The following is a guide for analyzing RNASeq data as received from the core. It has been optimized for Mac but should be able to be translated to Windows with some changes to the software used to access the terminal or remote server.

Download :

[“R”](https://www.r-project.org/)

[RStudio](https://www.rstudio.com/products/rstudio/) (the free version is all we need)

[iTerm2](https://iterm2.com/)

[Filezilla](https://filezilla-project.org/)

[Contact hpc3 and get a login account for the supercomputer. Instructions are here.](https://rcic.uci.edu/hpc3/hpc3-reference.html)

Fastq files are received from the sequencing core as txt.gz

These are zipped files and are large, leave them zipped. The code will unzip them as they use them.

1. First thing we must do is trim the adapters that are used for the sequencing reaction.

Code for this is found here:

[**Trimmomatic**](https://github.com/erichanse/RNASeqCode/blob/main/Kong_RNASeq/Trimming.py)

1. Next we will check the trimmed sequences for quality using the code found here:

[**Fastqc**](https://github.com/erichanse/RNASeqCode/blob/main/Kong_RNASeq/FastQC.R)

This will give us html breakdown of the sequencing quality before we proceed with the genome alignment.

1. Build the reference genome. Download the version of the genome you will be using to align your reads. Human, mouse etc. I found the version of the mouse genome at [UCSC](https://genome.ucsc.edu/cgi-bin/hgGateway).
2. For alignment we need to build the reference genome and load our samples onto the server. To do this, I use Filezilla (link is above and connect to the host: sftp://hpc3.rcic.uci.edu

Login using your uci credentials

Create a directory file and move your reference genome file, your gene annotation file and your trimmed and QC tested sequences to new and separate directories on the server.

Code for all of this is found here: [Alignment code](https://github.com/erichanse/RNASeqCode/blob/main/Kong_RNASeq/Alignment.sh)

STEP 1 is generating the genome reference assembly with annotations:

Open a SLURM session: In order for these giant files to be processed we have to use the UCI supercomputer and we have to get our job into the queue to use the processing power. This requires us to tell the supercomputer what we need.

Code for the SLURM is found here: [SLURM](https://github.com/erichanse/RNASeqCode/blob/main/Kong_RNASeq/Slurm_STAR.sh)

Save this file to your directory with the mycode.sh file extension

Run command:

sbatch mycode.sh

##where mycode is your SLURM directed version of STAR genome assembly

You will be given a jobID

Check progress with squeue -u $User

#where $User is your username

With the parameters I have in the code this process took a long time use more than one thread I think it should go faster.