

# RNA-Seq utilities

Tools and utilities to supplement [PRAGUI](#). Mainly pre-processing data preparation.

## Getting started

### Prerequisites

The [cell\\_bio\\_util](#) repository will need to be installed next to the directory holding this project. In the case of MRC-LMB's Mario-Xeon machine, this has already been done.

For Mario-Xeon, RNA-Seq utilities is located in:

```
/data2/utilities/RNA-Seq_utilities/
```

## Running the scripts

### Removing ribosomal RNA from .fastq files

The **rRNA\_remover.py** script will achieve this. In the terminal, simply run:

```
python3 /data2/utilities/RNA-Seq_utilities/rRNA_remover.py
```

with the following arguments:

Flag	Description
<code>-h, --help</code>	Show this help message and exit
<code>-d, --directory &lt;directory&gt;</code>	Specify the location of the RNA-Seq data
<code>-l, --rRNA_library &lt;file&gt;</code>	Specify location of the rRNA genome library. i.e. path to the <b>.fa</b> file. Default is the <i>C. elegans</i> library.
<code>-s, --single_end</code>	Flag if RNA-Seq data are single end reads. Mutually exclusive with the <code>-p / --paired_end</code> argument.
<code>-p, --paired_end &lt;pair_tag&gt;</code>	Flag if RNA-Seq data are paired end reads. Mutually exclusive with the <code>-s / --single_end</code> argument. Provide space separated pair tags, this will be the same as PRAGUI's "pair_tags" argument (e.g. <code>r_1 r_2</code> ).

### Example

```
python3 /data2/utilities/RNA-Seq_utilities/rRNA_remover.py -d /scratch/gurpreet/data/ -l /scratch/ribosomal_rna/worm/c_elegans_concat_rDNA.fa -p r_1 r_2
```

## Merging RNA-Seq data into one file per sample, per lane

The **rna\_seq\_lane\_merger.py** script will achieve this. In the terminal, simply run:

```
python3 /data2/utilities/RNA-Seq_utilities/rna_seq_lane_merger.py
```

with the following arguments:

Flag	Description
-h, --help	Show this help message and exit.
-f, --submission_form <file>	Path to the submission form provided (e.g. CRUKCI_SLX_Submission.xlsx) - Please provide full path and ensure this file is in same folder as the RNA-Seq files.
-l, --lane_tags <lane_tag>	Tags (space separated) that identify samples' RNA-Seq lanes e.g. s_1 s_2.
-s, --single_end	Flag if RNA-Seq data are single end reads. Mutually exclusive with the -p / --paired_end argument.
-p, --paired_end <pair_tag>	Flag if RNA-Seq data are paired end reads. Mutually exclusive with the -s / --single_end argument. Provide space separated pair tags, this will be the same as PRAGUI's "pair_tags" argument.

### Example

```
python3 /data2/utilities/RNA-Seq_utilities/rna_seq_lane_merger.py -f  
/scratch/gurpreet/rna_seq_data/CRUKCI_SLX_Submission.xlsx -l s_1 s_2 -p r_1 r_2
```

# Calculating mean and standard deviation of the TPM values

---

The **tpm\_standard\_deviation\_mean\_calculator.py** script will achieve this. In the terminal, simply run:

```
python3 /data2/utilities/RNA-Seq_utilities/tpm_standard_deviation_mean_calculator.py
```

with the following arguments:

Flag	Description
-h, --help	Show this help message and exit.
-t, --tpm_file <file>	Full path for the tpm.txt file of interest (often uses "samples.csv_tpm.txt" filename - where "samples.csv" refers to PRAGUI's input csv file).

## Example

```
python3 /data2/utilities/RNA-Seq_utilities/tpm_standard_deviation_mean_calculator.py -t /scratch/gurpreet/rna_seq_data/samples.csv_tpm.txt
```

---

# Downloading your files from the CRUK FTP server

The **cruk\_downloader.py** script will achieve this. In the terminal, simply run:

```
python3 /data2/utilities/RNA-Seq_utilities/cruk_downloader.py
```

with the following arguments:

Flag	Description
<code>-h, --help</code>	Show this help message and exit.
<code>-f, --submission_form &lt;file&gt;</code>	Full path to the submission form (CRUK format) e.g. CRUKCI_SLX_Submission.xlsx. Please ensure this file is in same folder as where you wish to download the RNA-Seq files to.

This script reads in the CRUKCI\_SLX\_Submission.xlsx form and automatically retrieves the SLX ID and list of your files with which it will download to a directory of your choosing.

## Example

```
python3 /data2/utilities/RNA-Seq_utilities/ cruk_downloader.py -f  
/scratch/gurpreet/rna_seq_data/CRUKCI_SLX_Submission.xlsx
```