Pre-processing the raw sequencing data

System requirement:

* macOS X Mojave Version 10.14 with python 2.7
* galaxy local version: 17.09
* FASTQ Trimmer version 1.1.1
* FASTQ Groomer version 1.1.1
* FastQC Hight Throughput Sequence QC Report version 0.11.8

Installations:

* Galaxy
  + https://galaxyproject.org/admin/get-galaxy/
  + configure the local serve according to the instruction
* FASTQ Trimmer, FASTQ Groomer
  + Install Galaxy tools according to the instructions
  + https://galaxyproject.org/admin/tools/add-tool-from-toolshed-tutorial/
* FastQC
  + https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

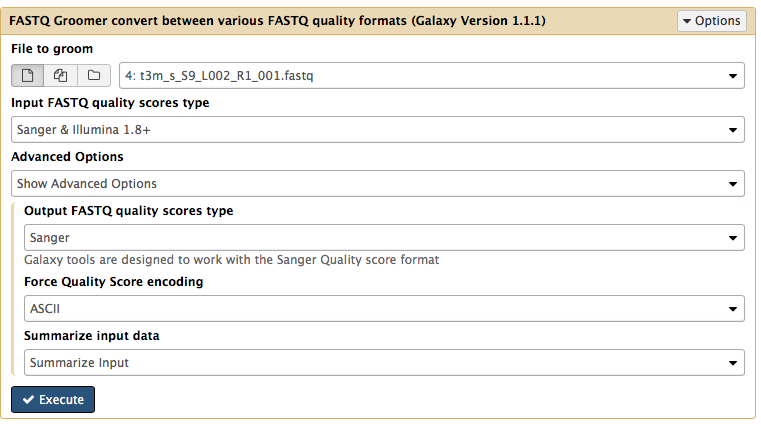
Instruction:

* open terminal
* cd galaxy (change directory to the local galaxy folder)
* sh run.sh (launch the local galaxy server)
* open a web browser, copy and paste the local server address:

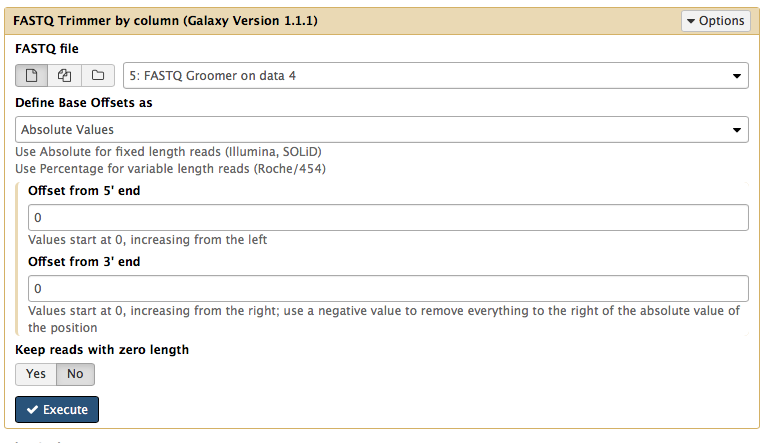
http://127.0.0.1:8080

* Upload the raw data into the Galaxy local server
* Run FASTQ Groomer (Galaxy)

1. File to groom: select the .fastq file
2. Input FASTQ quality scores type: Sanger & Illumina 1.8+
3. Advanced Options: Show Advanced Options
4. Output FASTQ quality scores type: Sanger
5. Force Quality Score encoding: ASCII
6. Summarized input data: Summarize input
7. Start by pressing "Execute"



* Run FASTQ Trimmer (Galaxy)
  1. Select the groomed FASTQ file
  2. Define Base Offsets as: Absolute Values
  3. Offset from 5' end: 3
  4. Offset from 3' end: 0
  5. Keep reads with zero length: No
  6. Start by pressing "Execute"



* Save the processed sequence files and check the quality by FastQC (command line)
  + fastqc $DIR/$file
  + $DIR: directory where the processed sequence located
  + $file: the name of the sequence file

Pseudo-alignment with Kallisto

System requirement:

* macOS X Mojave Version 10.14
* Kallisto version 0.45.1

Installation:

* https://pachterlab.github.io/kallisto/manual

Building the mouse index:

* The mouse reference genome was downloaded with:

http://www.ensembl.org/Mus\_musculus/Info/Index

* commend line for building the index:

kallisto index -i mouse-index-k21 --kmer-size=21 Mus\_musculus.GRCm38.cdna.all.fa.gz

Pseudoalignment:

* commend line:
* cd $DIR (change directory to where the processed sequence files located)
* run kallisto: (please note: the location of the index file)

kallisto quant -i dir/mouse-index-k21 -o Output-filename --single -l 48 -s 10 input\_sequence.fastq

example:

kallisto quant -i mouse-index-k21 -o Output-LP13\_S79\_L007\_R1\_001 --single -l 48 -s 10 LP13\_S79\_L007\_R1\_001.fastq

R analysis

System requirement:

* macOS Mojave 10.14.6
* R version 3.6.0 (2019-04-26)
* Platform: x86\_64-apple-darwin15.6.0 (64-bit)

Installation:

* R

https://cran.r-project.org/

* R studio

https://rstudio.com/products/rstudio/download/

Installation of R packages:

* Launch R studio:
  + install.packages("readr")
  + install.packages("tidyverse")
  + install.packages("dplyr")
  + install.packages("ggplot2")
  + install.packages("gplots")
  + install.packages("calibrate")
  + install.packages("fdrtool")
  + install.packages("BiocManager")
  + install.packages("RColorBrewer")
  + install.packages("vsn")
  + install.packages("tidyr")
  + install.packages("plotly")
  + install.packages("shiny")
  + install.packages("rmarkdown")
  + install.packages("ClassDiscovery")
  + BiocManager::install('tximport')
  + BiocManager::install('rhdf5')
  + BiocManager::install('DESeq2')
  + BiocManager::install('ensembldb')
  + BiocManager::install('GenomicFeatures')
  + BiocManager::install('biomaRt')
  + BiocManager::install("genefilter")
  + BiocManager::install("topGO")
  + BiocManager::install("org.Mm.eg.db")
  + BiocManager::install("Rgraphviz")

Instructions:

* Scripts with .R ending:
  + Launch R studio
  + source(“scriptname.R”)
* scripts with .Rmd ending:
  + Launch R studio
  + open the script in R studio, and click “Knit”, default output document is html, default output directory is the Document directory.