

Differential Expression Analysis

Center for Health Data Science



Health Data Science Sandbox¹

Overview

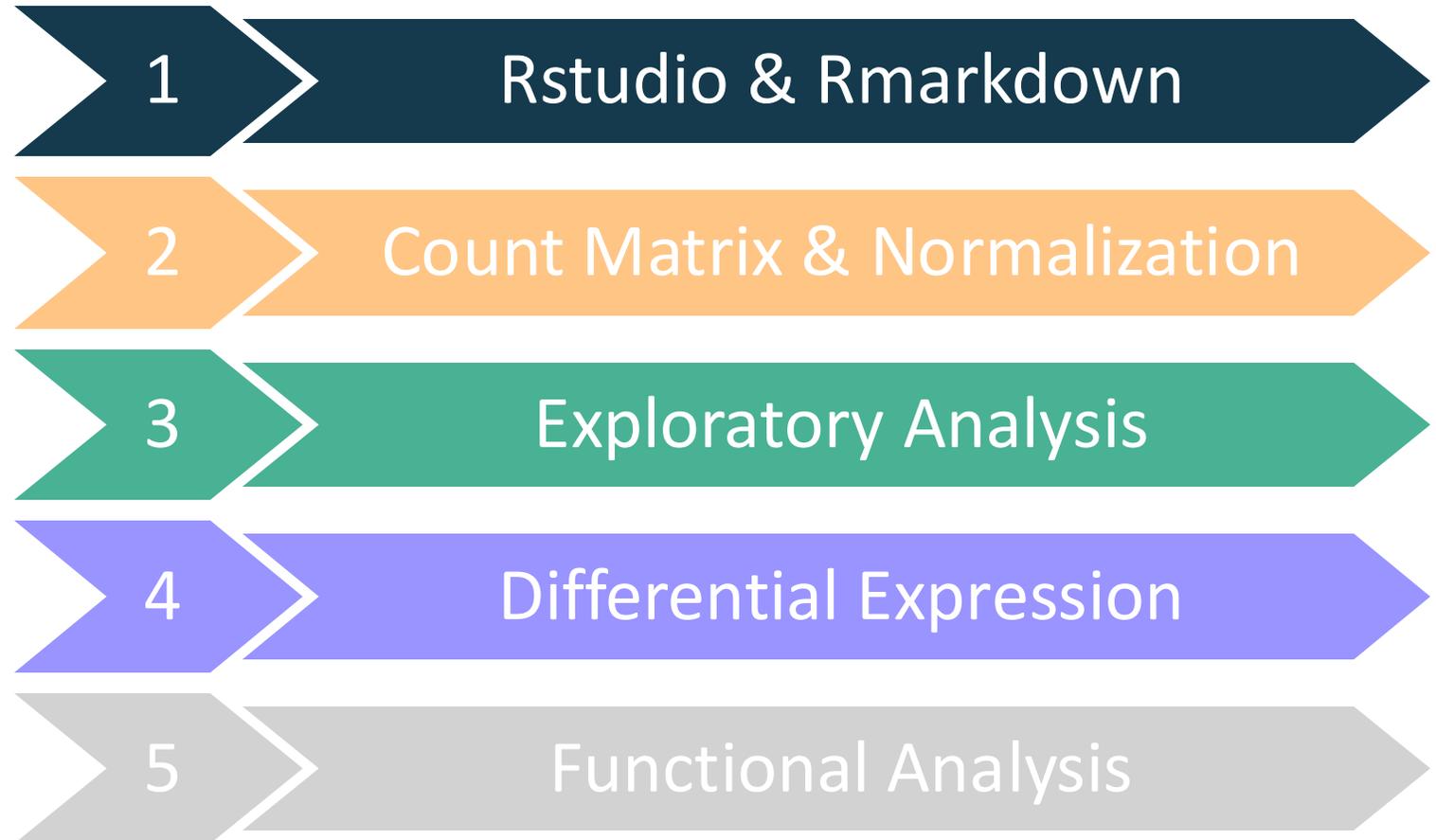


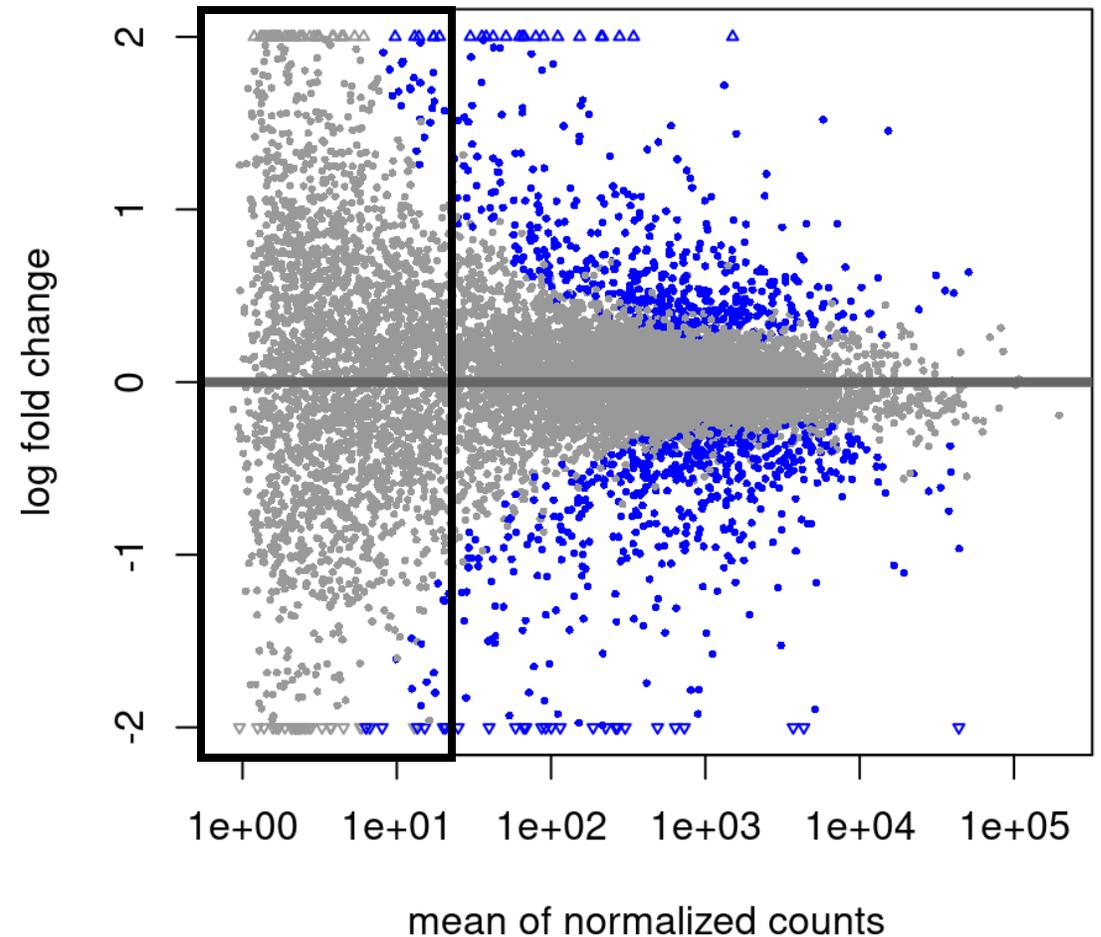
Table discussion – Recap Day 2

The results from your Vampirium intervention study have come back. Excellent! You rub your hands and do an evil laugh, then you consider how you can analyse your data.

- How do you need to prepare your data for analysis?
- What kind of exploratory analysis can you do?
- What do you learn from exploratory analysis?
- Which contrasts would you like to analyse in the following differential expression analysis?

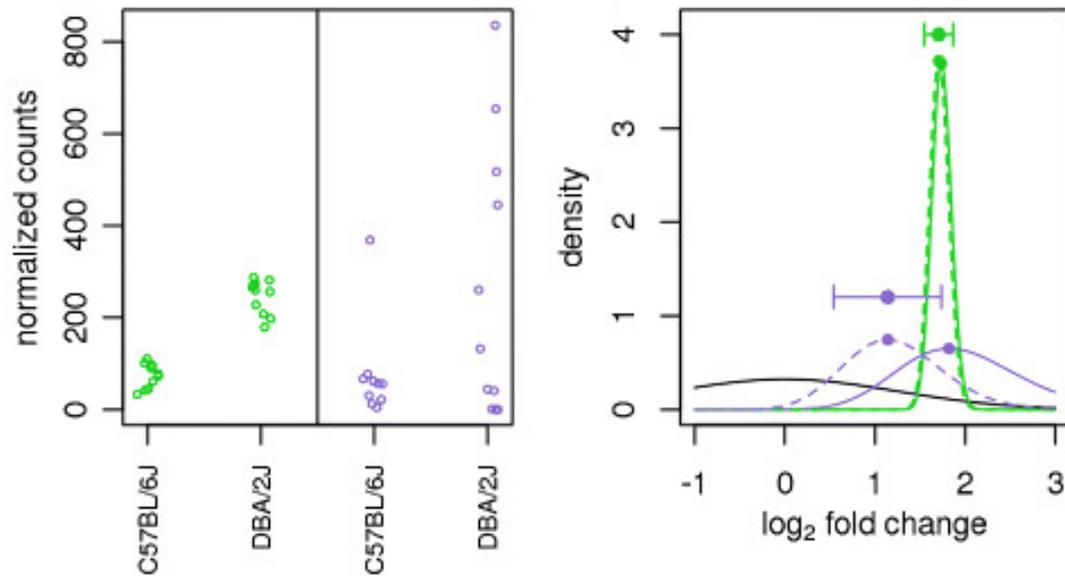
MA plot

- X-axis: mean expression of gene
 - How strongly expressed a gene is
- Y-axis: LFC
 - Difference between condition B/A
- Does not account for dispersion
- LFC from low count genes are very noisy
 - $2/1 \neq 200/100$



LFC and dispersion

The log fold change is calculated by comparing the **mean expression** of gene A in a condition (i.e. over/under-expression) versus a control.

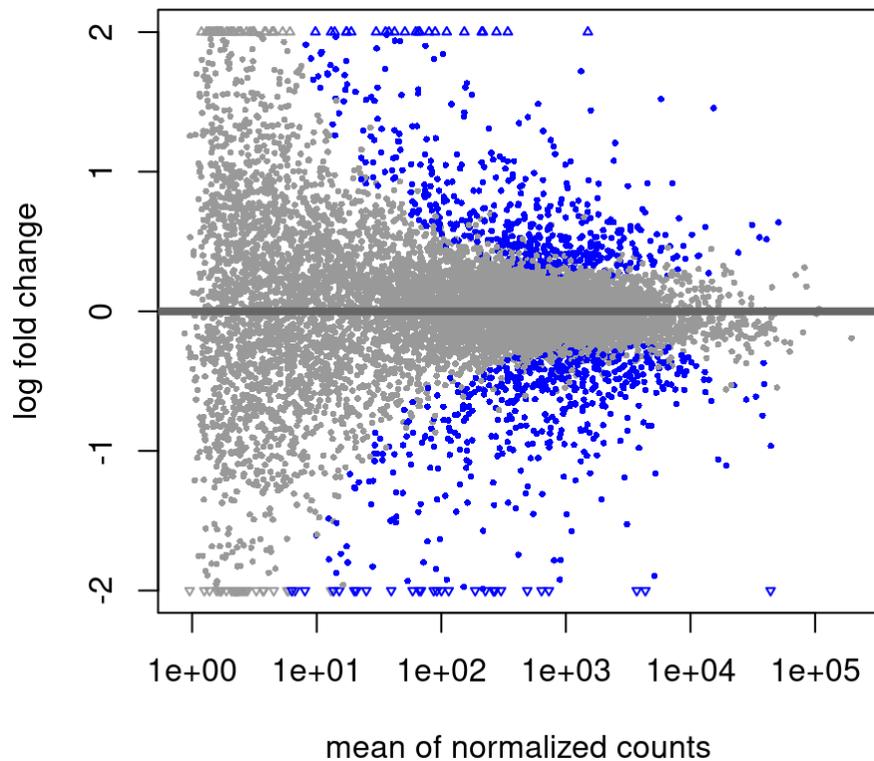


Not all means are created equal!

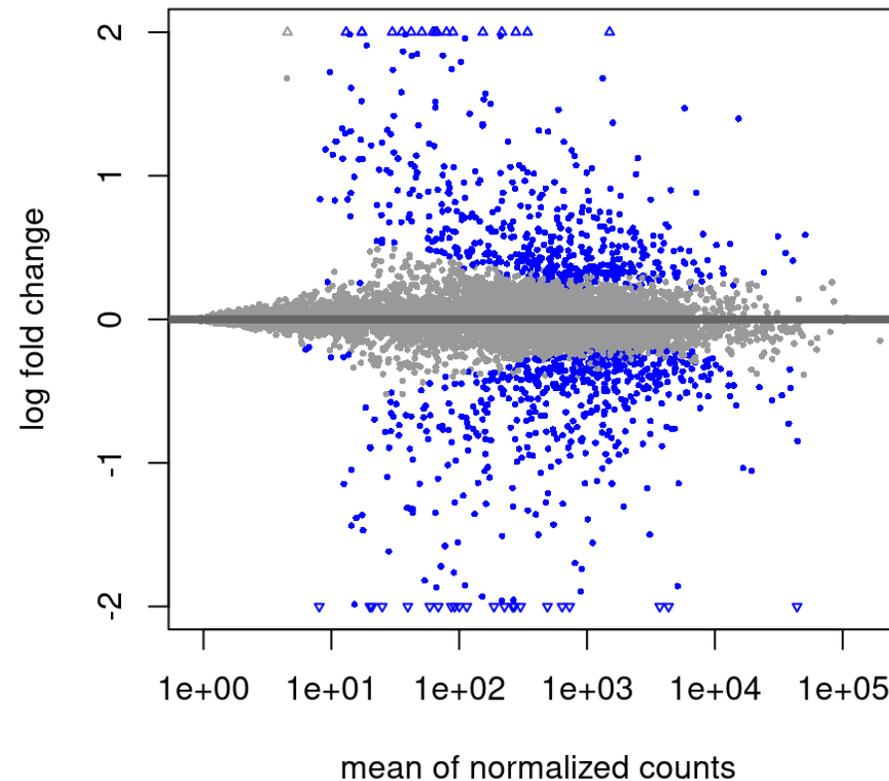
- Within group variance
- Lowly expressed genes might show large FC from a small difference in mean

MA plot: LFC Shrinkage

MA plot before shrinkage



MA plot after shrinkage

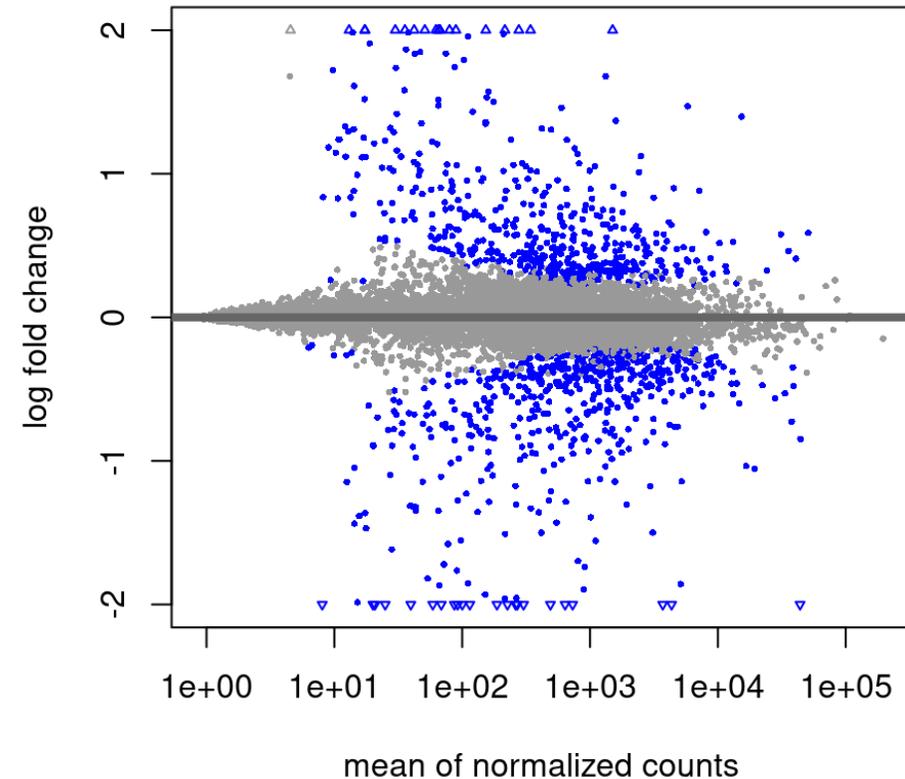


LFC Shrinkage

Use shrunken LFCs for:

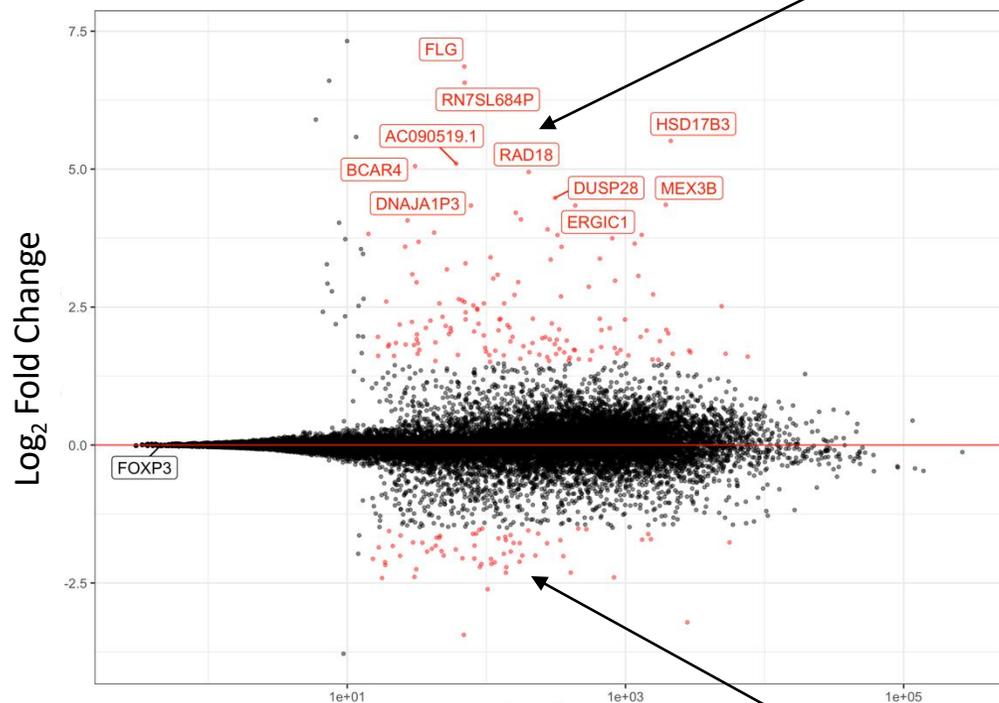
- Gene set enrichment Analysis (GSEA)
- Subsetting significant genes
- Comparing LFCs between conditions in the same experiment
- Reporting in papers (can put both)

MA plot after shrinkage

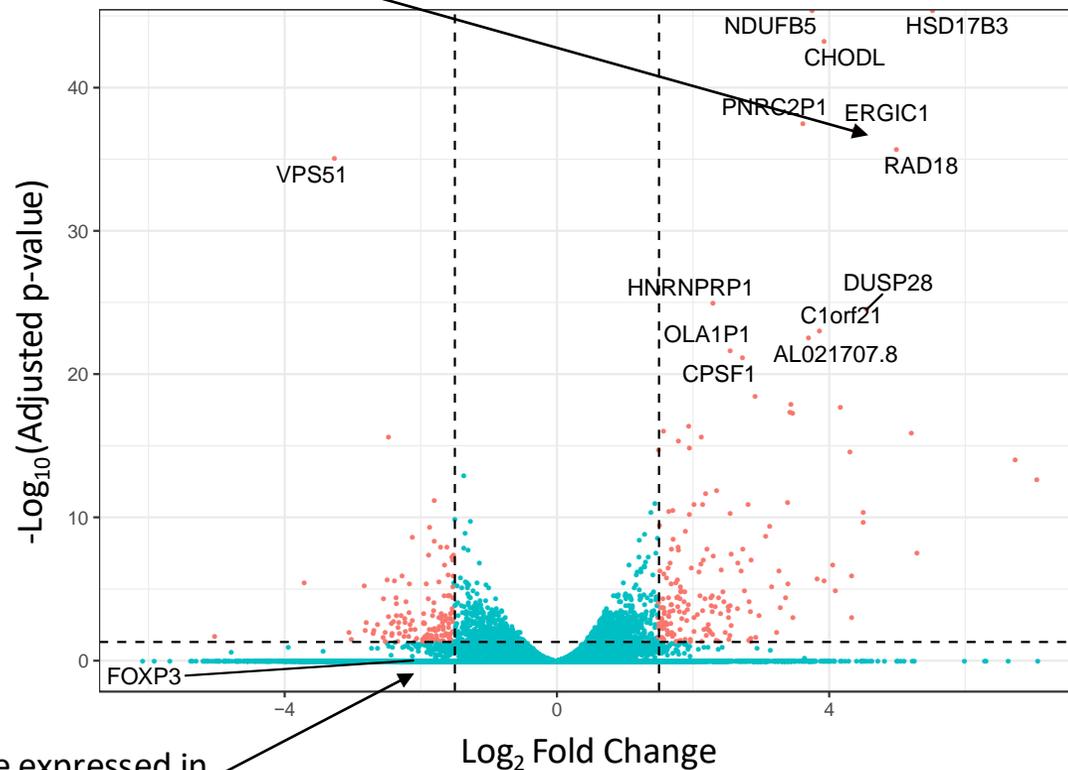


Volcano plot

MA plot

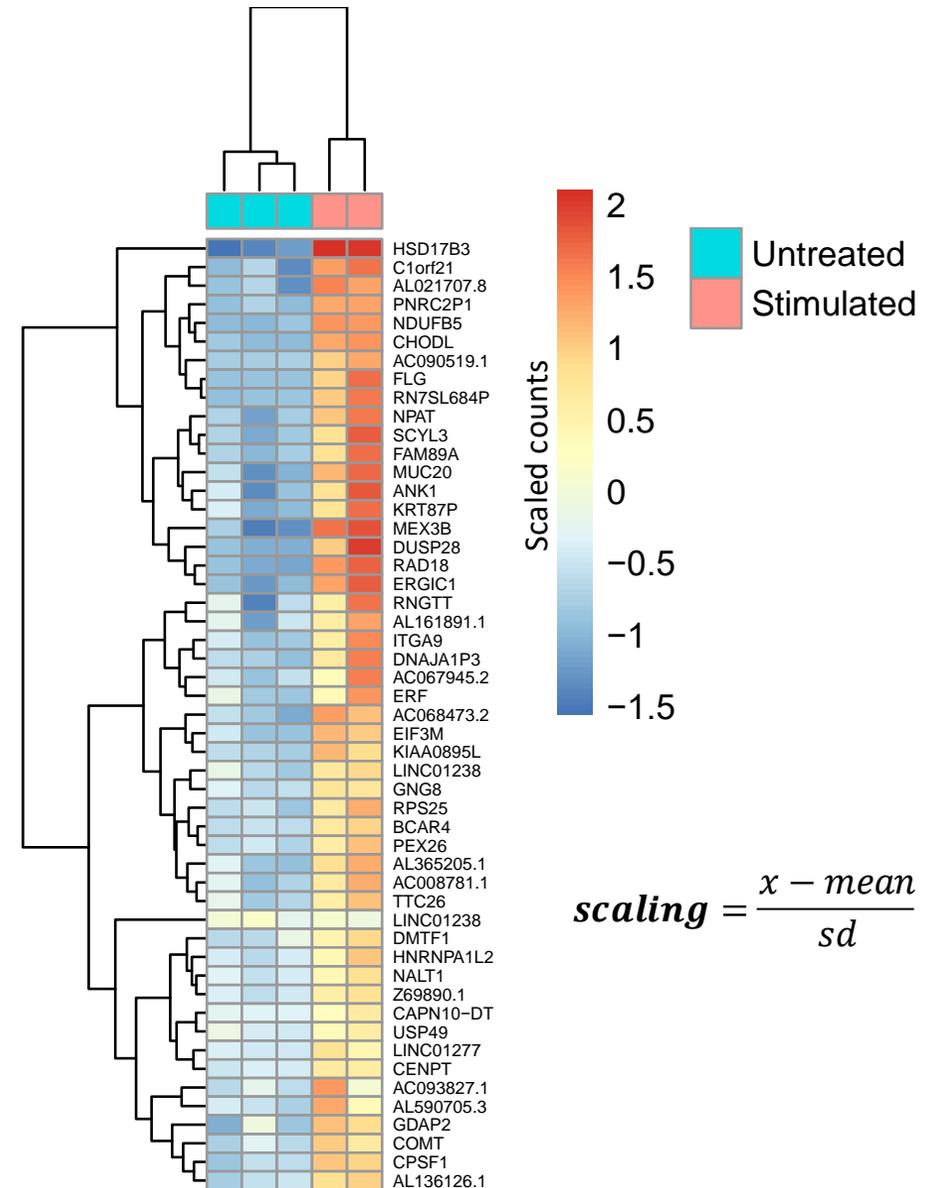


Volcano plot



Heatmap

- Heatmap of normalized expression counts
- Select a set of genes you are interested in!
- Here up-regulated genes (LFC > 1) in stimulated vs untreated cells
- **Scaling** is used for visualization only, NOT for downstream analysis!



Exercise

Let's visualize our DEA results!

Notebook:

- *07c_DEA_visualization.Rmd*

