

```

# RNA-seq FASTQ download and quality control (HPC workflow)
#This script downloads RNA-seq data from NCBI SRA, converts it to FASTQ format, and performs
quality control using FastQC.
#Input data:
# SRR_Acc_List.txt contains 6 SRR accession numbers
# SRR accessions were obtained from the Methods section of the associated research article
and originate from the NCBI Sequence Read Archive (SRA)

# -----
# 1) List and load available versions of sra-tools to the environment
# -----
module avail sra-tools/3.0.3-gcc-13.2.0
module load sra-tools/3.0.3-gcc-13.2.0

# -----
# 2) Create output directories
# -----
#sra: stores downloaded .sra files from NCBI
#fastq: stores converted FASTQ files

mkdir sra
mkdir fastq

# -----
# 3) Download .sra files using prefetch from NCBI
# -----

prefetch --option-file SRR_Acc_List.txt --output-directory sra/

#--option-file SRR_Acc_List.txt: Reads the list of 6 SRR accession numbers (one per line)
#--output-directory sra: Saves downloaded files into the sra directory
# Expected output:
# One subdirectory per SRR accession inside sra
# Each containing a .sra file

# -----
# 4) Convert .sra files to FASTQ format
# -----

fasterq-dump sra/*/*.sra --split-files --threads 4 -O fastq/

#fasterq-dump is used instead of fastq-dump for improved speed
#sra/*/*.sra (path for the files)
#--split-files: Splits paired-end reads into separate files (_1.fastq and _2.fastq),
indicating paired-end RNA-seq data
#--threads 4: Uses 4 CPU threads for faster conversion
# -O fastq/ : Writes FASTQ output files to the 'fastq' directory
#Expected output for each SRR accession: SRRxxxxxxx_1.fastq and SRRxxxxxxx_2.fastq

# -----
# 5) Running quality check on the FASTQ files
# -----
# running fastqc from the directory where the FASTQ files are (if not specify the path to the
FASTQ files/direcotry)

# Load FastQC module
module load fastqc

#Change to FASTQ directory before running FastQC
cd fastq

fastqc *.fastq -o /scratch/grp/msc_appbio/Group2_ABCC/RNAseq/trial/QC

# *.fastq runs FastQC on all FASTQ files in the current directory.
# -O Specifies the output directory where FastQC results will be written.

```

```
# Expected output (per FASTQ file):  
#- HTML report (<sample_number_fastqc.html)  
#- Compressed results archive (<sample_number_fastqc.zip)
```