**Readme**

**Description**

This workflow is designed for processing, aligning, and analyzing RNA-Seq data to identify differentially expressed genes and perform GO term enrichment analysis.

**Dataset** - Gene expression data from human respiratory cells under mock control and SARS-CoV-2 infection at two time points (24 and 72 hours).

**Tools and Modules Used**

SRA Toolkit: For downloading sequence data.

FastQC: For quality control checks on raw sequence data

Cutadapt: For trimming adapters and primers from sequence data

Bowtie2: For aligning sequences to the reference genome.

Samtools: For converting SAM to BAM format and sorting.

DESeq2: For differential gene expression analysis

clusterProfiler: For GO term enrichment analysis.

**Adapter Sequences**

Illumina Adapter: AGATCGGAAGAG

SOLiD Adapter: CGCCTTGGCCGT

Illumina Primer: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

**Trimming Process**

FASTQ files were trimmed to remove adapter sequences using Cutadapt with a minimum length cutoff of 1.

**Differential Expression Analysis (DESeq2)**

DESeq2 was employed to identify differentially expressed genes between conditions and time points.

**GO Term Enrichment Analysis**

GO term enrichment was performed on the sets of differentially expressed genes to identify biological processes significantly associated with gene expression changes.

**Scripts and Analysis**

The scripts directory contains all the necessary scripts for the analysis. It includes data preprocessing, alignment, differential expression analysis, and GO term enrichment.

**Execution Order:**

Quality control with FastQC

Trimming with Cutadapt

Alignment with Bowtie2

Conversion and sorting with Samtools

Differential expression analysis with DESeq2

GO term enrichment with clusterProfiler

**Data output files**

1. Quality\_mapping\_scores.xlsx
2. DEG\_condition.csv
3. DEG\_time.csv
4. GO\_condition.csv
5. GO\_time.csv
6. Plots related to GO annotation( barplot\_condition.pdf, barplot\_time.pdf, ego\_conditon.pdf, ego\_time.pdf)