**AR sequencing analysis pipeline**

#========= General location ==============

Download location from Basespace: ‘Sequencing/Runs/’

Default file format after download:

daisy-00-1D1P1C1P1C1\_L001-ds.fd23f8d8edd44ca6b50b75e6d7d74825/

Elliott-Daisy\_S1\_L001\_R1\_001.fastq.gz

Elliott-Daisy\_S1\_L001\_R2\_001.fastq.gz

Problem: For Interrogate, we need a portion from the folder (daisy-00-1D1P1C1P1C1) PLUS a portion from the file (\_R1\_001.fastq.gz)

#========= Copy scripts to correct location ==============

Go to:

https://github.com/theKellerLab/IntrgAnalysis

Clone or Download > Download ZIP

Unzip and copy to fastq files folder (e.g. 'Sequencing/)

#====== Combine fastq in one folder =============

bash combineAllFastqInOneFolder.sh

Output format: Elliott-Daisy\_S1\_L001\_R1\_001.fastq.gz

#====== run QC =============

mkdir CombinedFastqGzOriginalQC

for f in CombinedFastqGzOriginal/\*; do fastqc $f --outdir=CombinedFastqGzOriginalQC/; done

#====== run multiQC =============

#run multiqc

multiqc CombinedFastqGzOriginalQC/

mv multi\* CombinedFastqGzOriginalQCMulti/

#====== trim reads =====

https://bioinformaticsdotca.github.io/HTSeq\_2017\_IA\_lab

https://github.com/bioinformaticsdotca/HTSeq\_2017/blob/master/integrative\_assigment\_commands.sh

http://www.usadellab.org/cms/?page=trimmomatic

export SOFT\_DIR=/usr/local/

export TRIMMOMATIC\_JAR=$SOFT\_DIR/Trimmomatic-0.38/trimmomatic-0.38.jar

#===== trim multiple sequences SE ==========

mkdir CombinedFastqGzTrimmed15

for f in CombinedFastqGzOriginal/\*; do echo $f; filename="${f##\*/}"; echo $filename; \

java -jar $TRIMMOMATIC\_JAR SE -phred33 $f CombinedFastqGzTrimmed15/$filename \

ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36; \

done

mkdir CombinedFastqGzTrimmed30

for f in CombinedFastqGzOriginal/\*; do echo $f; filename="${f##\*/}"; echo $filename; \

java -jar $TRIMMOMATIC\_JAR SE -phred33 $f CombinedFastqGzTrimmed30/$filename \

ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:30 MINLEN:36; \

done

#===== trim multiple sequences PE ==========

Input Read Pairs: 84 Both Surviving: 51 (60.71%) Forward Only Surviving: 31 (36.90%) Reverse Only Surviving: 1 (1.19%) Dropped: 1 (1.19%)

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ wc Original/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

336 420 30829 Original/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ wc Original/Elliott-Daisy\_S1\_L001\_R2\_001.fastq

336 420 30903 Original/Elliott-Daisy\_S1\_L001\_R2\_001.fastq

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ sc SE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ wc SE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

124 155 8091 SE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ wc SE/Elliott-Daisy\_S1\_L001\_R2\_001.fastq

4 5 209 SE/Elliott-Daisy\_S1\_L001\_R2\_001.fastq

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ wc PE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

204 255 14837 PE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq

Stefans-MacBook-Pro-3:daisy-00-1D1P1C1P1C1 SKeller$ wc PE/Elliott-Daisy\_S1\_L001\_R2\_001.fastq

204 255 12943 PE/Elliott-Daisy\_S1\_L001\_R2\_001.fastq

path=”FastqGzOriginal/daisy-00-1D1P1C1P1C1\_L001-ds.fd23f8d8edd44ca6b50b75e6d7d74825/”

mkdir CombinedFastqGzTrimmed15

for f in CombinedFastqGzOriginal/\*; do echo $f; filename="${f##\*/}"; echo $filename; \

path=”FastqGzOriginal/daisy-00-1D1P1C1P1C1/”

#====change folder structure

targetFolder=”FASTQ\_Generation\_2018-12-19\_19\_32\_44Z-144643634\_test/\*”

for folder in $targetFolder; do echo $folder; done

java -jar $TRIMMOMATIC\_JAR PE -phred33 \

FastqGzOriginal/daisy-00-1D1P1C1P1C1/Original/Elliott-Daisy\_S1\_L001\_R1\_001.fastq.gz \

FastqGzOriginal/daisy-00-1D1P1C1P1C1/Original/Elliott-Daisy\_S1\_L001\_R2\_001.fastq.gz \

FastqGzOriginal/daisy-00-1D1P1C1P1C1/PE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq.gz \

FastqGzOriginal/daisy-00-1D1P1C1P1C1/SE/Elliott-Daisy\_S1\_L001\_R1\_001.fastq.gz \

FastqGzOriginal/daisy-00-1D1P1C1P1C1/PE/Elliott-Daisy\_S1\_L001\_R2\_001.fastq.gz \

FastqGzOriginal/daisy-00-1D1P1C1P1C1/SE/Elliott-Daisy\_S1\_L001\_R2\_001.fastq.gz \

ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

#========= redo QC ==========================

mkdir CombinedFastqGzTrimmed15QC

for f in CombinedFastqGzTrimmed15/\*; do fastqc $f --outdir=CombinedFastqGzTrimmed15QC/; done

CombinedFastqGzTrimmed30QC

for f in CombinedFastqGzTrimmed30/\*; do fastqc $f --outdir=CombinedFastqGzTrimmed30QC/; done

#====== run multiQC =============

multiqc CombinedFastqGzTrimmed15QC

mv multi\* CombinedFastqGzTrimmed15QCMulti/

multiqc CombinedFastqGzTrimmed30QC

mv multi\* CombinedFastqGzTrimmed30QCMulti/

#========= upload to Interrogate ============