APPLIED SEQUENCING INFORMATICS-ASSIGNMENT-2

APOORVA SHARMA

Step1: Download the datasets
What sequencing methodologies were employed? Are these single- or paired-end datasets?

SRR1523657

Experiment

Experiment	Library Name	Platform	Strategy	Source	Selection	Layout	Action
SRX660429		Illumina	RNA-Seq	TRANSCRIPTOMIC	cDNA	PAIRED	BLAST

SRX4037309

Experiment

Experiment	Library Name	Platform	Strategy	Source	Selection	Layout	Action
SRX4037309		Illumina	RNA-Seq	TRANSCRIPTOMIC	cDNA	PAIRED	BLAST

SRX4146457

SRX4146457: GSM3167509: Dap5_rep1_heavy_polysome; Homo sapiens; RNA-Seq
2 ILLUMINA (Illumina HiSeq 2000) runs: 32.7M spots, 3.3G bases, 1Gb downloads

Submitted by: NCBI (GEO)

Study: A widespread alternate form of cap-dependent mRNA translation initiation
PRJNA473917 • SRP149421 • All experiments • All runs
show Abstract

Sample: Dap5_rep1_heavy_polysome
SAMN09288426 • SRS3360077 • All experiments • All runs
Organism: Homo sapiens

Library:
Instrument: Illumina HiSeq 2000
Strategy: RNA-Seq
Source: TRANSCRIPTOMIC
Selection: cDNA
Layout: PAIRED

RNA-Seq methodology was employed using Illumina and it's a **paired end** data set.

```
#!/bin/bash
#SBATCH --job-name=Exercise1 # Job name
#SBATCH --mail-type=END,FAIL # Mail events (NONE, BEGIN, END, FAIL, ALL)
#SBATCH --mail-user=firstname.lastname@nyulangone.org # Where to send mail
#SBATCH --ntasks=1 # Run on a single CPU
#SBATCH --mem=8gb # Job memory request
#SBATCH --time=12:00:00 # Time limit hrs:min:sec
#SBATCH --output=Exercise1_%j.log # Standard output and error log
#SBATCH -p cpu_short # Specifies location to submit job

## [ EXERCISE 1 ] ###
#Load sratoolkit module
module load sratoolkit/2.10.9

# Run fastq-dump to download specified file from the SRA, split into R1 & R2
file (if appropriate), and compress with gzip. Note --origfmt restores fastq
header sequences to original form
fastq-dump --split-files SRR1523671 --gzip -O
/gpfs/scratch/kerberosid/Practicum2/ --origfmt


# remove temporary directory that is utilised by fastq-dump
rm -r ~/ncbi

module avail # find the Trim Galore module

module load trimgalore/0.5.0  # Load trimgalore module

module load python/cpu/2.7.15-ES

trim_galore -o /gpfs/scratch/as18818id/Practicum2/ --paired
/gpfs/scratch/as18818id/Practicum2/SRR1523671_1.fastq.gz
/gpfs/scratch/as18818id/Practicum2/SRR1523671_2.fastq.gz --fastqc
--fastqc_args "-o /gpfs/scratch/as18818id/Practicum2/" --q 30 --gzip



java -jar trimmomatic-0.39.jar PE input_forward.fq.gz input_reverse.fq.gz
output_forward_paired.fq.gz output_forward_unpaired.fq.gz
```

```
output_reverse_paired.fq.gz output_reverse_unpaired.fq.gz
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:True LEADING:3 TRAILING:3 MINLEN:36
```