**RNASeq Analysis**

**Download and install SRA toolkit**

**Download SRA accessions**

*prefetch SRR15838464 SRR15838465 SRR15838468 SRR15838469*

***Convert downloaded SRA files to FASTQ format***

*fastq-dump --gzip --split-files SRR15838464 SRR15838465 SRR15838468 SRR15838469*

**create a directory and save results of fastqc (QC)**

*mkdir /mnt/f/sra/fastqc\_results*

*fastqc /mnt/f/sra/\*.fastq.gz -o /mnt/f/sra/fastqc\_results*

**Trimming adapters and low-quality bases with Trimmomatic:**

**Write the shell script**

*chmod +x* ***trimming\_script.sh***

***Run the shell script trimming\_script.sh***

*./trimming\_script.sh SRR15838464\_1.fastq SRR15838464\_2.fastq SRR15838465\_1.fastq SRR15838465\_2.fastq SRR15838468\_1.fastq SRR15838468\_2.fastq SRR15838469\_1.fastq SRR15838469\_2.fastq*

**Build an index for the IRGSP-1.0 reference genome using HISAT2**

*hisat2-build IRGSP-1.0\_genome.fasta IRGSP-1.0\_index*

**Unzip the fastq fies for HISAT 2 alignment**

*gunzip \*.fastq.gz* ***or***

*gunzip SRR15838465\_1.fastq.gz*

*gunzip SRR15838465\_2.fastq.gz*

*gunzip SRR15838468\_1.fastq.gz*

*gunzip SRR15838468\_2.fastq.gz*

*gunzip SRR15838469\_1.fastq.gz*

*gunzip SRR15838469\_2.fastq.gz*

**Align Paired-end Reads to the Reference Genome using HISAT2:**

*hisat2 -x IRGSP-1.0\_index -1 SRR15838464\_1.fastq -2 SRR15838464\_2.fastq -S aligned\_reads\_SRR15838464.sam*

*Repeat this command for each pair of FASTQ files, replacing the filenames and output SAM file names accordingly.*

***Convert SAM to BAM and Sort:***

*samtools view -bS aligned\_reads\_SRR15838464.sam | samtools sort -o aligned\_reads\_sorted\_SRR15838464.bam*

*Repeat this command for each SAM file, replacing the filenames and output BAM file names accordingly.*

***Index the Sorted BAM files:***

*samtools index aligned\_reads\_sorted\_SRR15838464.bam*

*Repeat this command for each sorted BAM file.*

**featureCounts to quantify gene expression from the sorted BAM files.**

*featureCounts -p -a IRGSP-1.0\_representative\_transcript\_exon\_2024-01-11.gtf -o counts\_data.txt -T 4 aligned\_reads\_sorted\_SRR15838464.bam aligned\_reads\_sorted\_SRR15838465.bam aligned\_reads\_sorted\_SRR15838468.bam aligned\_reads\_sorted\_SRR15838469.bam*

***Differential Expression Analysis:***

*Use a tool like DESeq2 or edgeR for differential expression analysis. You'll need to load the gene count data into R and perform differential expression analysis. Here's a basic example using DESeq2 in R:*

*library(DESeq2)*

*# Load count data*

*countData <- as.matrix(read.table("gene\_counts\_SRR15838464.txt", header = TRUE, row.names = 1))*

*colData <- read.table("sample\_metadata.txt", header = TRUE, row.names = 1)*

*# Create DESeqDataSet object*

*dds <- DESeqDataSetFromMatrix(countData = countData, colData = colData, design = ~ condition)*

*# Perform differential expression analysis*

*dds <- DESeq(dds)*

*res <- results(dds)*

*Repeat this analysis for each of your samples.*