RNA-Seq Analysis Pipeline for SRA Accessions

This document outlines the complete RNA-Seq analysis pipeline used for processing SRA accessions. Each section details the tool, its purpose, and the exact commands (with comments for clarity and reproducibility).

Table of Contents

1. Data Acquisition and Conversion

Tools:

- SRA Toolkit (`prefetch`)
- HISAT2 Genome Index
- Ensembl GTF File

Purpose:

Download RNA-Seq data, reference genome, and GTF file for downstream analysis.

```bash
#!/bin/bash

SECONDS=0

cd /mnt/d/NGS/RNA_Seq

# 1. Download RNA-seq data from SRA
# SRR Accession numbers for RNA-seq data (controlled = SRR11412217, infected = SRR11412227)
# (healthy = SRR11517718, infected = SRR11517728)

prefetch SRR11412217 --progress # controlled
prefetch SRR11412227 --progress # infected
prefetch SRR11412218 --progress # controlled
prefetch SRR11412228 --progress # infected
echo "Downloaded RNA-seq data from SRA"

# 2. Download reference genome
```

```
mkdir -p Reference_Genome
cd Reference_Genome
wget https://genome-idx.s3.amazonaws.com/hisat/grch38_genome.tar.gz
tar -xvzf grch38_genome.tar.gz
echo "Genome indices successfully downloaded"

# 3. Download GTF file for featureCounts
wget http://ftp.ensembl.org/pub/release-106/gtf/homo_sapiens/Homo_sapiens.GRCh38.106.gtf.gz
echo "GTF file successfully downloaded"
cd -

duration=$SECONDS
echo "$(($duration / 60)) minutes and $(($duration % 60)) seconds elapsed."
```