

RNA-seq analysis part II

Workshop Lead: Ariel Madrigal Aguirre

Registration link: NA

Approximate duration: 2 hours

Prerequisites:

1. Familiarity with R
2. Access to Rstudio IDE

Summary: (2-3 sentences summarizing the workshop)

Participants will learn how to use DESeq2 to identify differentially expressed genes and become familiar with design matrices. The session will also introduce concepts essential for working with real-world data, such as incorporating covariates and interaction terms into models. Additionally, participants will learn how to apply Over-Representation analysis (ORA) and Gene Set Enrichment Analysis (GSEA) for uncovering biologically meaningful patterns. Hands-on examples and exercises will help reinforce concepts and guide participants through a typical RNA-seq workflow.

Learning Objectives: (List 2-5 learning objectives participants will learn upon completion of this workshop)

1. Understand the statistical framework used by DESeq2
2. Design and configure differential expression models appropriately for RNA-seq data
3. Perform differential gene expression analysis and gene set enrichment analysis (GSEA)
4. Visualize and interpret results from differential expression and enrichment analyses

Dataset:

This workshop will use processed RNA-seq counts from this study (<https://doi.org/10.1016/j.stem.2021.04.028>), where they studied the transcriptional response to SARS-CoV-2 infection in three tissues (cornea, limbus, sclera). The dataset consists of 18 samples, with three replicates per tissue and condition (mock vs CoV-2). To perform GSEA analysis, we will also use annotated gene sets from MsigDB (<https://www.gsea-msigdb.org/gsea/msigdb>).

Content:

1. Differential expression (1 hour)

- a. Statistical modelling concepts used in DESeq2 and its differences with respect to edgeR
- b. Introduction to design matrices for gene expression experiments
- c. Multi-factor designs
 - i. Covariates
 - ii. Interactions
- d. Log-fold shrinkage
- e. Hands-on activity (Rmarkdown)
 - i. Quality control steps (PCA, correlation, distributions of total counts)
 - ii. Standard workflow with DESeq2
 - iii. Exploring results (MA-plots, volcano plots, boxplots, heatmaps)
 - iv. Log-fold shrinkage
 - v. Variations to the standard workflow (covariates and interactions)
 - vi. Exercises:
 - a) Modify the design matrix to include a covariate term
 - b) Access contrasts in DESeq2

2. Break (5 min)

3. GSEA and ORA (45 min)

- a. Statistical concepts in Over-Representation Analysis
- b. Statistical concepts in Gene Set Enrichment Analysis
- c. When to use GSEA or ORA?
- d. Hands-on activity (Rmarkdown)
 - i. Perform GSEA with fgSEA
 - ii. Visualization of GSEA results
 - iii. Perform ORA with EnrichR
 - iv. Visualization of EnrichR results
 - v. Exercises:
 - a) Perform GSEA and ORA with different databases.

4. Closing remarks (10 mins)