From: Koth, Laura
To: Koth, Laura
Subject: FW: rna seq current pipeline
Date: Thursday, April 9, 2020 3:21:01 PM

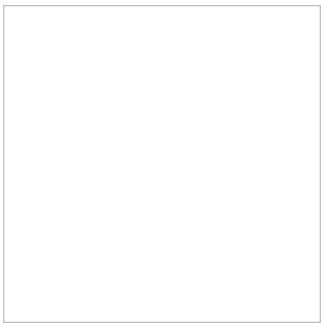
From: Nerella, Srilaxmi <Srilaxmi.Nerella@ucsf.edu>
Sent: Tuesday, April 7, 2020 10:30 AM
To: Koth, Laura <Laura.Koth@ucsf.edu>
Cc: Christenson, Stephanie <Stephanie.Christenson@ucsf.edu>
Subject: Re: rna seq current pipeline

Hi Laura,
This is a summary of our pipeline(Read me file). Please feel free to ask me if you have any questions:

Step 1: FastQC: quality control checks on raw sequence data
This is the first step of Q/C. We need to run the tool and go through the results of the FastQC tool for all samples
Syntax to run FastQC:
fastqc -o <directory to write output files> <directory path of fastq file>
e.g: fastqc -o /raid/data/vap_dataset_2/fastqc_dir
/raid/data/vap_dataset_2/fastq_files/Sample_A04_R1_001.fastq.gz
- prepare a list of samples(e.g: Sample_A04) and give it as a command-line argument to the bash script 'run_fastqc.sh'
e.g: sh run_fastqc.sh sample_list
check Fastqc output for all samples to make sure that the quality is good.

Step2: Cutadapt, Sickle and STAR steps:

Command: sh run_pipeline.sh samples
Cutadapt: finds and removes adapter sequences, primers, etc., from fastq reads
Sickle: A windowed adaptive trimming tool for .fastq files using quality
STAR: Spliced Transcripts Alignment to a reference



Step 3: script to concatenate all gene counts files
Once the star alignment runs for a batch of samples is completed, the gene counts can be concatenated or merged into a single file using the script 'concatenate_gene_counts.py'.
command: python concatenate_gene_counts.py
The above python script looks in the 'Alignments' directory for star gene counts file(files ending with ReadsPerGene.out.tab) and merges them into a 'counts.txt' file. The other output file when the script was run is 'stats.txt'

Step 4: make the summary file:
* use the merge_log_final_out_files.pl script to put all the output from the log files together(the script prints the commands to paste and enter on the command line). count the raw read counts and put them together.

Thanks,
Sri