Exercises for Workshop W5a: RNA-seq I Analysis

**PRIOR TO CLASS:**

1. Make sure you have access to hoffman2. You can do this easily here: <https://www.hoffman2.idre.ucla.edu/Accounts/Requesting-an-account.html>
2. Once you have an account, we need a few software packages that are not available on hoffman2. I’ll provide step-by-step instructions.

First, we have to install FastQC:

* In your home directory, run the following command:

wget https://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc\_v0.11.9.zip --no-check-certificate

unzip fastqc\_v0.11.9.zip

To check if FastQC installed properly:

~/FastQC/fastqc --help

Next, we have to install trimmomatic:

wget http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.39.zip

unzip Trimmomatic-0.39.zip

To check if trimmomatic installed properly:

qrsh -l h\_rt=0:30:00,h\_data=5G

module load java

java -jar Trimmomatic-0.39/trimmomatic-0.39.jar

Lastly, we have to install Salmon:

wget https://github.com/COMBINE-lab/salmon/releases/download/v1.5.1/salmon-1.5.1\_linux\_x86\_64.tar.gz

tar xzvf salmon-1.5.1\_linux\_x86\_64.tar.gz

To check if Salmon installed properly:

salmon-1.5.1\_linux\_x86\_64/bin/salmon

Exercise 1: UNIX warm-up

1A: Logging in (connecting) to Hoffman-2

To log into the cluster, first open a terminal (OSX: Terminal app; Windows: Ubuntu app, for instance),

then enter the following command:

ssh my\_username@hoffman2.idre.ucla.edu

# (Note: Replace `my\_username` with your personal Hoffman2 user/account

↪ name.)

Important: Every time you connect to the Hoffman-2 cluster, you will initially have a session on the

login node (login server). Because this computer is shared among all users and has limited computing resources, the first thing you should do after logging into the cluster is to ask for an interative session

on a worker node, by running the command:

qrsh -l h\_rt=2:00:00,h\_data=8G

# This asks for a 2-hour session, allowing up to 8 GB of memory.

After a few seconds, you should have been given an interactive session on a worker node. You can

tell that you have moved by looking at the server name in your command prompt: the name of the

Hoffman-2 login node is login2, whereas worker nodes have names like n2190 or n2236.

1B: Obtaining the workshop data

Once you are connected to the cluster and logged into a worker node, make a copy of the workshop

data in your “home” directory:

cp -R /u/project/collaboratory/nrochett/QCBio\_RNAseq1/ ~/

There should now be a QCBio\_RNAseq1/ directory within your home directory, with the following

contents: