Instructions for Initial Setup and Installation of the Analysis Tools on AWS Linux Instance

Run the following commands in the terminal of your AWS Linux instance:

1. Initial setup

```
sudo adduser fawzi
sudo usermod -aG sudo fawzi
su - fawzi
mkdir sequencing
mkdir sequencing/tools
mkdir sequencing/projects
mkdir sequencing/projects/RNA-seq_FRDA
sudo chmod -R ugo+rwx /data
cd sequencing/tools
```

2. Installing SRA Toolkit

```
wget http://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
tar -xvzf ~/sequencing/tools/sratoolkit.current-ubuntu64.tar.gz -C ~/sequencing/tools/
```

Edit /home/fawzi/.bashrc to insert the followig line at the end of the file:

```
export PATH=$PATH:/home/fawzi/sequencing/tools/sratoolkit.2.9.6-1-ubuntu64/bin
source ~/.bashrc
```

3. SAMtools

```
wget https://github.com/samtools/samtools/releases/download/1.9/samtools-1.9.tar.bz2
tar -vxjf ~/sequencing/tools/samtools-1.9.tar.bz2 -C ~/sequencing/tools/
```

Install updates and required packages on Ubuntu 18.04 (for running make):

```
cd ~/sequencing/tools/samtools-1.9/
sudo apt-get update
sudo apt-get install gcc
sudo apt-get install make
sudo apt-get install libbz2-dev
sudo apt-get install zlib1g-dev
sudo apt-get install libncurses5-dev
sudo apt-get install libncursesw5-dev
sudo apt-get install liblzma-dev
make
```

Edit /home/fawzi/.bashrc to insert the followig line at the end of the file:

```
export PATH=$PATH:/home/fawzi/sequencing/tools/samtools-1.9
source ~/.bashrc
```

4. HISAT2

```
wget http://ccb.jhu.edu/software/hisat2/dl/hisat2-2.1.0-Linux_x86_64.zip -P ~/sequencing/tools/
unzip ~/sequencing/tools/hisat2-2.1.0-Linux_x86_64.zip -d ~/sequencing/tools
```

Edit /home/fawzi/.bashrc to insert the followig line at the end of the file:

```
export PATH=$PATH:/home/fawzi/sequencing/tools/hisat2-2.1.0
source ~/.bashrc
```

5. Installing UCSC Genome hg38 index

```
wget  ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/hg38.tar.gz -P /data
tar -xzvf /data/hg38.tar.gz -C /data
```

6. Installing gencode v28 hg38 gene annotations

```
cd /home/fawzi/sequencing/projects/RNA-seq_FRDA
wget ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_28/gencode.v28.annotation.gtf.
gunzip gencode.v28.annotation.gtf.gz
```

7. HTseq

```
pip install HTSeq
```

8. Generating gene symbols (names) of gencode IDs

```
sh scripts/gencode.v28.symbols.awk > gencode.v28.symbols.txt
```

Instructions for the Execution of Analysis Tools

Run the following commands sequentially:

9. Download the reads of each sample from SRA using SRA Toolkit.

```
nohup sh scripts/download-sra-reads.sh > download-sra-reads.out &
```

10. Align (map) the reads of each sample to the Human Genome hg32 using HISAT2.

```
nohup sh scripts/hisat2.sh > hisat2.out &
```

11. Quantify the abundance of genes in each sample using HTseq.

```
nohup sxh scripts/htseq-count.sh > htseq-count.out &
```

12. Analysis using R:

- Merging the Count Files:
  Run the script merge_counts.R
- Exploratory Analysis:
  Run the script exploratory_analysis.R
- Differential Expression Analysis: Run the script differential_expression_analysis.R
- Epigenetics and Expression Analysis:
  Run the script epigenetics_expression.R