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"

RStudio (the free version is all we need)

iTerm2

Filezilla

Contact hpc3 and get a login account for the supercomputer. Instructions are here.

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

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

    Fastqc

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

3. 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.

4. For alignment we need to build the reference genome and load our samples onto the server. To do this, I use Filezilla and connect to the host: sftp://hpc3.rcic.uci.edu
    Login using your uci credentials
    Create a directory file and move your reference genome files and your trimmed and QC tested sequences to new directories on the server
    Code for all of this is found here: Alignment code