**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)**
   1. Statistical modelling concepts used in DESeq2 and its differences with respect to edgeR
   2. Introduction to design matrices for gene expression experiments
   3. Multi-factor designs
      1. Covariates
      2. Interactions
   4. Log-fold shrinkage
   5. Hands-on activity (Rmarkdown)
      1. Quality control steps (PCA, correlation, distributions of total counts)
      2. Standard workflow with DESeq2
      3. Exploring results (MA-plots, volcano plots, boxplots, heatmaps)
      4. Log-fold shrinkage
      5. Variations to the standard workflow (covariates and interactions)
      6. Exercises:
         1. Modify the design matrix to include a covariate term
         2. Access contrasts in DESeq2
2. **Break (5 min)**
3. **GSEA and ORA (45 min)**
   1. Statistical concepts in Over-Representation Analysis
   2. Statistical concepts in Gene Set Enrichment Analysis
   3. When to use GSEA or ORA?
   4. Hands-on activity (Rmarkdown)
      1. Perform GSEA with fGSEA
      2. Visualization of GSEA results
      3. Perform ORA with EnrichR
      4. Visualization of EnrichR results
      5. Exercises:
         1. Perform GSEA and ORA with different databases.
4. **Closing remarks (10 mins)**