Linux Commands for RNAseq Analysis. (Done in WSL1)

Installed WSL1, fastQC, Trimmomatic, hisat2, samtools and subread as Feature counts is available through subreads.

A fastq file was provided for demo. Pasted it in the WSL1 folder.

$ fastqc demo.fastq

To view the file

$ wslview demo\_fastqc.html

For trimming the low quality regions and adapter reads

$ java -jar /home/yourusername /Trimmomatic-0.39/trimmomatic-0.39.jar SE -phred33 -threads 4 demo.fastq trimmed-demo.fq LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:25

quality control of the trimmed file

$ fastqc trimmed-demo.fq

To view the file

$ wslview trimmed-demo\_fastqc.html

Alignment – Go to UCSC Browser> Downlaods> genome data> (+)human genomes> take any updated file > genome seq files and select annotations > Std genome sequence files > scroll down > A list of genome will be there hg38.fa.gz can be chosen for complete genome, in this case Chromosome 1 was downloaded for the remaining analysis

$ wget \_\_\_\_\_\_paste the link for specific chromosome\_\_\_\_\_\_\_

$ gunzip chr1.fa.gz

Building Index of reference genome (in this case chromosome1)

$ hisat2-build chr1.fa genome\_index

Aligning of reads (trimmed) to indexed genome

$ hisat2 -x genome\_index -U trimmed-demo.fq -S aligned-demo.sam

$ samtools view -bS aligned-demo.sam > aligned-demo.bam

Sorting the BAM file **by genomic coordinates** (i.e., the order of positions along the genome) and writes the sorted output to a new file.

$ samtools sort aligned-demo.bam -o sorted\_demo.bam

Gene Annotation : **Assigning aligned reads to genomic features** (like genes or exons) based on a reference annotation file (like chr1.gtf), so that we can **count how many reads map to each gene.**

$ featureCounts -a chr1.gtf -o counts.txt -t exon -g gene\_id sorted-demo.bam