

Processing Pipeline of EOAD and NDC samples

1. Run FastQC on raw FASTQ samples

```
#Run fastqc for the FASTQ files
(base) [prvaldes@tscc-login2 ~]$ qsub -l -q condo -l walltime=8:00:00 -l
nodes=2:ppn=24
qsub: waiting for job 27215286.tscc-mgr7.local to start
qsub: job 27215286.tscc-mgr7.local ready
(base) [prvaldes@tscc-13-37 Chen_Foundation_FASTQ]$ fastqc *.fastq.gz
Started analysis of CN19009_sc15_b2_AD_S68_L004_R1_001.fastq.gz
Approx 5% complete for CN19009_sc15_b2_AD_S68_L004_R1_001.fastq.gz
Approx 10% complete for CN19009_sc15_b2_AD_S68_L004_R1_001.fastq.gz
[example]
Output = .fastqc.html files and .fastqc.zip files
```

2. Run MultiQC on raw fastqc samples

```
#Run MultiQC for the raw FastQC files
(base) [prvaldes@tscc-login2 ~]$ source multqc_env/bin/activate
(multqc_env) (base) [prvaldes@tscc-login12 FastQC]$ multiqc
/home/prvaldes/scratch/Chen_Foundation_FASTQ/FastQC
[WARNING]      multiqc : MultiQC Version v1.11 now available!
[INFO ]      multiqc : This is MultiQC v1.8
[INFO ]      multiqc : Template   : default
[WARNING]      multiqc : You are running MultiQC with Python 2.7.5
[WARNING]      multiqc : Please upgrade! MultiQC will soon drop support for
Python < 3.6
[INFO ]      multiqc : Searching   :
/home/prvaldes/scratch/Chen_Foundation_FASTQ/FastQC
[INFO ]      fastqc : Found 54 reports
[INFO ]      multiqc : Compressing plot data
[INFO ]      multiqc : Report     : multiqc_report.html
[INFO ]      multiqc : Data      : multiqc_data
[INFO ]      multiqc : MultiQC complete
```

3. Run Trimgalore!

Batch #1

```
(base) [prvaldes@tscc-login11 ~]$ qsub -l -q condo -l walltime=8:00:00 -l
nodes=2:ppn=24
qsub: waiting for job 27224483.tscc-mgr7.local to start
qsub: job 27224483.tscc-mgr7.local ready

(base) [prvaldes@tscc-0-49 Trimgalore]$ bash Trimgalore_RNA_EOAD.NDC.sh
```

4. Run MultiQC on Trimmed, Validated Files

Processing Pipeline of EOAD and NDC samples

```
(base) [prvaldes@tscc-login2 ~]$ (base) [prvaldes@tscc-login11  
clean_Chen_Foundation_FastQC_ALL]$ source  
/home/prvaldes/multqc_env/bin/activate  
(multqc_env) (base) [prvaldes@tscc-login11  
clean_Chen_Foundation_FastQC_ALL]$ multqc  
/home/prvaldes/scratch/Trimgalore/clean_Chen_Foundation_FastQC_ALL
```

5. Make transcriptome index with a kmer length of k=31

#Get cDNA file

```
(base) [prvaldes@tscc-login11 homo_sapiens_104]$ wget  
ftp://ftp.ensembl.org/pub/release-  
104/fasta/homo\_sapiens/cdna/Homo\_sapiens.GRCh38.cdna.all.fa.gz
```

#Get non-coding RNA file

```
(base) [prvaldes@tscc-login11 homo_sapiens_104]$ wget  
ftp://ftp.ensembl.org/pub/release-  
104/fasta/homo\_sapiens/ncrna/Homo\_sapiens.GRCh38.ncrna.fa.gz
```

#Concatenate both files together into one file

#Notes from here: <https://www.biostars.org/p/81924/>

```
(base) [prvaldes@tscc-login11 homo_sapiens_104]$ cat  
Homo_sapiens.GRCh38.cdna.all.fa.gz Homo_sapiens.GRCh38.ncrna.fa.gz >  
Homo_sapiens.GRCh38.cdna.all.ncrna.fa.gz
```

#Start the screen in a new window

screen

#Submit interactive jobs to the home-shankar

```
[prvaldes@tscc-login2 ~]$ qsub -l -q condo -l walltime=8:00:00 -l nodes=2:ppn=24  
qsub: waiting for job 27223067.tscc-mgr7.local to start  
qsub: job 27223067.tscc-mgr7.local ready
```

#Build the transcriptome index using kmer count of 31 using Kallisto

#kallisto index builds an index from a FASTA formatted file of target sequences.

```
(base) [prvaldes@tscc-0-49 homo_sapiens_104]$  
/home/prvaldes/programs/kallisto/kallisto index -k 31 -i  
Homo_sapiens.GRCh38.cdna.all.release-104_k31.idx  
/home/prvaldes/programs/homo_sapiens_104/Homo_sapiens.GRCh38.cdna.all.ncrn  
a.fa.gz
```

6. Run Kallisto

```
(base) [prvaldes@tscc-login11 ~]$ qsub -l -q home-shankar -l walltime=48:00:00  
-l nodes=1:ppn=24
```

```
(base) [prvaldes@tscc-2-13 Kallisto]$ bash KallistoScript_RNA_EOAD.NDC.sh
```

7. Run MultiQC on Kallisto Files

Note: runs on kallisto.log files

```
(base) [prvaldes@tscc-login2 ~]$ (base) [prvaldes@tscc-login11  
clean_Chén_Foundation_FastQC_ALL]$ source  
/home/prvaldes/multiqc_env/bin/activate  
(multiqc_env) (base) [prvaldes@tscc-login11  
clean_Chén_Foundation_FastQC_ALL]$ multiqc  
/home/prvaldes/scratch/KallistoOut_RNA_Chén
```

8. Proceed with downstream quantified transcript counts from Kallisto RNA- using EOAD.RNAseq.Kimma.nVenn.Analysis.Rmd script file.