

# Output description for differ expression analysis

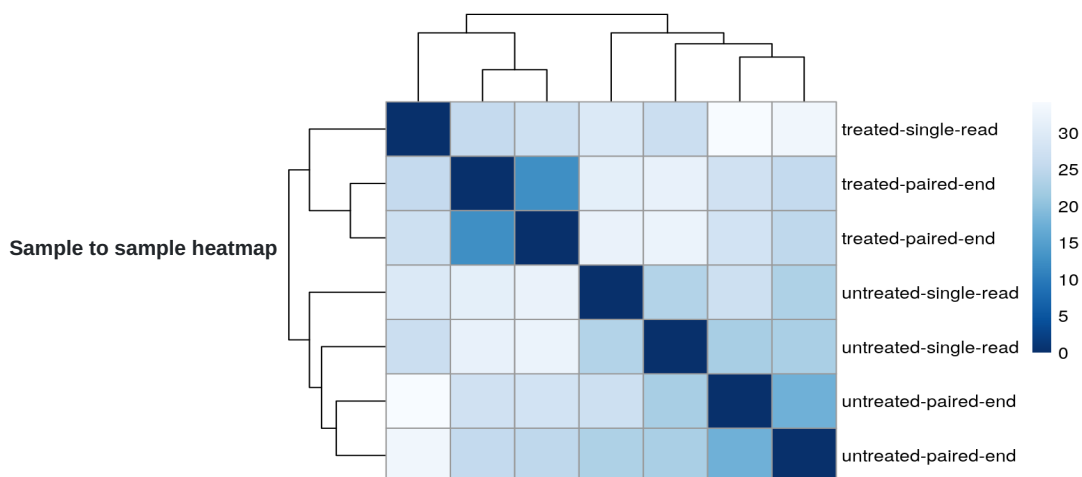
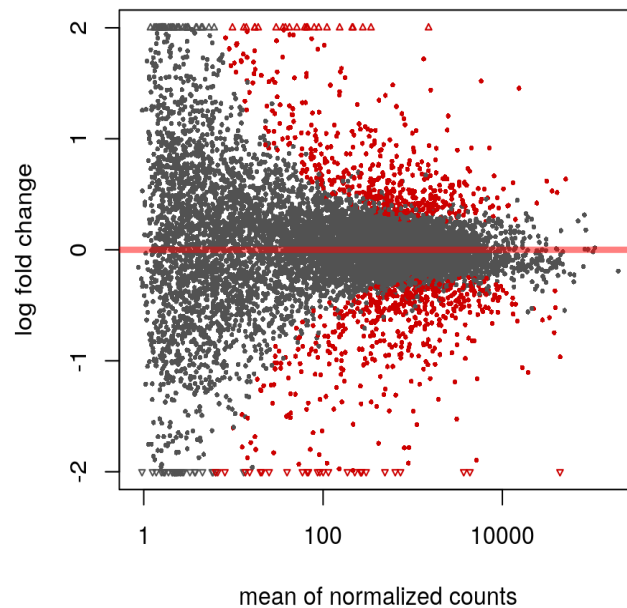
## Pipeline overview

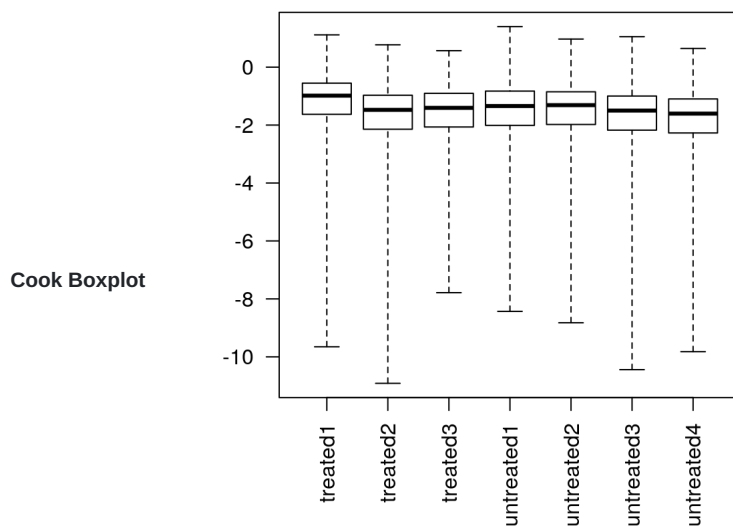
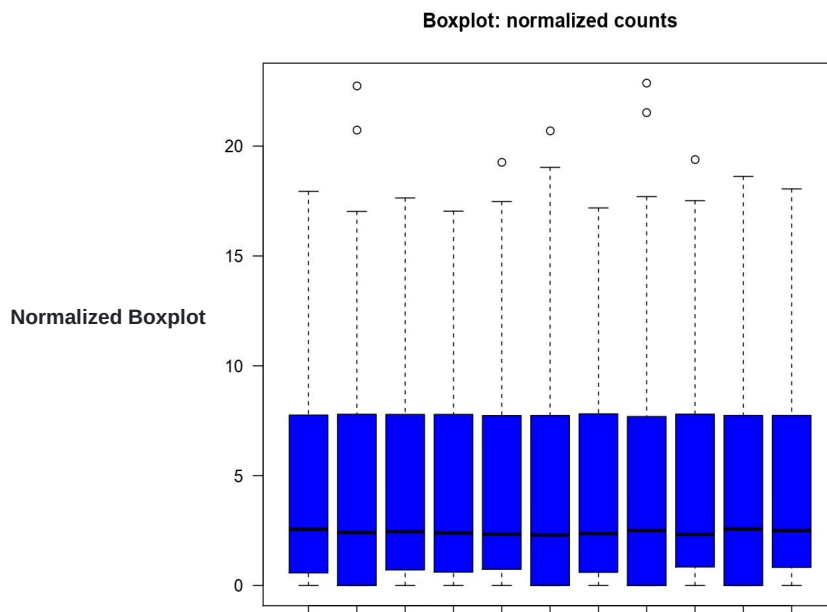
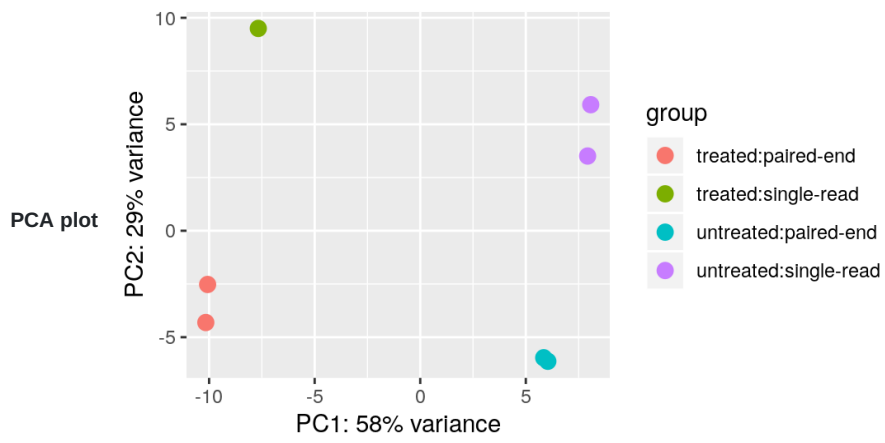
- [DESeq2 v1.18.1](#) - Differential expression analysis and plots

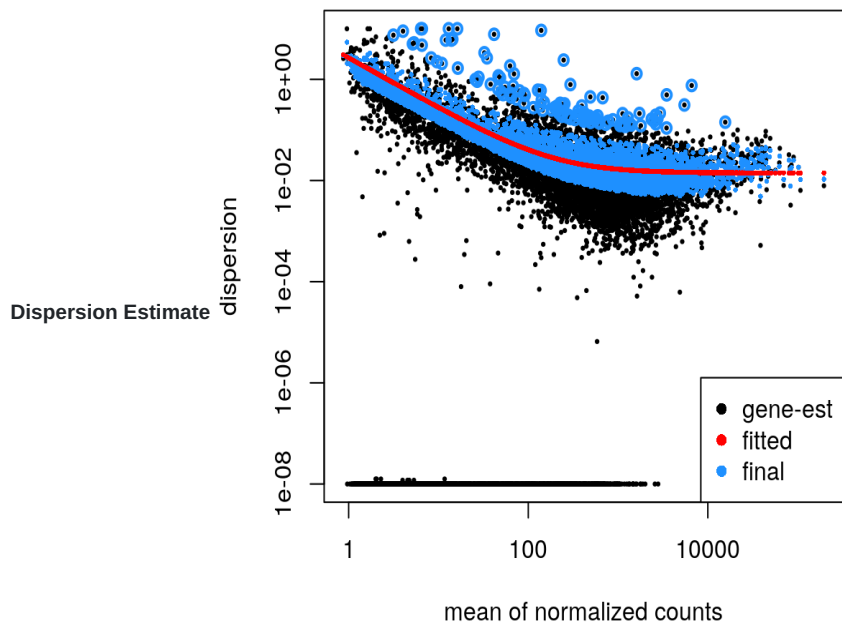
### DESeq2

Differential expression analysis with DESeq2. [DESeq2 \[12\]](#) is a Bioconductor package for R used for RNA-seq data analysis. The script included in the pipeline uses DESeq2 to normalize read counts and create a heatmap / dendrogram showing pairwise euclidean distance (sample similarity). It also creates other plots to evaluate the sample dispersion. It also provides PCA plots to evaluate sample grouping.

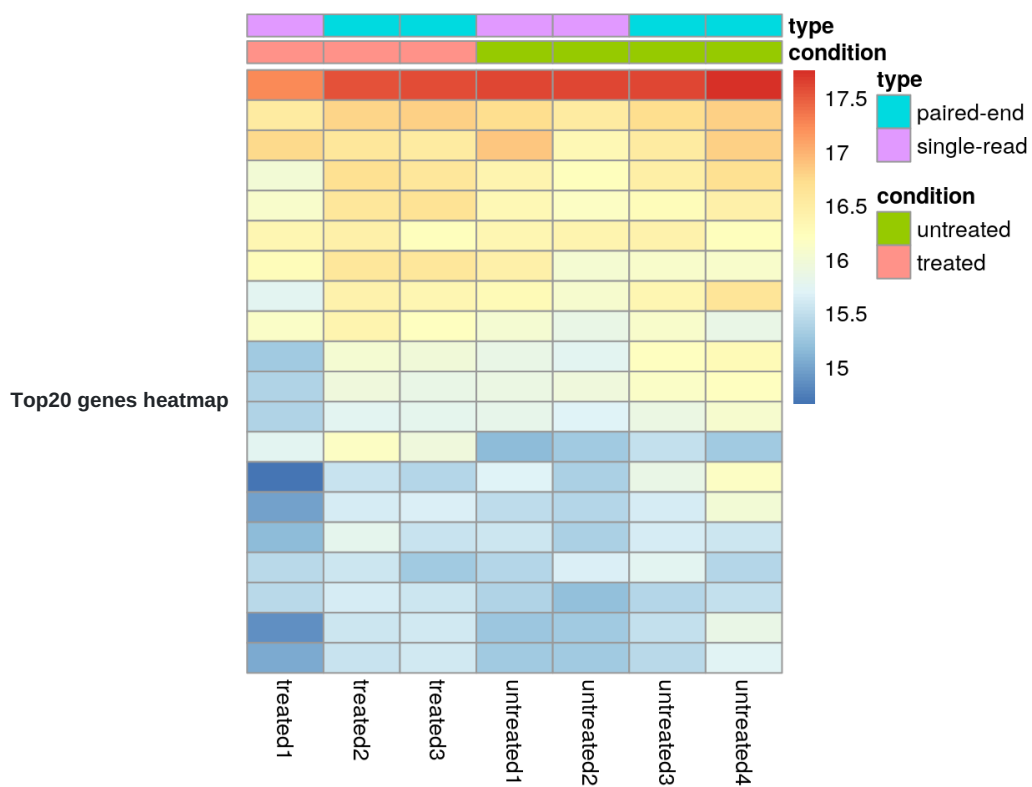
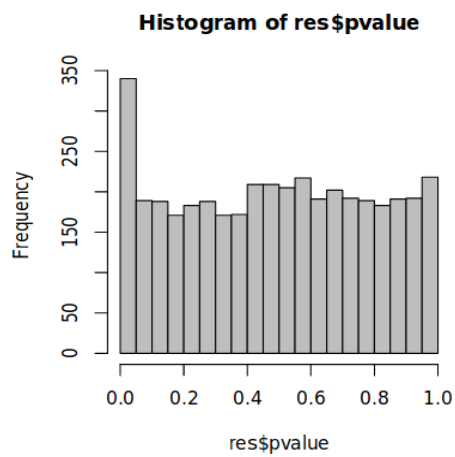
### MA plot



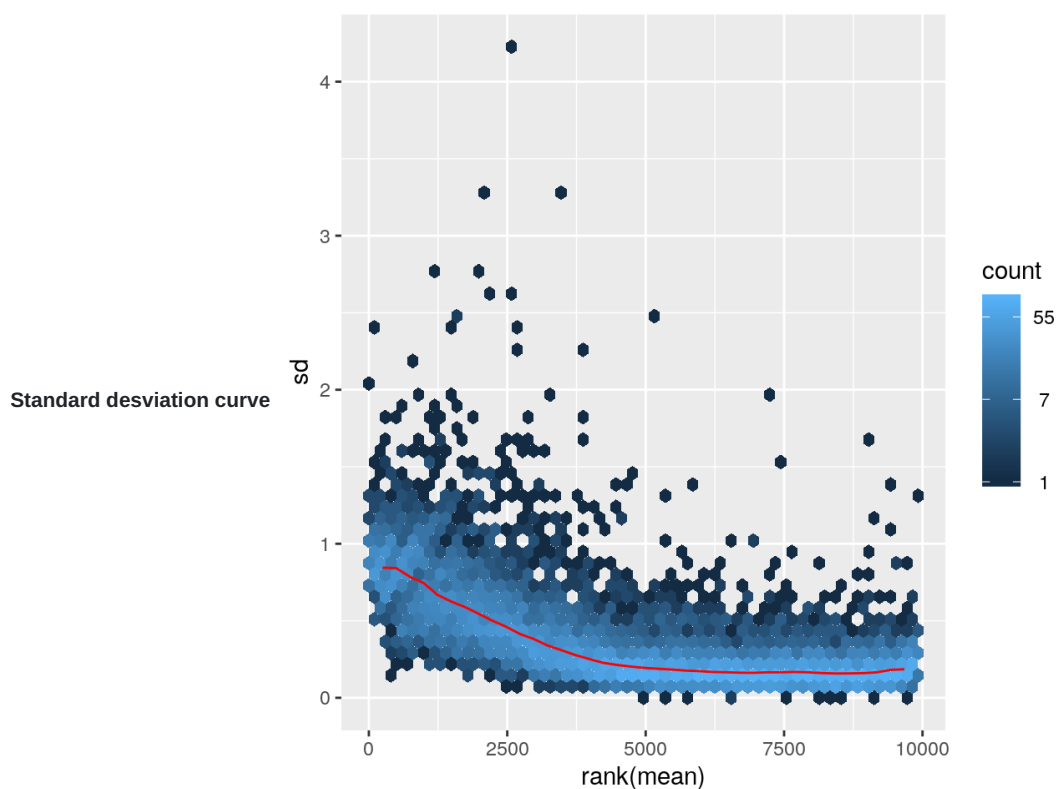
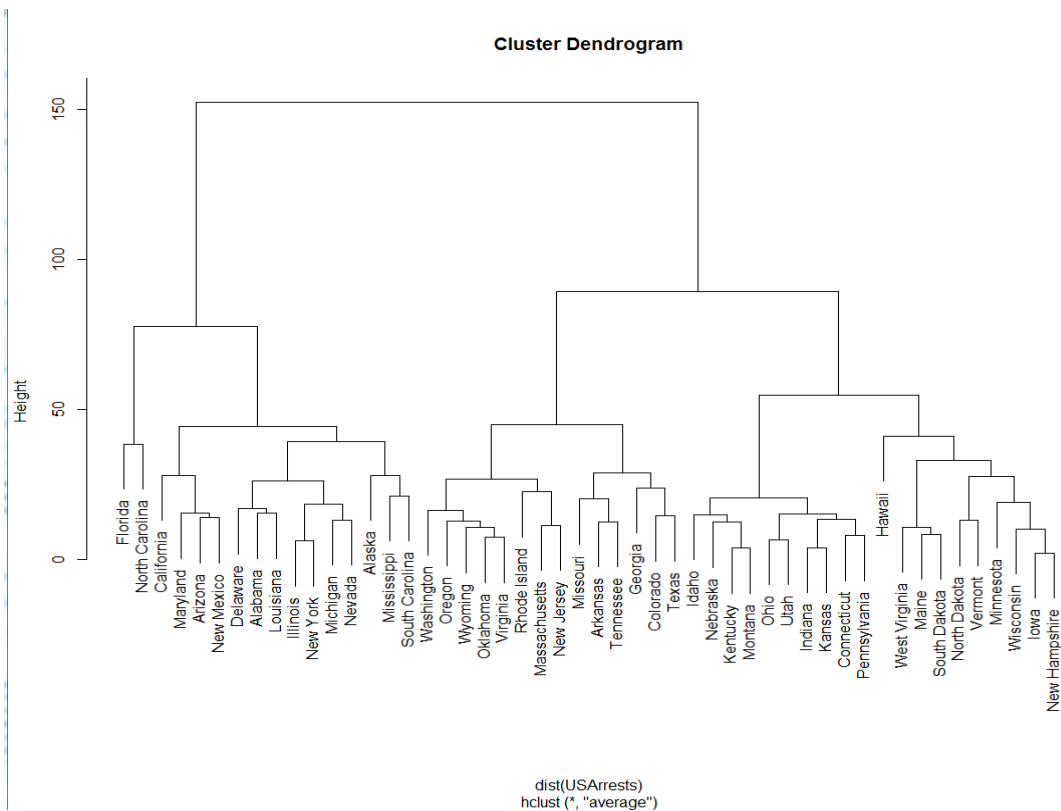




Pvalue test histogram



Hierarchical clustering



#### Output files:

- DE\_matrix.txt
  - Comparative table with the differential expression of two conditions.
- maPlot.pdf
  - MA plot of the DESeq analysis results for all the samples
- heatmap\_sample\_to\_sample.pdf
  - Heatmap with the euclidean distance between samples.
- plotPCA.pdf
  - PCA plot of the samples for the rlog and the vsd.
    - rlog refers to the regularized log transformation, which transforms the count data to the log2 scale in a way which minimizes differences between samples for rows with small counts, and which normalizes with respect to library size.
    - vsd refers to variance stabilizing transformation (VST), which calculates a variance stabilizing transformation (VST) from the fitted dispersion-mean relation(s) and then transforms the count data (normalized by division by the size factors or

normalization factors), yielding a matrix of values which are now approximately homoskedastic (having constant variance along the range of mean values). The transformation also normalizes with respect to library size.

- boxplot.pdf
  - PDF file with the box\_plots
    - Box plot of the normalized Counts
    - Box plot of the counts cook distances to see if one sample is consistently higher than others.
- plotDispersions.pdf
  - PDF file with plots to analyze the dispersion of the samples
    - Dispersion calc is the per-gene dispersion estimate together with the fitted mean-dispersion relationship.
    - Histogram with the test of the differential expression pvalues
- hierarchical\_clustering.pdf
  - PDF file with the hierarchical clustering of the samples. The input data comes from the normalization of the counts. For the normalization DESeq uses the normalization of the ratios where the counts are divided by sample-specific size factors determined by median ratio of gene counts relative to geometric mean per gene.
- heatmapCount\_top20.pdf
  - Heat map of the top 20 genes with the higher normalized mean count. The normalization is the same as the one of the hierarchical clustering.
- heatmapCount\_all\_genes.pdf
  - Heatmap of the normalized counts of all the genes.
- plotSD.pdf
  - Standard deviation of the transformed data, across samples, against the mean, using the shifted logarithm transformation.--->