Homework #1: Module 1 (RNA-seq)

For your homework, you will be tasked with using featureCounts (part of the Subread package) to generate a counts matrix to be used for DESeq2 analysis. This will involve reading through the featureCounts documentation (http://bioinf.wehi.edu.au/subread-package/SubreadUsersGuide.pdf) in the read summarization step. This will involve writing your own shell script and submitting it to the hotel queue. This will not designed be a copy-and-paste job, so be sure to read the manual and think carefully about your command.

Specific Instructions:

1. Write a shell script called featureCounts.sh that will read in either the .sam files you aligned yourself OR the .sam files that have been provided in the shared folder( /oasis/tscc/scratch/biom200/cmm262/star\_alignment/ ). For the job, you will require 1 node and 2 processors per node. You can allot 1 hour of walltime. Carefully review the program usage arguments and include those that you decide are necessary for out dataset. Remember the technical details of our experiment (library preparation method, sequencing run type, etc.). Provide us with the full path to the directory of your submission script, as well as the location of your .out and .err files.
2. Successfully submit your job to the cluster. Make sure that your job runs and finishes successfully. If it does not run to completion, assess your error files, correct your mistakes and resubmit the job. Put the output somewhere meaningful and provide us with the **full path** to the directory that contains your final featureCounts output.
3. In one paragraph, describe the flags that you chose to include, as well as the reason that you decided to include them. Refer to your lecture notes for hints.

**General questions:**

1. How is stranded information preserved in an RNA-seq library? Why might you want to preserve stranded information?
2. What are the similarities and differences between RPKM and TPM? Which one might be considered preferable and why?

Ryan Marina: Office hours

Monday (01/15/18): SCRM Lobby 4 – 6 PM