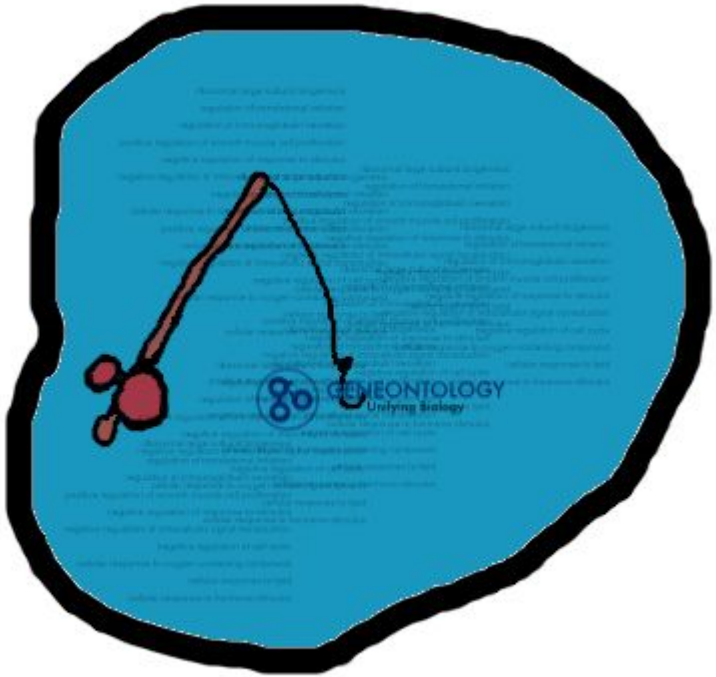
# Multiple test correction

- When we are searching for significance across multiple hypotheses, we have to adjust our p-value threshold downward
- Recall that the p-value is the probability of obtaining results at least as extreme as the results observed
- Important for both DGE and pathway enrichment

1. Bonferroni correction: $\alpha/n$
2. Benjamini-Hochberg: $P_{(k)} \leq \alpha \cdot k / n$

# Fishing for enriched pathways

- Given a set of genes, we want to find what pathways are associated with them
- Every pathway is a hypothesis and gene matches are evidence that the hypothesis is true



# Count matrix paper for R tutorial

BIOINFORMATICS ARTICLE

## SNP-adjacent super enhancer network mediates enhanced osteogenic differentiation of MSCs in ankylosing spondylitis

Wenhui Yu[1,‡], Keng Chen[1,‡], Guiwen Ye[3,‡], Shan Wang[2], Peng Wang[1], Jinteng Li[1], Guan Zheng[1], Wenjie Liu[1], Jiajie Lin[1], Zepeng Su[1], Yunshu Che[1], Feng Ye[3], Mengjun Ma[1], Zhongyu Xie[1,*] and Huiyong Shen[1,*,†]

[1]Department of Orthopedics, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen 518003, P.R. China, [2]Center for Biotherapy, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen 518003, P.R. China and [3]Department of Orthopedics, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, P.R. China

# Other resources

The Gene Ontology Resource
http://geneontology.org/