RNA-SEQ ANALYSIS REPORT

Aminata NDIAYE

1. Download 4 sequence datasets deposited to the EBI ENA:

Sequence datasets were taken into run section by typing in the search bar (http://www.ebi.ac.uk/ena/data/view) sequence IDS.

2. Extract fastq files

Fastq files have been extracted through their ftp addresses found in TXT results. After connecting with putty on a remote server, I used the following commands to download the fastq files into a folder.

ERR990557s.fastq

- wget ftp.sra.ebi.ac.uk/vol1/fastq/ERR990/ERR990557/ERR990557.fastq.gz

ERR990558s.fastq

- wget ftp.sra.ebi.ac.uk/vol1/fastq/ERR990/ERR990558/ERR990558.fastq.gz

ERR990559s.fastq

- wget ftp.sra.ebi.ac.uk/vol1/fastq/ERR990/ERR990559/ERR990559.fastq.gz

ERR990560s.fastq

- wget ftp.sra.ebi.ac.uk/vol1/fastq/ERR990/ERR990560/ERR990560.fastq.gz

3. For each file, select 8,000,000 (8 millions) of sequence reads and generate the following sample files:

To select only 8 millions of sequence reads I firstly use gzip to dezipp the original fastq files. Then I used the seqtk tools to sample the amount of sequence reads wanted in my fastq file. These are the following commands used (examples for ERR990557)

- Dezipping zipped files:

gzip -d ERR990557.fastq.gz

- Sampling 8 millions of sequence reads:

sqtk sample ERR990557.fastq 8000000 > ERR990557.selected.fastq

4. Align these read datasets to the reference genome by any appropriate mean, and generate a sorted bam alignment file.

I first downloaded the Drosophila reference genome through the flybase database. Subsequently I used bwa mem for the alignment of reads. For their ordination I used the tool samtoolssort (ordination by coordinates) which also allowed me to convert the sam files into bam.

5. Count reads aligning to genome's genes by any appropriate mean

To count reads aligned to each genome gene, I used the samtools tool. Through the following commands, the numbers of the mapped reads found are as follows.

- samtools view -F 0x904 -c ERR990557.bam

  3349122

- samtools view -F 0x904 -c ERR990558.bam

  3639848

- samtools view -F 0x904 -c ERR990559.bam

  2999175

- samtools view -F 0x904 -c ERR990560.bam

  3465490