Session 5 Recitation

1) Log into a development node. Make a directory CMSE890_Sec303/Session5 either in your home directory or scratch. Copy the human genome reference files to your directory (home or scratch) on the HPCC (runWget_References.sh). Look through the script and the files to satisfy yourself that you understand what kind of info they contain. If you wanted to submit this script to run on the cluster instead of just running it on a dev-node, what additional lines would you add to the script? Try it.

2) Use **module spider** to look for the Salmon genome mapping software. Look through the script runSalmon_index.sh and figure out how to convert it to work in your directory. Submit the job to the queue and figure out how to keep track of it. If it finishes during class time, go through the log files (runSalmon_index.sh.eJobID or .oJobID). If it finishes successfully, you should see a file called versionInfo.json in the results directory.

3) Take the raw data files in /mnt/research/CMSE-bioinformatics/week3/fastq and attempt to align one set of them with Salmon. Use the script runSalmon_single.sh as a guide and then try changing the resources. You should see a quant.sf file in the results directory when it is finished.

4) Look at the runSalmon_single.sh script used to submit one sample set at a time. How would you modify the script to do multiple samples at once? Then consider whether you would want to instead write a bash script to submit several single sample jobs automatically.