

BIOS6660

Homework 6: RNA-Seq Preprocessing

Due: Tuesday March 12, 2019 at 10:30 am

In this assignment, you will be able to pre-processing paired-end RNA-Seq data into a dataset ready to be analyzed.

Instructions for turning in assignment: Submit a GitHub URL through Canvas

Data: On yampa under `/BIOS6660/Homework6/rawReads` there are 4 samples fastq files from a paired-end read RNA-Seq experiment. The naming convention follows the general format: `sampleName_R1_subset.fastq.gz`

R1 for the first read and R2 are for files with the following paired read.

Please note I took a subset of reads from the original files to speed up running time.

I also have created index files needed for hisat2 and rsem. They are located on yampa under: `/BIOS6660/Homework6/indexFiles`

Assignment: Prepare the 4 samples RNA-Seq data to be analyzed. More specifically, trim, check for quality, align to the genome and quantitate. Put together a cohesive Rmarkdown report showing these steps and the code used. If you are referencing a bash script, please upload the bash script as well to canvas.

The following software packages are installed on yampa: cutadapt (for trimming), fastqc (for QC), hisat2 (for alignment) and rsem (for quantitation).

I do not care what software packages you use for each of the processing steps, but I will not help you install packages that are currently not on yampa.