Marques et al 2021 SMARTSeq Preprocessing

Contents

Preprocessing of data in bash

```
ls
```

Adapter trimming using fastp

```
find -name "*.fastq.gz" | while read A; \
do fastp --adapter_fasta adapters.fa -i $A -o "${A}_trimmed.fastq.gz" \
-h "${A}_report.html" --thread 10; done
```

rename the trimmed files and reports

```
for x in $(find . -name "*.fastq.gz_trimmed.fastq.gz"); \
do
  mv $x $(echo "$x" | sed 's/\.fastq.gz_trimmed.fastq.gz$/_trimmed.fastq.gz/')
done
```

```
for x in $(find . -name "*.fastq.gz_report.html"); do
  mv $x $(echo "$x" | sed 's/\.fastq.gz_report.html$/_fastp_report.html/')
done
```

FASTQC and MulitQC after adapter trimming

```
pwd
mkdir -p fastqc_trimmed

fastq_files=${fastq_file_dir}
inputFiles=Sample_S*/*_trimmed.fastq.gz
task (){

echo "running fastqc on ${1}"

fastqc ${1} --outdir fastqc_trimmed
echo "fastqc for ${1} is done"
}



N=10
(
for SAMPLE in $inputFiles
do
((i=i%N)); ((i++==0)) && wait
task "$SAMPLE" &
```

```
done
)
```

```
multiqc ./fastqc_trimmed/ --outdir fastqc_trimmed
```

Alignment and Gene counts

Alignment using STAR using the genome with fluorophore prepared Ensmble GRCz11 DanRer11 v102 and gene counts using feature counts

```
mkdir -p ./star_alignments_and_counts

PATH=$PATH:/home/prateek/Mercader_Lab/STAR-2.7.1a/source

# with counts from star
STAR --genomeLoad LoadAndExit --genomeDir ./Ensmbl/GRCz11/v102/star_index/GRCZ11_Ensmbl_v102_star_index
for i in $(ls Sample_S*/*_trimmed.fastq.gz | sort -u); do
STAR --genomeDir ./Ensmbl/GRCz11/v102/star_index/GRCZ11_Ensmbl_v102_star_index \
--readFilesIn ${i} \
--runThreadN 10 \
--outFileNamePrefix ./star_alignments_and_counts/${i:33:-9}_star_ \
--outSAMtype BAM SortedByCoordinate \
--outSAMunmapped Within \
--quantMode GeneCounts \
--readFilesCommand zcat \
--sjdbGTFfile ./Ensmbl/GRCz11/v102/Danio_rerio.GRCz11.102_fp_validated.gtf \
--outSAMattributes Standard;

#for i in $(ls Sample_S*/*_trimmed.fastq.gz | sort -u); do
STAR --genomeDir ./Ensmbl/GRCz11/v102/star_index/GRCZ11_Ensmbl_v102_star_index \
--readFilesIn ${i} \
--runThreadN 10 \
--outFileNamePrefix ./star_alignments_and_counts/${i:33:-9}_star_ \
--outSAMtype BAM SortedByCoordinate \
--outSAMunmapped Within \
--quantMode GeneCounts \
--readFilesCommand zcat \
```

```
--sjdbGTFfile ./Ensmbl/GRCz11/v102/Danio_rerio.GRCz11.102_fp_validated.gtf \
--outSAMattributes Standard; done


STAR --genomeLoad Remove --genomeDir ./Ensmbl/GRCz11/v102/star_index/GRCZ11_Ensmbl_v102_star_index

PATH=$PATH:/home/prateek/Mercader_Lab/subread-2.0.1-source/bin

featureCounts -T 10 -a ./Ensmbl/GRCz11/v102/Danio_rerio.GRCz11.102_fp_validated.gtf -t exon -g gene_id \
-o featurecounts_counts_all.txt ./star_alignments_and_counts/*.bam
```

Rename feature counts files

```
mv featurecounts_counts_all.txt SMART_Laura_featurecounts_counts_all.txt
mv featurecounts_counts_all.txt.summary SMART_Laura_featurecounts_counts_all_summary.txt
```