DATA THEARPY

Workflow : RAW-reads to Counts file

March 8, 2019
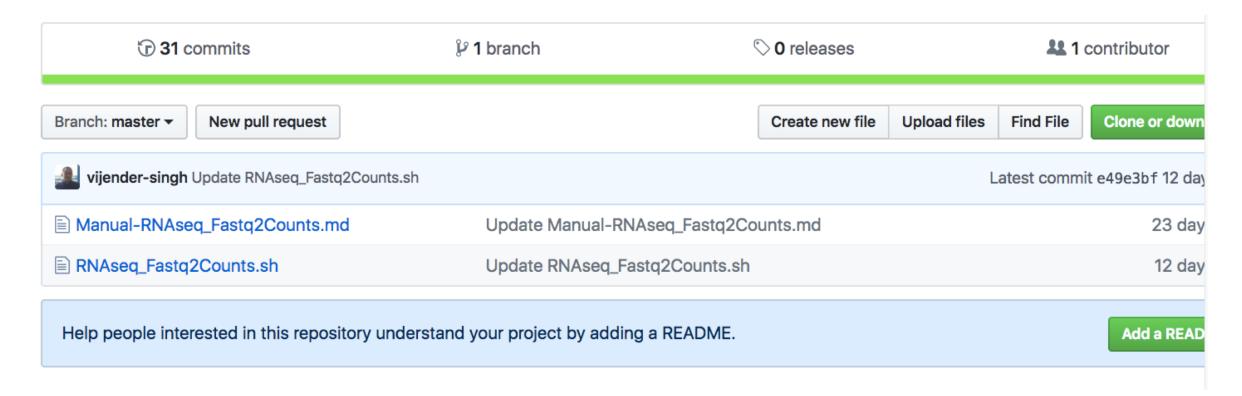
mRNA Data Analysis Pipeline

RNAseq_Fastq2Counts.sh

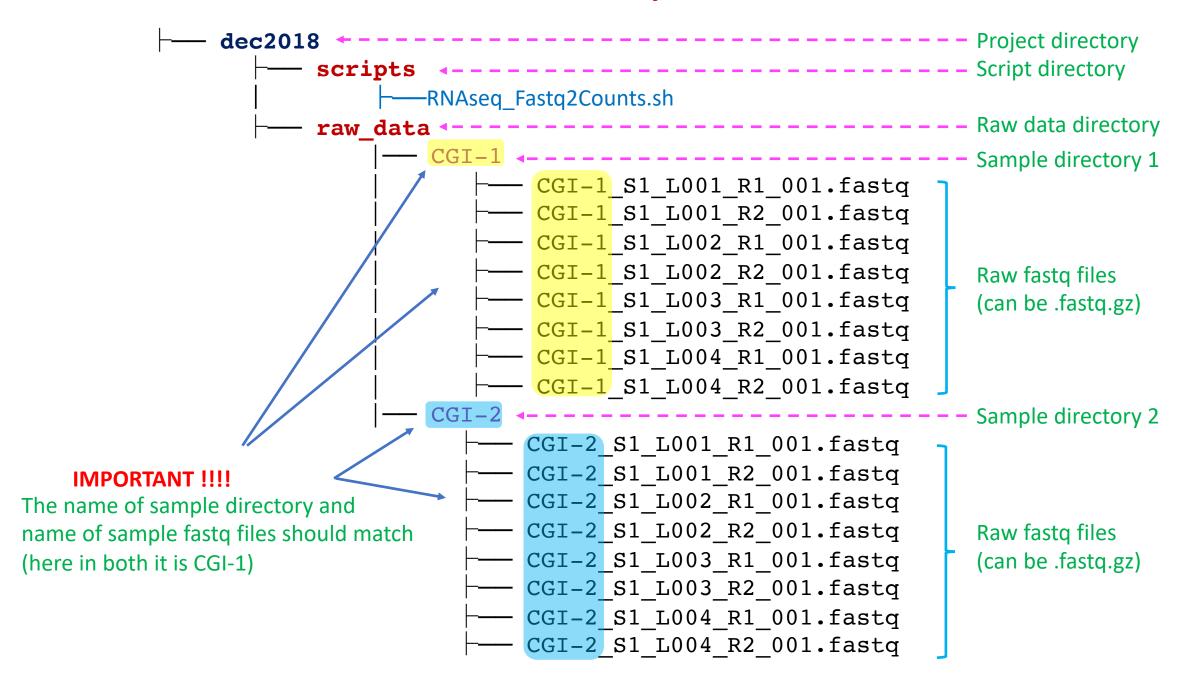|	
Merging Reads	
Quality Assessment	FastQC — Babraham Bioinformatics
Trimming	Trimmomatic/Sickle
Quality Assessment	FastQC — Babraham Bioinformatics
Mapping to a Reference	Bowtie2/HISAT2/STAR
Counting reads per gene	HTSeq
Differential Gene Expression	DESeq2 — Bioconductor OPEN SOURCE SOFTWARE FOR BIOINFORMATICS
GO Term Enrichment	BiNGO
Submit to SRA	NCBI

Workflow script

https://github.com/vijender-singh/Workflow_Scripts
ON Xanadu cluster: /UCHC/LABS/CBC/workflow/RNAseq/RNAseq_Fastq2Counts.sh

Requires a directory architecture explained in next slide for 2 samples CGI-1 and CGI-2

Directory Architecture

```
sbatch RNAseq_Fastq2Counts.sh \
    -s <SampleName> \
    -p <Path/to/project directory> \
    -m <human|mouse>
```

```
e.g.
sbatch RNAseq_Fastq2Counts.sh -s CGI-1 -p /UCHC/LABS/vsingh/dec2018 -m mouse

sbatch RNAseq_Fastq2Counts.sh -s CGI-2 -p /UCHC/LABS/vsingh/dec2018 -m mouse
```

```
├──── dec2018
       ├─── scripts
       ├─── raw_data
       ├─── logs
              ├─── CGI-1_log_trace
              ├ ── CGI-1_OUT
              ├─── CGI-2_log_trace
              ├─── CGI-2_OUT
```

CGI-1_log_trace

```
CMD1=OK
CMD2=OK
CMD3=OK
CMD4=OK
CMD5=OK
CMD6=OK
CMD7=OK
CMD8=OK
CMD9=OK
CMD10=OK
CMD11=OK
CMD12=OK
CMD13=OK
CMD14=OK
```

CGI-1_OUT

```
UNCOMPRESSSING STEP: Either Uncompressing of fastq files was successfully
 executed in previous run OR there are no compressed files.
MERGING R1 FASTQ FILES: Merging of R1 reads was successfully executed in
previous run OR NOT REQUIRED
MERGING R2 FASTQ FILES: Merging of R2 reads was successfully executed in
previous run OR NOT REQUIRED
VALIDATING READS IN PAIRED FILES: File merging Authenticated in previous
run
FASTQC, quality check was executed in previous run
SICKLE : Read trimming was successfully executed in previous run
VALIDATING READ PAIRS POST SICKLE: Read counts post Trimming was successf
ully executed in previous run
FastQC post trimming was successfully executed in previous run
=========================================================

 ANALYSIS COMPLETED
```

NOTE: In log_trace files the status of the commands (CMD) can be either **OK** or **FAIL.**

OK : Command executed successfully

FAIL: Either the command failed due to error or that command is not executed yet. If script stop running check OUT file
      and SLURM out/error files for details.

```
├──── dec2018
      ├──── counts
      ├──── fastqc_processed
      ├──── fastqc_raw
      ├──── logs
      ├──── mapping
      ├──── merged_fastq
      ├──── raw_data
      ├──── scripts
      ├──── tmp
      └──── trimmed_reads
```