Homework day 4 – FastQC, Trimmomatic, HISAT2 and IGV.

Authors: Mary Allen & Daniel Ramirez

FastQC & Trimmomatic

1. Edit the template.sbatch to run fastQC on the files:
Day4HW_1.fastq

found in the directory /scratch/Workshop/4_qc/fastq/

Output the FastQC file to your user directory.

What can you say about the reads?
Is there adaptor content that needs to be trimmed out?

2. Edit the template.sbatch to run Trimmomatic on the files:
Day4HW_1.fastq

Trim the reads so that they have a minimum length of 30 base pairs.

Save the output files as 'Day4HW_1.trimmed.fastq'

These are single-end read files, with phred 33 quality scores and the library was made using TruSeq3-SE adapters, whose fasta file is on /opt/trimmomatic/0.36/adapters

3. Run FastQC again on the output file:
Day4HW_1.trimmed.fastq

Are these files trimmed well? Is there adaptor content? What parameters could you adjust to make trimming more stringent?

Hisat2 & IGV

1. Edit the template.sbatch to run map the two paired-end fastq files:
Day4HW_R1.fastq
Day4HW_R2.fastq
found in the directory /scratch/Workshop/SR2019/4_qc/homework

2. Visualize the resulting sorted.bam files using IGV.
What happens to the mapping efficiency if you play with hisat2 preset options?
(e.g. changing --very-sensitive to –very-fast)