

# Independent RNA-seq Analysis

Now that you've learned how to do RNA-seq, you're going to do it entirely on your own. Go to [https://docs.google.com/spreadsheets/d/1dal19nlCrX4PoVzh\\_rjLUfnR1PmBom\\_JwmGS8A3uPQc/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1dal19nlCrX4PoVzh_rjLUfnR1PmBom_JwmGS8A3uPQc/edit?usp=sharing) to sign up for and download one of the datasets. No repeats please. On the same spreadsheet you'll also sign up for a presentation practice time with me Thursday afternoon. You may notice looking at the GEO repository that there are more samples than you downloaded; use only the samples I gave you, trying to keep this simple.

This will be due **Thursday 7/29 by 9AM**. To close out the program, everyone will give short 10 minute presentations, similar to your exploratory data analysis presentations, on your results. The presentation will be the following day **Friday 7/30 at 12PM** and all Coriell employees will be invited.

---

## Analysis and the Report

You will need to submit the **table of your differential expression results** along with an **Markdown document** containing your report. The report should show all the code you used for the analysis as well as describe your conclusions on the data and the differential expression results. The report will need to (at minimum) contain the following:

### A. Introduction

1. Needs to include a short description of the data and the experimental set up, including, but not limited to
  - What drug or inhibitory strategy is used?
  - What protein is inhibited and what does it normally do?
  - What cell line is used?
  - What disease is being studied?
  - What are the conditions tested and how long are the cells treated for (if known)?
2. Also needs to contain a table or list of the samples

### B. Analysis

1. Must cluster the data to check for batch effects and clustering by treatment. In addition to the plots, should have a short paragraph describing your conclusions from the clustering. For this section you must plot:
  - Hierarchical clustering
  - PCA
2. Must do differential expression testing. The resulting table must be submitted along with your report
3. For the differential expression results, you must plot the following along with a sentence or two describing the conclusions you drew from the plots:

- Volcano plot
  - MA plot
  - 4. Finally, to synthesize your results, you must conduct gene set enrichment analysis and include a plot showing the topmost down- and up- regulated pathways, both with all pathways and with collapsed pathways. You should also include a few sentences describing the conclusions you drew from the gene set enrichment analysis results.
- C. Conclusion
1. The report needs to end with a short paragraph describing the results of your analysis
- D. References
1. Must include a reference to the GEO repository
  2. If there's a published paper using the data that must be referenced as well
  3. If you used any other sources to understand the data, paper, or results, they must be referenced as well
- 

## Presentation

Your presentation must include:

- Background on the data
- Initial clustering, either PCA or hierarchical clustering
- Show results using an MA plot, volcano plot, and/or heatmap
- Show overlap of results from different conditions using an UpSet plot if applicable
- Gene set enrichment analysis
- Conclusion and significance of the results
- Should have a slide listing references as well and if any references are applicable to the contents of the slide, please reference them using *Author, Year* in the bottom right corner