

Differential gene expression analysis

DGE Analysis

Data

- RNA-Seq data
 - At least 3 biological replicates per condition + control samples!
 - Infection experiment
 - Different strains and treatment(s)
 - ...
- Genome Assembly or transcriptome assembly of the species

Pre-processing

- Quality control of RNA-Seq data
 - FastQC
 - Contaminations?
- Trimming of RNA-Seq data
 - Trimmomatic
- Mapping against genome assembly or transcriptome assembly
 - HISAT2, Star, Kallisto,...
 - Against genome assembly a spliced alignment must be computed
- Count mapped reads with FeatureCounts or just use Kallisto

Overview & Normalization methods

- https://hbctraining.github.io/Training-modules/planning_successful_rnaseq/lessons/sample_level_QC.html
- RPKM, FPKM, TPM nicely explained: <https://www.rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/>

DGE Analysis with TPMs

$$\text{TPM} = A \times \frac{1}{\sum(A)} \times 10^6$$

$$\text{Where } A = \frac{\text{total reads mapped to gene} \times 10^3}{\text{gene length in bp}}$$

TPM is proportional to RPKM ⁴,

$$\text{TPM} = \frac{RPKM}{\sum(RPKM)} \times 10^6$$

- Normalizes for gene length & sequencing depth
- Python Code for TPM: https://www.reneshbedre.com/blog/expression_units.html
- **Does not replace statistical analysis with DeSeq2 or EdgeR, but often needed for visualization!**

DGE Analysis with DeSeq2 or EdgeR

- DeSeq2 or EdgeR
- R packages!
- DeSeq2 is very powerful, but difficult to use/understand
 - Needs 3 Replicates!
 - Large community and many documentation/ how to's
 - Template script:
<https://gist.github.com/stephenturner/f60c1934405c127f09a6>
 - Explanation of script: <https://stephenturner.github.io/deseq-to-fgsea/>
- Result: Log2 fold changes (LFC), normalized counts,...

Visualization

- Principal Component Analysis (PCA): <https://www.reneshbedre.com/blog/principal-component-analysis.html>
- Volcano plot: <https://www.reneshbedre.com/blog/volcano.html>
- Heatmap: <https://www.reneshbedre.com/blog/heatmap-python.html>
- UpsetR plot: <https://cran.r-project.org/web/packages/UpSetR/vignettes/basic.usage.html>
- Plot expression of a differentially expressed gene for treatment and control → best for RNA-Seq data of a time course

Pathway analysis

- Pathway enrichment analysis: KAAS <https://www.genome.jp/kegg/kaas/>
- GO term enrichment analysis: TopGo (R package); goatools (python)
 - <https://github.com/tanghaibao/goatools>