Here we go to analyze RNA seek data.   
Following the steps of “Dariya Sydykova”

1. Log in to phylocluster.   
   (phylocluster is a cluster of CCBB)  
     
   How to log in?  
   Type “ssh phylocluster.ccbb.utexas.edu” into terminal.  
   Note to log into lonestar we were writing  
   “ssh ucaglar@lonestar.tacc.utexas.edu”  
     
   Change into /share/WilkeLab/AG3C/data/experiments/ directory. Change into the directory for the desired time course. Copy the data structure in that directory using this function:  
     
   tar -cvf hydra\_data\_structure.tar sample\*/RNA/  
   tar function combines multiple files and directories into a single file  
     
   Tar is something like “zip”  
   Here are some properties  
     
   -c: Create a new archive containing the specified items.  
   -v: Produce verbose output. In create and extract modes, tar will list each file name as it is read from or written to the archive. In list mode, tar will produce output similar to that of ls(1). Additional -v options will provide additional detail.  
   -f: file. Read the archive from or write the archive to the specified file. The filename can be - for standard input or standard output.
2. Log in to @lonestar.tacc.utexas.edu. Change into scratch and make a "data" directory. Change into "data" and make a folder for a given time course (ex:glucose). Copy the data structure file from hydra to that directory:  
    *cd $SCRATCH  
   mkdir data  
   cd data  
   mkdir glucose\_time\_course  
   cd glucose\_time\_course  
   scp username@phylocluster.ccbb.utexas.edu:/share/WilkeLab/AG3C/data/experiments/\*\_time\_course/hydra\_data\_structure.tar .*  
     
   The scp function corresponds to secure copy  
   -- secure copy (remote file copy program)  
     
   scp: copies files between hosts on a network.   
   It uses ssh(1) for data transfer, and uses the same authentication and provides the same security as ssh(1). Unlike rcp(1), scp will ask for passwords or passphrases if they are needed for authentication.  
     
   File names may contain a user and host specification to indicate that the file is to be copied to/from that host. Local file names can be made explicit using absolute or relative pathnames to avoid scp treating file names containing `:' as host specifiers. Copies between two remote hosts are also permitted
3. Once you copied the tar file into that directory, unfold the tar file using:  
     
   *tar -xvf hydra\_data\_structure.tar*Tar -x: Extract to disk from the archive. If a file with the same name appears more than once in the archive, each copy will be extracted, with later copies overwriting (replacing) earlier copies.

1. You should now have the exact same data structure in $SCRATCH/data/\*\_time\_course/ as in the /share/WilkeLab/AG3C/data/experiments/\*\_time\_course. Make a folder unanalyzed\_raw\_reads in the same directory to store unanalyzed raw reads pertaining to that time course. Copy all of the reads into that directory.

5. Change into Ecoli\_RNAseq/src/. Run this shell script to make sample\*/RNA/\*depleted.raw and move the raw reads there (change TIME\_COURSE\_DIR variable in move\_raw\_reads.sh to match the time course directory your data is in) by typing:

./move\_raw\_reads.sh

6. Concatenate reads, if multiple fastq files are present for one sample's each read. Run this function in the same directory (Ecoli\_RNAseq/src/) to write a paramlist file that will concatenate all the reads:

./write\_paramlist\_concatenate.sh

7. Change into $SCRATCH/data/ and run the concatenation:

chmod +x paramlist\_concatenate

./paramlist\_concatenate

This will take about 20 mins. The concatenated reads would appear in $SCRATCH/\*\_time\_course/sample\*/RNA/\*depleted.processed/

7. Now you can make a paramlist file to run the bowtie pipeline in parallel for each sample. Change into Ecoli\_RNAseq/src/ and run this function (change TIME\_COURSE\_DIR variable in write\_paramlist\_bowtie.sh to match the time course directory your data is in):

./write\_paramlist\_bowtie.sh

8. Change into Ecoli\_RNAseq/job\_submissions\_files/. Correct the line in bowtie\_launcher.sge that says "setenv CONTROL\_FILE paramlist\_bowtie\_\*". "paramlist\_bowtie\_\*" is the file that was written by running ./write\_paramlist\_bowtie.sh. "\*" here stands for the time course you are analyzing.

9. Now you are ready to submit the job to TACC. Change into $SCRATCH/data/ and run:

qsub ~/Ecoli\_RNAseq/job\_submissions\_files/bowtie\_launcher.sge

You can check the status of your job by entering "qstat" function. All of the files generated by the pipeline will appear in SCRATCH/\*\_time\_course/sample\*/RNA/\*depleted.processed/.

10. Once the pipeline is done running, you need to combine all the quality control files into one file. Change into the \*\_time\_course (ex: glucose\_time\_course) and run:

cat sample\*/RNA/\*depleted.processed/\*\_quality\_control.txt > quality\_control\_glucose.csv

Open this file in excel and delete every other line that contains headers from each sample (cat function adds all of the lines in each sample).

11. To copy the new data generated to hydra, change into Ecoli\_RNAseq/src/. Change DATA\_DIR, SAMPLES, DEST\_DIR, and TAR\_NAME in both make\_tar\_processed.sh and make\_tar\_raw.sh. Run both of them:

./make\_tar\_processed.sh

./make\_tar\_raw.sh

This scripts will write tar files for raw and processed read files (ex:TAR\_NAME\_processed.tar, TAR\_NAME\_raw.tar).

12. Step 11 also generates md5sum files (TAR\_NAME.md5sum). These are needed to make sure that the tar files copied to hydra are identical to the ones on lonestar. Change into the directory with tar files and copy them to hydra:

scp TAR\_NAME\_processed.tar TAR\_NAME\_raw.tar username@phylocluster.ccbb.utexas.edu:/share/WilkeLab/AG3C/data/experiments/\*\_time\_course/\*

13. Log in to hydra and change into the directory where tar files are. Run:

md5sum TAR\_NAME\_processed.tar > TAR\_NAME\_processed.md5sum

md5sum TAR\_NAME\_raw.tar > TAR\_NAME\_raw.md5sum

Check that the value given in the TAR\_NAME\_\*.md5sum matches the value given in the TAR\_NAME\_\*.md5sum on lonestar directory. If they are not the same, tar files need to be recopied unto hydra.

14. Unfold the tar files on hydra:

tar -xvf TAR\_NAME\_processed.tar

tar -xvf TAR\_NAME\_raw.tar