Pre-lab

The next workshop in class is for analyzing bulk RNA-Seq gene expression data. We will be using one of the publicly available datasets on GEO: [**GSE230679**](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE230679). This dataset includes bulk RNA-Seq data of mouse Nucleus Accumbens (NAc) at select time points after treatment with LSD, Ketamine, MDMA, or cocaine or saline controls.

Go to the GEO dataset link below and explore the dataset, and then download the TPM data from the GitHub link.

GEO dataset: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE230679>

TPM data: <https://github.com/melsadany/BfB-M5_RNA-seq-tut/blob/main/data/GSE230679_20230425_psychadelic_study_tpm_matrix.csv>

Feel free to explore iDEP before you come to class:

iDEP: <http://bioinformatics.sdstate.edu/idep/#tab-4975-1>

# iDEP tutorial

1. Got to the iDEP server: <http://bioinformatics.sdstate.edu/idep/#tab-4975-1>
2. Click on “select” next to Species and choose Mus Musculus from the new window. Then, click on Dismiss, and make sure it shows next to species. If it doesn’t, repeat this step again.

A screenshot of a computer

Description automatically generated

1. Click on “Browse” to upload the CSV file you downloaded

A screenshot of a computer

Description automatically generated

1. Once the upload is complete, navigate to the “Pre-process” page, and explore the different plots generated. These plots are to give you an idea about the sequencing depth and any potential bias in your data.

A screenshot of a graph

Description automatically generated

Click on the QC tab and check all given plots. Anything to report here?

1. Click on the PCA tab and report discuss these findings with your table.

A screenshot of a computer

Description automatically generated

1. To do differential gene expression analysis, click on the DEG1 tab and change the fold change value to 1. Then, choose the groups of interest you want to compare. For the in-class demonstration, we will use cocaine14 vs. saline14. You can choose any other group of interest when you upload your results. After checking the box, click on “Submit”.

A screenshot of a computer

Description automatically generated

1. Click on the results tab and compare the number of identified differentially expressed genes (DEGs) with your table.

A screenshot of a computer

Description automatically generated

1. Explore the different tabs in DEG2 tab. Save both the volcano plot and the MA plot with the conditions you’re comparing as the file name with the plot name (e.g., MA-plot\_cocaine14-vs-saline14). You can choose different color and add gene names if you want.

A screenshot of a computer

Description automatically generated

1. Check the different outputs in the Pathway tab. Try the ReactomePA and GSEA (preranked fgsea) methods. Discuss the difference between both methods with your table. Discuss with your table which one of the Pathway databases would be suitable for your research question. Save the network plot with the same naming format mentioned earlier(name: pathway-network\_cocaine14-vs-saline14).

A screenshot of a computer

Description automatically generated

# ICON questions

1. How confident do you feel about processing any RNA-Seq data from GEO on iDEP? (give a number from 1-10; 1 not confident at all and 10 very confident)
2. What research question did you choose? (i.e., what conditions did you compare?)
3. Upload the three plots from your results: volcano plot, MA plot, and the pathway network plot.
4. Which one of these you should use for plotting gene expression?
   1. Raw gene counts
   2. Normalized gene counts
5. What does fold change value mean?
6. Give an example of a possible use for each of these plots in the context of RNA-Seq analysis:
   1. MA plot
   2. Boxplot
   3. Volcano plot
   4. Heatmap
   5. Density plot
7. What are possible areas of improvement for this workshop? (i.e., what concepts could’ve been explained better or in another approach?)
8. Rate this workshop (give a number from 1-10; 1 needs a lot of work and 10 great)

# Pre-lab

* Based on a quick Google search, how many types/methods/techniques are there for RNA-Seq?
* What’s a batch effect? What are the possible sources for it?
* Go to GEO and search for datasets that includes bulk RNA-Seq data for psychedelics. Filter the data to only include data from high throughput sequencing and homo sapiens. Upload a screenshot of your search results.
* Bonus: What’s your understanding of a PCA