

Project

Name: Aris Podotas

University: National and Kapodistrian University of Athens

Program: Data Science and Informaion Technologies

Specialization: Bioinformatics - Biomedical Data

Lesson: Introduction To Bioinformatics

Date: January 2025

Paper: Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis

Link: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3477283/>

Contents



1 Preface

The following analysis was done a Linux computer and all the paths will be of the relative file scheme on that computer. This was not done on the virtual machine for there is a desire to apply this analysis on seperate problems in the future and a installation that would be able to do so locally is preferred.

Along with this report file there will be a

file for the script that generate the results,

file that has the log of the outputs, all the files used for the analysis (such as the .sra files).

2 Description of the data

2.1 Getting the data

```
1 #!/bin/bash
2
3 # Chip seq pipleine
4
5 # Installing the sra toolkit
6 wget --output-document sratoolkit.tar.gz https://ftp-trace.ncbi.nlm.nih
   .gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
7 tar -vxzf sratoolkit.tar.gz
8
9 # Downloading files
10 prefetch SRR452931
11 mv ./SRR452931/SRR452931.sra ./
12 rm ./SRR452931/
13 fastq-dump -v --split-3 SRR452931
14 prefetch SRR524934
15 mv ./SRR524934/SRR524934.sra ./
16 rm ./SRR524934/
17 fastq-dump -v --split-3 SRR524934
18 prefetch SRR524936
19 mv ./SRR524936/SRR524936.sra ./
20 rm ./SRR524936/
21 fastq-dump -v --split-3 SRR524936
22 prefetch SRR524939
23 mv ./SRR524939/SRR524939.sra ./
24 rm ./SRR524939/
25 fastq-dump -v --split-3 SRR524939
26
27 # Getting the genome
28 wget -a https://hgdownload.cse.ucsc.edu/goldenpath/hg18/bigZips/hg18.
   chrom.sizes
29 wget -a https://hgdownload.cse.ucsc.edu/goldenpath/hg18/bigZips/hg18.fa
   .gz
30 gzip -d hg18.fa.gz
31
32 bowtie-build hg18.fa hg18
```



3 Protocol

4 Statistics

5 References

6 Results