

**TEXT in GREEN is a specific argument for each function/parameter**

**Download sra for ubuntu**

wget -output-document SRAtoolkit.tar.gz <https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz>

**Extract the contents of the tar file:**

tar -vxzf sratoolkit.tar.gz

##### Append the path to the binaries to your PATH environment variable:

export PATH=$PATH:$PWD/sratoolkit.3.0.0-mac64/bin

##### Verify that the binaries will be found by the shell:

##### which fastq-dump

##### Download fastqc files and unzip contents

##### ls

##### srrxxxx\_fastq.gz1 srrxxxx\_fastq.gz2

##### gunzip \*.fastqc.gz\*



*First we need to inspect for quality*

**Install “fastqc” binary**

sudo apt-get update

sudo apt-get install fastqc

**Perform Quality Control using fastqc**

fastqc <my\_file.ext> -o <output\_folder>



*Next we can trim off adapters and sequence we don’t need with “porechop”*

**Install “Porechop” binary**

sudo apt-get install Porechop

**Trim the sequences using “Porechop”**

Porechop my\_fastq\_file > trimmed\_output\_filename



*We ready to align but first well need an indexed reference geneome and convert to binary. We can index with “Bowtie” and align our reads with “bwa” and convert this human readable outputs to binary using samtools. This will allow us to Compare differential expression of our data to things like a GTF or GFF files for a given genome.*

**Install “bowtie” and bwa binaries**

sudo apt get install bowtie

sudo apt get install bwa

**Index a genome using “bwa”**

bwa index -p genome <my\_genome\_file.fa>

**Align trimmed reads to indexed Genome**

bwa mem -t 2 trimmed\_output\_filename > output\_name.sam

**Convert human readable “sam” format to Binary “bam” format**

sudo apt install samtools

samtools view -S -b output\_name.sam > my\_alignment.bam

samtools sort -O <my\_sort.bam> <my\_alignment.bam>

**Obtain and decompress a “gtf” files**

*If you’re looking for information on gene structure or RNA expressions for a geneome gtf.gz files contain this*

gunzip Sarrchoromyces\_cerevisiaw.R64-1-1.107.gtf.gz

**Mark or index the sorted binary file**

samtools index <my\_sort.bam>

*From here we can perform Differential expression using our annotated genome sequence binary which includes our reads and our unpacked GTF or GFF file using htseq*

**Install htseq**

sudo apt get install python.htseq

**check if the sam file is sorted using samtools**

samtools view -f 4 my\_alignment.sam

**Count the reads using the gtseq-count and save to a file**

**htseq-count -–f bam** <my\_sort.bam> | Sarrchoromyces\_cerevisiaw.R64-1-1.107.gtf > my\_output.ham