**BIOM200 Module 1- Part 1**

Please answer the following questions in 3-4 sentences. You may answer directly in this word document. Only one completed document is needed per group.

1. Describe the differences between single-end and paired-end sequencing. When might we opt to perform a paired-end experiment over a single-end, and vice-versa? What is an index sequence and why do we include it our library?
2. In sequencing experiments, what do we mean by cluster density and library complexity? How might these factors impact the quality of a sequencing experiment? What measures can be taken to alleviate these problems?
3. How is stranded information maintained in an RNA-Seq library prep? What is the difference between a stranded and unstranded library?
4. Before library preparation and sequencing, we need to perform an additional stage of RNA selection. Why? What different methods can be used for RNA selection, and how might the specific type of technique influence the profile of your final library?
5. What are the strengths and limitations of RNA-seq and ATAC-seq? Why might you choose to use one assay over the other?

**BIOM200 Module 1- Part 2**

The following are instructions for your group project. Please perform the following steps as a group. For files and notebooks generated, please provide a full path to their location. You may write the responses in this Word document for submission. Only one copy of both files and responses are needed from each group.

1. Choose two experimental datasets that you wish to compare from the ENCODE website and download them as before. Please use two replicates per condition if possible. Briefly describe the dataset (library type/read length, how RNA was selected, sequencing instrument, strandedness, etc.) and what you aim to learn with your analysis.
2. Perform alignment using STAR using the reference genomes in the shared folder (I have made both mouse and human indexes in the shared folder under (/oasis/tscc/scratch/biom200/bms\_2018/annotations/). Take a look at the Log file you get after successfully running STAR on your samples and make a table summarizing the most important metrics (include in this word document). Please provide a full path to the directory of your STAR mapping results (1 per group). Lastly, in 3-4 sentences, comment on the parameters that you decided to use and why.
3. After running featureCounts and getting you counts matrix, quantify differential expression using DESeq2. How many significantly expressed genes did you identify? Comment on what significance cutoffs you decided to use and why. Generate an MA-plot to summarize your differential expression results. Generate a clustermap for differentially expressed genes and briefly conclude what we can learn from this figure. Copy both images into this word file.