

#PART 4 OF THE HANDS ON, TASK 1:

```
#
#
docker start -ai ae19e0cf92bc
PS1="ae19e0cf92bc:$ "
ae19e0cf92bc:$ ls -l
total 484
drwxr-xr-x 5 root root 4096 Jun 4 15:27 analyses
drwxr-xr-x 2 root root 4096 Jun 4 15:28 annotation
drwxrwxr-x 5 1000 1000 4096 Jun 4 11:02 chip-nf
-rw-rw-r-- 1 1000 1000 59 Jun 4 08:58 cmd.txt
drwxrwxr-x 7 1000 1000 4096 Jun 4 15:27 data
-rw-r--r-- 1 root root 474513 Jun 4 09:18 metadata.tsv
ae19e0cf92bc:$ ls -l | awk '/^d/ {print $9}'
analyses
annotation
chip-nf
data
ae19e0cf92bc:$ cd ../analyses/
ae19e0cf92bc:$ ls -l | awk '/^d/ {print $9}' | xargs -n 1 -I {} mkdir ../../ATAC-seq/analyses/{}
ae19e0cf92bc:$ cd ../../ATAC-seq/
ae19e0cf92bc:$ ls
aggregation.plot analyses annotation chip-nf cmd.txt data peaks.analysis scatterplot.correlation
ae19e0cf92bc:$ rmdir chip-nf/ peaks.analysis/ aggregation.plot/ scatterplot.correlation/
ae19e0cf92bc:$ ls analyses/
aggregation.plot peaks.analysis scatterplot.correlation
ae19e0cf92bc:$ ls data
bed.files bigBed.files bigWig.files fastq.files tsv.files
ae19e0cf92bc:$ cd ..
ae19e0cf92bc:$ ls
ATAC-seq ChIP-seq bin docker files.txt handsOn_images install.dependencies.txt test
#
#
#
```

#PART 4 TASK 2 (FILE ALREADY DOWNLOADED):

```
ae19e0cf92bc:$ head ./ATAC-seq/files.txt
"https://www.encodeproject.org/metadata/?-
replicates.library.biosample.donor.uuid=d370683e-81e7-473f-8475-7716d027849b&status=released&status=submitted&status=in+progress&assay_title=ATAC-
seq&biosample_ontology.term_name=sigmoid+colon&biosample_ontology.term_name=stomach&type=Experiment"
https://www.encodeproject.org/files/ENCFF332SCG/@download/ENCFF332SCG.fastq.gz
https://www.encodeproject.org/files/ENCFF591BAY/@download/ENCFF591BAY.fastq.gz
https://www.encodeproject.org/files/ENCFF591QAG/@download/ENCFF591QAG.bigWig
https://www.encodeproject.org/files/ENCFF778VJL/@download/ENCFF778VJL.bam
https://www.encodeproject.org/files/ENCFF997HH0/@download/ENCFF997HH0.bigWig
https://www.encodeproject.org/files/ENCFF049WJI/@download/ENCFF049WJI.bigWig
https://www.encodeproject.org/files/ENCFF189YSM/@download/ENCFF189YSM.bed.gz
https://www.encodeproject.org/files/ENCFF697MUB/@download/ENCFF697MUB.bed.gz
https://www.encodeproject.org/files/ENCFF628ARL/@download/ENCFF628ARL.bigWig
ae19e0cf92bc:$ ATAC_metadata=$(awk 'NR==1 {gsub("/", "");print}' ./ATAC-seq/files.txt)
ae19e0cf92bc:$ echo $ATAC_metadata
https://www.encodeproject.org/metadata/?
```

```
replicates.library.biosample.donor.uuid=d370683e-81e7-473f-8475-7716d027849b&status=released&status=submitted&status=in+progress&assay_title=ATAC-seq&biosample_ontology.term_name=sigmoid+colon&biosample_ontology.term_name=stomach&type=Experiment
ae19e0cf92bc:$ ./bin/download.metadata.sh $ATAC_metadata
The name is too long, 256 chars total.
Trying to shorten...
New name is index.html?-
replicates.library.biosample.donor.uuid=d370683e-81e7-473f-8475-7716d027849b&status=released&status=submitted&status=in+progress&assay_title=ATAC-seq&biosample_ontology.term_name=sigmoid+colon&biosample_ontology.term_name=sto.
--2024-06-04 17:15:14-- https://www.encodeproject.org/metadata/?-
replicates.library.biosample.donor.uuid=d370683e-81e7-473f-8475-7716d027849b&status=released&status=submitted&status=in+progress&assay_title=ATAC-seq&biosample_ontology.term_name=sigmoid+colon&biosample_ontology.term_name=stomach&type=Experiment
Resolving www.encodeproject.org (www.encodeproject.org)... 34.211.244.144
Connecting to www.encodeproject.org (www.encodeproject.org)|34.211.244.144|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/tsv]
Saving to: 'index.html?-
replicates.library.biosample.donor.uuid=d370683e-81e7-473f-8475-7716d027849b&status=released&status=submitted&status=in+progress&assay_title=ATAC-seq&biosample_ontology.term_name=sigmoid+colon&biosample_ontology.term_name=sto'
```

```
index.html?replicates.library.biosamp      [ <=> ] 24.00K --.-KB/s in 0.001s
```

```
2024-06-04 17:15:16 (28.5 MB/s) - 'index.html?-
replicates.library.biosample.donor.uuid=d370683e-81e7-473f-8475-7716d027849b&status=released&status=submitted&status=in+progress&assay_title=ATAC-seq&biosample_ontology.term_name=sigmoid+colon&biosample_ontology.term_name=sto' saved [24579]
```

```
ae19e0cf92bc:$ ls
ATAC-seq ChIP-seq bin docker files.txt handsOn_images install.dependencies.txt metadata.tsv test
ae19e0cf92bc:$ mv metadata.tsv ./ATAC-seq/
ae19e0cf92bc:$ cd ATAC-seq/
ae19e0cf92bc:$ head metadata.tsv
File_accession File_format File_type File_format_type Output_type File_assembly Experiment_accession Assay Donor(s) Biosample_term_id
Biosample_term_name Biosample_type Biosample_organism Biosample_treatments Biosample_treatments_amount Biosample_treatments_duration
Biosample_genetic_modifications_methods Biosample_genetic_modifications_categories
Biosample_genetic_modifications_targetsBiosample_genetic_modifications_gene_targets Biosample_genetic_modifications_site_coordinates
Biosample_genetic_modifications_zygosity Experiment_target Library_made_from Library_depleted_in Library_extraction_method Library_lysis_method
Library_crosslinking_method Library_strand specific Experiment_date_released Project RBNS_protein_concentration Library_fragmentation_method
Library_size_range Biological_replicate(s) Technical_replicate(s) Read_length Mapped_read_length Run_type Paired_end Paired_with Index_of
Derived_from Size Lab md5sum dbxrefs File_download_URL Genome_annotation Platform Controlled_by File_Status s3_uri Azure_URL
File_analysis_title File_analysis_status Audit_WARNING Audit_NOT_COMPLIANT Audit_ERROR
ENCFF332SCG fastq fastq reads ENCSR0860GH ATAC-seq /human-donors/ENCDO451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo sapiens DNA 2016-06-15 ENCODE 150-250 1 1 1 100 paired-ended 2 /files/-
ENCFF591BAY/ 4829455575 Michael_Snyder,_Stanford 46c728282f436846dda2e6fd06a05d0f SRA:SRR14305482 https://www.encodeproject.org/files/-
ENCFF332SCG/@download/ENCFF332SCG.fastq.gz Illumina_HiSeq_4000 released s3://encode-public/2016/04/26/9d9bcafc-8027-44b0-bc78-7d680e1af4fd/-
ENCFF332SCG.fastq.gz https://datasetencode.blob.core.windows.net/dataset/2016/04/26/9d9bcafc-8027-44b0-bc78-7d680e1af4fd/ENCFF332SCG.fastq.gz?-
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D
moderate_library_complexity,_moderate_number_of_reproducible_peaks,_mild_to_moderate_bottlenecking low FRiP_score
ENCFF591BAY fastq fastq reads ENCSR0860GH ATAC-seq /human-donors/ENCDO451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo sapiens DNA 2016-06-15 ENCODE 150-250 1 1 1 100 paired-ended 1 /files/-
ENCFF332SCG/ 4677656905 Michael_Snyder,_Stanford 615445218c4d845e93fea578a7230422 SRA:SRR14305482 https://www.encodeproject.org/files/-
ENCFF591BAY/@download/ENCFF591BAY.fastq.gz Illumina_HiSeq_4000 released s3://encode-public/2016/04/26/ec15f652-2fde-4deb-b54d-6b4ce5e6a7db/-
ENCFF591BAY.fastq.gz https://datasetencode.blob.core.windows.net/dataset/2016/04/26/ec15f652-2fde-4deb-b54d-6b4ce5e6a7db/ENCFF591BAY.fastq.gz?
```

sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF591QAG bigWig bigWig signal_p-value GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 /files/-
ENCF778VJL/ 1229149264 Anshul_Kundaje,Stanford 4a83f2ec8ed07194aae16fb2125d50b3 https://www.encodeproject.org/files/ENCF591QAG/@download/-
ENCF591QAG.bigWig released s3://encode-public/2016/12/22/a9441946-951f-41f9-a235-8222a5d73587/ENCF591QAG.bigWig https://-
datasetencode.blob.core.windows.net/dataset/2016/12/22/a9441946-951f-41f9-a235-8222a5d73587/ENCF591QAG.bigWig?
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D Lab_custom_GRCh38 archived
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF778VJL bam bam alignments GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 101 /files/-
ENCF591BAY/, /files/ENCF332SCG/ 6014216214 Anshul_Kundaje,Stanford 4f04671a6fc6aaf7a1f52ce381e0bbc6 https://www.encodeproject.org/files/-
ENCF778VJL/@download/ENCF778VJL.bam released s3://encode-public/2016/12/22/d18a1e9c-ef63-4d63-9a7c-cb8836d02f39/ENCF778VJL.bam https://-
datasetencode.blob.core.windows.net/dataset/2016/12/22/d18a1e9c-ef63-4d63-9a7c-cb8836d02f39/ENCF778VJL.bam?
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D Lab_custom_GRCh38 archived
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF997HHO bigWig bigWig fold_change_over_control GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 /files/ENCF778VJL/
1396147459 Anshul_Kundaje,Stanford 689b9a5828c53a594c75f6534a324a6c https://www.encodeproject.org/files/ENCF997HHO/@download/ENCF997HHO.bigWig
released s3://encode-public/2016/12/22/2d0c959d-4abc-49e1-b7a8-739bdd97aad/ENCF997HHO.bigWig https://datasetencode.blob.core.windows.net/dataset/-
2016/12/22/2d0c959d-4abc-49e1-b7a8-739bdd97aad/ENCF997HHO.bigWig?sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D
Lab_custom_GRCh38 archived moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF049WJI bigWig bigWig fold_change_over_control GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 /files/ENCF182ZGK/
1442938137 ENCODE_Processing_Pipeline c179a0d9c939eefc012267902905f65d https://www.encodeproject.org/files/ENCF049WJI/@download/-
ENCF049WJI.bigWig released s3://encode-public/2021/02/25/d4b6163d-f6c3-41ed-bf90-cd2270617924/ENCF049WJI.bigWig https://-
datasetencode.blob.core.windows.net/dataset/2021/02/25/d4b6163d-f6c3-41ed-bf90-cd2270617924/ENCF049WJI.bigWig?
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D ENCODE4_v1.9.1_GRCh38 released
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF189YSM bed_idr_ranked_peak bed_idr_ranked_peak IDR_ranked_peaks GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159
sigmoid_colon tissue Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 /files/-
ENCF182ZGK/ 5824121 ENCODE_Processing_Pipeline 5e5b640fd3c54d6fc9504d103f1532b7 https://www.encodeproject.org/files/ENCF189YSM/@download/-
ENCF189YSM.bed.gz released s3://encode-public/2021/02/25/575596e6-f176-432d-96ba-81fda37b7ec3/ENCF189YSM.bed.gz https://-
datasetencode.blob.core.windows.net/dataset/2021/02/25/575596e6-f176-432d-96ba-81fda37b7ec3/ENCF189YSM.bed.gz?
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D ENCODE4_v1.9.1_GRCh38 released
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF697MUB bed_narrowPeak bed_narrowPeak IDR_thresholded_peaks GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159 sigmoid_colon
tissue Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 /files/-
ENCF356LFX/, /files/ENCF182ZGK/ 1381456 ENCODE_Processing_Pipeline 685cfcb082f5f5b2f79b59b768af6f59 https://www.encodeproject.org/files/ENCF697MUB/
/@download/ENCF697MUB.bed.gz released s3://encode-public/2021/02/25/e1e14e86-f3f6-4ff8-b189-092ffe36440f/ENCF697MUB.bed.gz https://-
datasetencode.blob.core.windows.net/dataset/2021/02/25/e1e14e86-f3f6-4ff8-b189-092ffe36440f/ENCF697MUB.bed.gz?
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D ENCODE4_v1.9.1_GRCh38 released
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
ENCF628ARL bigWig bigWig signal_p-value GRCh38 ENCSR0860GH ATAC-seq /human-donors/ENC0451RUA/ UBERON:0001159 sigmoid_colon tissue
Homo_sapiens DNA 2016-06-15 ENCODE 150-250 1 1_1 /files/-
ENCF182ZGK/ 1066372356 ENCODE_Processing_Pipeline ec51ecf40b45a9d87a60f322aff8b669 https://www.encodeproject.org/files/ENCF628ARL/@download/-
ENCF628ARL.bigWig released s3://encode-public/2021/02/25/28217bb8-30cb-4592-b735-f8294abf142d/ENCF628ARL.bigWig https://-
datasetencode.blob.core.windows.net/dataset/2021/02/25/28217bb8-30cb-4592-b735-f8294abf142d/ENCF628ARL.bigWig?
sv=2019-10-10&si=prod&sr=c&sig=9qSQZo4ggrCNpybBExU8SypuUZV33igI1lw0P7rB3c%3D ENCODE4_v1.9.1_GRCh38 released
moderate_library_complexity, moderate_number_of_reproducible_peaks, mild_to_moderate_bottlenecking low_FrIP_score
#

```
#
#
ae19e0cf92bc:$ grep -F ATAC-seq metadata.tsv | grep -F "bigBed_narrowPeak" | grep -F "pseudoreplicated_peaks" | grep -F "GRCh38" | awk 'BEGIN{FS=OFS="\t"}{print $1, $11}' | sort -k2,2 -k1,1r | sort -k2,2 -u > analyses/bigBed.peaks.ids.txt
#
#
#
ae19e0cf92bc:$ cut -f1 analyses/bigBed.peaks.ids.txt | \
> while read filename; do
>   wget -P data/bigBed.files "https://www.encodeproject.org/files/$filename/@download/$filename.bigBed"
> done
--2024-06-04 17:34:04-- https://www.encodeproject.org/files/ENCFF287UHP/@download/ENCFF287UHP.bigBed
Resolving www.encodeproject.org (www.encodeproject.org)... 34.211.244.144
Connecting to www.encodeproject.org (www.encodeproject.org)|34.211.244.144|:443... connected.
HTTP request sent, awaiting response... 307 Temporary Redirect
Location: https://encode-public.s3.amazonaws.com/2021/02/25/1b862318-d014-4eeb-9e2b-cf0aafdd7298/ENCFF287UHP.bigBed?response-content-
disposition=attachment%3B%20filename%3DENCFF287UHP.bigBed&AWSAccessKeyId=ASIATGZNGCNX7REBACWN&Signature=j45QhK5parJvD0z%2BJnc9yCyksbM%3D&x-amz-security-
token=IQoJb3JpZ2luX2VjECEaCXVzLXdlc3QtMiJIMEYCIQC34cXvPnFsGaHAwMraPAwxvIG%2Bh2yqRfJDNXpgFQGjkwIhA0h0IypcnfH6Hv6v7RCvg3CmrL47UJM2UjwFmI6EmDWKrwFCKr%2F%2F%2F%
[following]
--2024-06-04 17:34:05-- https://encode-public.s3.amazonaws.com/2021/02/25/1b862318-d014-4eeb-9e2b-cf0aafdd7298/ENCFF287UHP.bigBed?response-content-
disposition=attachment%3B%20filename%3DENCFF287UHP.bigBed&AWSAccessKeyId=ASIATGZNGCNX7REBACWN&Signature=j45QhK5parJvD0z%2BJnc9yCyksbM%3D&x-amz-security-
token=IQoJb3JpZ2luX2VjECEaCXVzLXdlc3QtMiJIMEYCIQC34cXvPnFsGaHAwMraPAwxvIG%2Bh2yqRfJDNXpgFQGjkwIhA0h0IypcnfH6Hv6v7RCvg3CmrL47UJM2UjwFmI6EmDWKrwFCKr%2F%2F%2F%
Resolving encode-public.s3.amazonaws.com (encode-public.s3.amazonaws.com)... 52.92.128.65, 52.92.161.17, 52.92.137.33, ...
Connecting to encode-public.s3.amazonaws.com (encode-public.s3.amazonaws.com)|52.92.128.65|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6569988 (6.3M) [binary/octet-stream]
Saving to: 'data/bigBed.files/ENCFF287UHP.bigBed'

ENCFF287UHP.bigBed          100%[=====>] 6.27M  958KB/s   in 7.9s

2024-06-04 17:34:15 (808 KB/s) - 'data/bigBed.files/ENCFF287UHP.bigBed' saved [6569988/6569988]

--2024-06-04 17:34:15-- https://www.encodeproject.org/files/ENCFF762IFP/@download/ENCFF762IFP.bigBed
Resolving www.encodeproject.org (www.encodeproject.org)... 34.211.244.144
Connecting to www.encodeproject.org (www.encodeproject.org)|34.211.244.144|:443... connected.
HTTP request sent, awaiting response... 307 Temporary Redirect
Location: https://encode-public.s3.amazonaws.com/2021/02/25/4c185617-5c6e-48b2-9816-011d22463561/ENCFF762IFP.bigBed?response-content-
disposition=attachment%3B%20filename%3DENCFF762IFP.bigBed&AWSAccessKeyId=ASIATGZNGCNX7REBACWN&Signature=TL7LBm1X3dQWi0zS1uE3KBPrZ34%3D&x-amz-security-
token=IQoJb3JpZ2luX2VjECEaCXVzLXdlc3QtMiJIMEYCIQC34cXvPnFsGaHAwMraPAwxvIG%2Bh2yqRfJDNXpgFQGjkwIhA0h0IypcnfH6Hv6v7RCvg3CmrL47UJM2UjwFmI6EmDWKrwFCKr%2F%2F%2F%
[following]
--2024-06-04 17:34:16-- https://encode-public.s3.amazonaws.com/2021/02/25/4c185617-5c6e-48b2-9816-011d22463561/ENCFF762IFP.bigBed?response-content-
disposition=attachment%3B%20filename%3DENCFF762IFP.bigBed&AWSAccessKeyId=ASIATGZNGCNX7REBACWN&Signature=TL7LBm1X3dQWi0zS1uE3KBPrZ34%3D&x-amz-security-
token=IQoJb3JpZ2luX2VjECEaCXVzLXdlc3QtMiJIMEYCIQC34cXvPnFsGaHAwMraPAwxvIG%2Bh2yqRfJDNXpgFQGjkwIhA0h0IypcnfH6Hv6v7RCvg3CmrL47UJM2UjwFmI6EmDWKrwFCKr%2F%2F%2F%
Resolving encode-public.s3.amazonaws.com (encode-public.s3.amazonaws.com)... 52.92.128.65, 52.92.161.17, 52.92.137.33, ...
Connecting to encode-public.s3.amazonaws.com (encode-public.s3.amazonaws.com)|52.92.128.65|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6390157 (6.1M) [binary/octet-stream]
Saving to: 'data/bigBed.files/ENCFF762IFP.bigBed'

ENCFF762IFP.bigBed          100%[=====>] 6.09M  976KB/s   in 7.3s
```

2024-06-04 17:34:25 (852 KB/s) - 'data/bigBed.files/ENCFF762IFP.bigBed' saved [6390157/6390157]

```
#
#
#
ae19e0cf92bc:$ ../bin/selectRows.sh <(cut -f1 analyses/bigBed.peaks.ids.txt) metadata.tsv | cut -f1,46 > data/bigBed.files/md5sum.txt
ae19e0cf92bc:$ cat data/bigBed.files/md5sum.txt | while read filename original_md5sum; do md5sum data/bigBed.files/"$filename".bigBed |\
> awk -v filename="$filename" -v original_md5sum="$original_md5sum" 'BEGIN{FS=" "; OFS="\t"}{print filename, original_md5sum, $1}'
> done > tmp
#
#
#
ae19e0cf92bc:$ mv tmp data/bigBed.files/md5sum.txt
ae19e0cf92bc:$ awk '$2!=$3' data/bigBed.files/md5sum.txt
#CONVERSION FROM BIGBED TO BED
ae19e0cf92bc:$ cut -f1 analyses/bigBed.peaks.ids.txt | while read filename; do bigBedToBed data/bigBed.files/"$filename".bigBed data/bed.files/"$filename".bed;
done
#
#
#
#PART 4: TASK 3.1, BEDTOOL INTERSECTION WITH PROMOTER REGIONS
ae19e0cf92bc:$ cut -f-2 analyses/bigBed.peaks.ids.txt |while read filename tissue; do echo "$tissue"; bedtools intersect -a data/bed.files/"$filename".bed
b ../ChIP-seq/annotation/genecode.v24.protein.coding.non.redundant.TSS.bed -u | awk '!seen[$1"\t"$2]++' | wc -l; done
sigmoid_colon
21500
stomach
22211
#
#
#
#YOU CAN ALSO USE SORT, AND IT IS PROBABLY MORE EFFICIENT THAN AWK, BUT IT IS A GOOD TRAINING USING AWK FOR TEXT PROCESSING,
#THIS PROGRAM IS ABLE TO PARSE FILES OF ENORMOUS SIZE, EVEN OF DIFFERENT GIGABYTES, AND THE SLICING IS EASY AND STRAIGHTFORWARD
#THE ARRAY "SEEN" IF CACHES ANOTHER COMBINATION OF $1"\t"$2 AVOID PRINTING, ASSOCIATE TO THAT COMBINATION A NEW ADDITIONAL NUMBER AND GOES ON,
#OTHERWISE IT PRINTS THE RECORD, SO $0 IN AWK TERMS. THE FIRST FIELD IS THE CHROMOSOME, THE SECOND FIELD IS THE STARTING POSITION
#
ae19e0cf92bc:$ cut -f-2 analyses/bigBed.peaks.ids.txt |while read filename tissue; do echo "$tissue"; bedtools intersect -a data/bed.files/"$filename".bed
b ../ChIP-seq/annotation/genecode.v24.protein.coding.non.redundant.TSS.bed -u | sort -u -k1,1 -k2,2 | head; done
sigmoid_colon
chr1 100037869 100038981 Peak_27878 321 . 4.54433 32.19514 29.45567 967
chr1 100132473 100133456 Peak_49540 146 . 3.49259 14.62506 12.14781 870
chr1 1001849 1002073 Peak_30448 285 . 3.71640 28.51370 25.81558 115
chr1 100249509 100250300 Peak_1270 1000 . 21.73774 323.83047 319.51874 405
chr1 100265547 100266806 Peak_101561 58 . 2.51046 5.81125 3.63623 267
chr1 10032367 10033447 Peak_1496 1000 . 18.96693 310.39404 306.16052 442
chr1 100351281 100353312 Peak_103210 56 . 2.03606 5.66158 3.49373 730
chr1 100538921 100539666 Peak_144377 44 . 2.42157 4.40977 2.36383 160
chr1 100894751 100895546 Peak_21817 444 . 5.59701 44.47607 41.62366 114
chr1 100895615 100896647 Peak_166588 39 . 1.90678 3.97255 1.98963 782
stomach
chr1 100037754 100038958 Peak_24494 264 . 4.72546 26.49679 23.68646 597
chr1 100132473 100133430 Peak_140009 45 . 2.30292 4.54862 2.46205 104
chr1 100249320 100250079 Peak_2465 1000 . 15.71092 166.10545 162.15907 590
```

```
chr1 100265547 100266861 Peak_10439 660 . 9.92556 66.08572 62.86540 631
chr1 10032375 10033532 Peak_1832 1000 . 16.02310 185.39249 181.30231 448
chr1 100351295 100353177 Peak_11364 610 . 6.04190 61.01293 57.83438 401
chr1 100539133 100540185 Peak_49481 120 . 3.80854 12.02153 9.52162 491
chr1 100894745 100895228 Peak_17554 386 . 5.25164 38.61178 35.64590 111
chr1 100895403 100896869 Peak_22227 294 . 4.52464 29.49813 26.64439 359
chr1 101025551 101026405 Peak_237330 30 . 2.00382 3.07693 1.25215 58
```

```
#
#
#
```

#PART 4: TASK 3.2, BEDTOOL INTERSECTION WITH WHOLE GENE BODY

```
ae19e0cf92bc:$ cut -f-2 analyses/bigBed.peaks.ids.txt | while read filename tissue; do echo "$tissue"; bedtools intersect -a data/bed.files/"$filename".bed -
b ../ChIP-seq/annotation/gencode.v24.protein.coding.gene.body.bed -v | awk '!seen[$1"\t"$2]++' | wc -l ; done
```

```
sigmoid_colon
```

```
25635
```

```
stomach
```

```
25665
```

```
ae19e0cf92bc:$ #THIS IS THE RIGHT COMMAND TO INCLUDE INTO THE REPORT, SECTION 4.3.2
```

```
ae19e0cf92bc:$ cut -f-2 analyses/bigBed.peaks.ids.txt | while read filename tissue; do echo "$tissue"; bedtools intersect -a data/bed.files/"$filename".bed -
b ../ChIP-seq/annotation/gencode.v24.protein.coding.gene.body.bed -v | awk '!seen[$1"\t"$2]++' | awk 'END{print NR}' ; done
```

```
sigmoid_colon
```

```
25635
```

```
stomach
```

```
25665
```

```
#
#
#
```

#PART 5: TASK 1, CREATING A REGULATORY_ELEMENTS FOLDER

```
docker start -ai ae19e0cf92bc
```

```
PS1="ae19e0cf92bc:$ "
```

```
ae19e0cf92bc:$ pwd
```

```
/home/pc/master/deadlines_assignments_papers_lectures/epigenomics/epigenomics_first_task_19_03_2024/epigenomics_uvic
```

```
ae19e0cf92bc:$ mkdir regulatory_elements
```

```
ae19e0cf92bc:$ ls
```

```
ATAC-seq bin files.txt install.dependencies.txt test
```

```
ChIP-seq docker hands0n_images regulatory_elements
```

```
#
#
#
```

#CREATION OF THE BED FILES, THAT WERE NOT INITIALLY CREATED DURING PART 4:

```
#
```

```
ae19e0cf92bc:$ cd ATAC-seq/
```

```
ae19e0cf92bc:$ ls
```

```
analyses annotation cmd.txt data files.txt metadata.tsv
```

```
ae19e0cf92bc:$ cut -f-2 analyses/bigBed.peaks.ids.txt | while read filename tissue; do echo "$tissue"; bedtools intersect -a data/bed.files/"$filename".bed -
b ../ChIP-seq/annotation/gencode.v24.protein.coding.gene.body.bed -v | awk '!seen[$1"\t"$2]++' | awk 'END{print NR}' ; done
```

```
sigmoid_colon
```

```
25635
```

```
stomach
```

```
25665
```

```
#JUST AS A RECAP
```

```

ae19e0cf92bc:$ cut -f-2 analyses/bigBed.peaks.ids.txt | while read filename; do bedtools intersect -a data/bed.files/"$filename".bed -b ../ChIP-seq/-
annotation/gencode.v24.protein.coding.gene.body.bed -v | awk '!seen[$1"\t"$2]++' > data/bed.files/ATAC_seq_peaks_outside_genes_"$tissue".bed ; done
#
#
#PART 5: TASK 2, OVERLAP PEAKS OF H3K27AC AND H3K4ME1 IN THE CORRESPONDING TISSUE
#SEARCHING THE FILES TO DOWNLOAD
ae19e0cf92bc:$ for mod in H3K27ac H3K4me1; do echo $mod; grep -F $mod ../ChIP-seq/metadata.tsv | grep -F "bigBed_narrowPeak" | grep -F "pseudoreplicated_peaks"
| grep -F "GRCh38" | awk 'BEGIN{FS=OFS="\t"}{print $1, $11}' | sort -k2,2 -k1,1r | awk '!seen[$2]++' ; done
H3K27ac
ENCF872UHN sigmoid_colon
ENCF977LBD stomach
H3K4me1
ENCF724ZOF sigmoid_colon
ENCF844XRN stomach
ae19e0cf92bc:$ ls
ae19e0cf92bc:$ mkdir analyses
ae19e0cf92bc:$ for mod in H3K27ac H3K4me1; do echo $mod; grep -F $mod ../ChIP-seq/metadata.tsv | grep -F "bigBed_narrowPeak" | grep -F "pseudoreplicated_peaks"
| grep -F "GRCh38" | awk 'BEGIN{FS=OFS="\t"}{print $1, $11}' | sort -k2,2 -k1,1r | awk '!seen[$2]++' > analyses/bigBed."$mod".peaks.ids.txt ; done
H3K27ac
H3K4me1
ae19e0cf92bc:$ ls analyses/
bigBed.H3K27ac.peaks.ids.txt bigBed.H3K4me1.peaks.ids.txt
ae19e0cf92bc:$ for mod in H3K27ac H3K4me1; do mkdir -p data/"$mod"; done
ae19e0cf92bc:$ ls data/
H3K27ac H3K4me1
#THIS FOR CYCLE WAS NOT NECESSARY, IT WOULD HAE BEEN BETTER TO USE JUST MKDIR
#
#
#DOWNLOADING THE BIGBED FILES
ae19e0cf92bc:$ for mod in H3K27ac H3K4me1; do
> cut -f1 analyses/bigBed."$mod".peaks.ids.txt | while read filename ;
> do
> echo $filename ; wget -P data/"$mod" https://www.encodeproject.org/files/"$filename"/@@download/"$filename".bigBed
> done
> done
ENCF872UHN
--2024-06-05 08:45:56-- https://www.encodeproject.org/files/ENCF872UHN/@download/ENCF872UHN.bigBed
Resolving www.encodeproject.org (www.encodeproject.org)... 34.211.244.144
Connecting to www.encodeproject.org (www.encodeproject.org)|34.211.244.144|:443... connected.
HTTP request sent, awaiting response... 307 Temporary Redirect
Location: https://encode-public.s3.amazonaws.com/2020/09/21/c38cd923-008e-4c39-a49f-feaf05263d93/ENCF872UHN.bigBed?response-content-
disposition=attachment%3B%20filename%3DENCF872UHN.bigBed&AWSAccessKeyId=ASIATGZNGCNX6TSVMQLY&Signature=Sf9qzBklwrEFxe3NQ6w%2Fg6fNIT0%3D&x-amz-security-
token=IQoJb3JpZ2luX2VjEDAAaCXVzLXdlc3Q0tMiJIMEYCIQckugIq2xlp9tzhANCxKu%2BCWa0J%2FZCTTMSU%2FS2GR6qigIhAInQtJLNGbBTicpusoj%2FPA9q0mF1gBMjIYuDT0SU0VaaKrwFCLn%2F%2F%
[following]
--2024-06-05 08:45:57-- https://encode-public.s3.amazonaws.com/2020/09/21/c38cd923-008e-4c39-a49f-feaf05263d93/ENCF872UHN.bigBed?response-content-
disposition=attachment%3B%20filename%3DENCF872UHN.bigBed&AWSAccessKeyId=ASIATGZNGCNX6TSVMQLY&Signature=Sf9qzBklwrEFxe3NQ6w%2Fg6fNIT0%3D&x-amz-security-
token=IQoJb3JpZ2luX2VjEDAAaCXVzLXdlc3Q0tMiJIMEYCIQckugIq2xlp9tzhANCxKu%2BCWa0J%2FZCTTMSU%2FS2GR6qigIhAInQtJLNGbBTicpusoj%2FPA9q0mF1gBMjIYuDT0SU0VaaKrwFCLn%2F%2F%
Resolving encode-public.s3.amazonaws.com (encode-public.s3.amazonaws.com)... 52.92.144.105, 52.92.209.185, 52.92.227.105, ...
Connecting to encode-public.s3.amazonaws.com (encode-public.s3.amazonaws.com)|52.92.144.105|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2980089 (2.8M) [binary/octet-stream]

```


Saving to: 'data/H3K27ac/ENCFF872UHN.bigBed'


```
ae19e0cf92bc:$ ls analyses/
```

```

bigBed.H3K27ac.peaks.ids.txt bigBed.H3K4me1.peaks.ids.txt
ae19e0cf92bc:$ for mod in H3K27ac H3K4me1 ;
> do echo "$mod" ; cut -f1 analyses//bigBed."$mod".peaks.ids.txt | while read filename ;
> do bigBedToBed data/"$mod"/"$filename".bigBed data/"$mod"/"$filename".bed
> done
> done
H3K27ac
H3K4me1
#
#
#
#INTERSECTING GENE PEAKS WITH H3K27AC AND H3K4ME1 PEAKS
#
ae19e0cf92bc:$ for mod in H3K27ac H3K4me1 ; do cut -f-2 analyses/bigBed."$mod".peaks.ids.txt | while read filename tissue ; do echo "$tissue $filename
$mod" ; bedtools intersect -a ../ATAC-seq/data/bed.files/ATAC_seq_peaks_outside_genes_"$tissue".bed -b data/"$mod"/"$filename".bed -u > data/-
common_peaks_"$mod"_"$tissue".bed; done; done
sigmoid_colon ENCF872UHN H3K27ac
stomach ENCF977LBD H3K27ac
sigmoid_colon ENCF724ZOF H3K4me1
stomach ENCF844XRN H3K4me1
ae19e0cf92bc:$ ls data/
H3K27ac H3K4me1 common_peaks_H3K27ac_sigmoid_colon.bed common_peaks_H3K27ac_stomach.bed common_peaks_H3K4me1_sigmoid_colon.bed
common_peaks_H3K4me1_stomach.bed
ae19e0cf92bc:$ head data/common_peaks_H3K27ac_sigmoid_colon.bed
chr1 778339 779193 Peak_175893 36 . 2.20588 3.66321 1.70835 781
chr1 817296 818240 Peak_23845 396 . 7.14286 39.65236 36.84049 208
chr1 818645 819307 Peak_57080 119 . 3.77937 11.91532 9.50266 404
chr1 826914 828166 Peak_33159 254 . 4.43365 25.44272 22.78304 377
chr1 904265 904921 Peak_22173 435 . 7.26132 43.57509 40.73002 498
chr1 923679 924127 Peak_21162 463 . 6.25197 46.38569 43.51812 148
chr1 1122088 1122426 Peak_20098 497 . 7.93796 49.74874 46.85592 212
chr1 1157380 1158638 Peak_233916 27 . 1.69724 2.73909 0.98229 1124
chr1 1208073 1208673 Peak_223436 30 . 1.91693 3.06306 1.24451 530
chr1 1362059 1362294 Peak_37551 214 . 3.96324 21.40214 18.79951 120
ae19e0cf92bc:$ head data/common_peaks_H3K27ac_stomach.bed
chr1 778500 779017 Peak_2029 1000 . 24.25669 178.99461 174.95253 280
chr1 817333 818325 Peak_148510 41 . 2.38230 4.16539 2.10929 505
chr1 826954 828100 Peak_163653 41 . 2.11480 4.11076 2.08755 154
chr1 1067682 1068269 Peak_25860 249 . 3.88170 24.92832 22.14262 271
chr1 1068516 1069461 Peak_19319 346 . 4.48689 34.61676 31.69713 267
chr1 1079493 1080378 Peak_24039 270 . 5.22928 27.02500 24.20646 525
chr1 1124797 1125018 Peak_32494 192 . 3.54911 19.25146 16.56763 106
chr1 1125097 1125536 Peak_106662 55 . 2.09450 5.51351 3.32675 326
chr1 1128531 1128793 Peak_39761 152 . 3.01140 15.27100 12.67715 120
chr1 1132532 1132876 Peak_35758 173 . 3.10234 17.32507 14.68308 203
ae19e0cf92bc:$ bedtools intersect -a data/common_peaks_H3K27ac_sigmoid_colon.bed -b data/common_peaks_H3K4me1_sigmoid_colon.bed -u |wc -l
8627
ae19e0cf92bc:$ bedtools intersect -a data/common_peaks_H3K4me1_sigmoid_colon.bed -b data/common_peaks_H3K27ac_sigmoid_colon.bed -u > data/-
common_peaks_H3K27ac_H3K4me1_sigmoid_colon.bed
ae19e0cf92bc:$ bedtools intersect -a data/common_peaks_H3K4me1_stomach.bed -b data/common_peaks_H3K27ac_stomach.bed -u > data/-
common_peaks_H3K27ac_H3K4me1_stomach.bed

```

```

ae19e0cf92bc:$ wc -l data/common_peaks_H3K27ac_H3K4me1_stomach.bed
5148 data/common_peaks_H3K27ac_H3K4me1_stomach.bed
#
#
#
#THESE ARE THE CANDIDATE DISTAL REGULATORY ELEMENTS FOR SIGMOID_COLON AND STOMACH RESPECTIVELY
#
#
#
#PART 5, TASK 3: GENERATE A FILE REGULATORY.ELEMENTS.STARTS.TSV THAT CONTAINS THE NAME OF THE REGULATORY REGION AND THE START (5') COORDINATE OF THE REGION
ae19e0cf92bc:$ ls
ATAC-seq ChIP-seq bin docker files.txt handsOn_images install.dependencies.txt regulatory_elements test
ae19e0cf92bc:$ cd regulatory_elements/
ae19e0cf92bc:$ mkdir regulatory_elements_starts
ae19e0cf92bc:$ for tissue in sigmoid_colon stomach ; do echo "$tissue" ; grep chr1 data/common_peaks_H3K27ac_H3K4me1_"$tissue".bed | awk 'BEGIN{OFS="\t"}{print $4,$2}' | head; done
sigmoid_colon
Peak_23845 817296
Peak_22173 904265
Peak_21162 923679
Peak_20098 1122088
Peak_233916 1157380
Peak_223436 1208073
Peak_37551 1362059
Peak_28508 1430253
Peak_59235 1433282
Peak_22309 1434331
stomach
Peak_25860 1067682
Peak_19319 1068516
Peak_24039 1079493
Peak_32494 1124797
Peak_106662 1125097
Peak_13453 1136406
Peak_18461 1144096
Peak_22657 1149943
Peak_13897 1157884
Peak_143416 1433207
ae19e0cf92bc:$ for tissue in sigmoid_colon stomach ; do echo "$tissue" ; grep chr1 data/common_peaks_H3K27ac_H3K4me1_"$tissue".bed | awk 'BEGIN{OFS="\t"}{print $4,$2}' > regulatory_elements_starts_"$tissue".tsv ; done
#
#
#
#NOTICE HOW THIS TIME THERE IS NO NEED TO IMPOSE FS="\t", BECAUSE THESE BED FILES WERE PROCESSED BEFORE WITH BEDTOOLS, WHICH ACCEPTS ONLY FILES WITH
#FIELDS SEPARATED EACH ONE BY A TAB INSTEAD OF A SPACE, THE GREP COMMAND CAN ALSO BE AVOIDED (SEE BELOW)
#
#
#
#PART 5, TASK 4: FILE CALLED GENE.STARTS.TSV WHICH WILL STORE THE NAME OF THE GENE IN THE FIRST COLUMN, AND THE START COORDINATE OF THE GENE ON THE SECOND COLUMN
ae19e0cf92bc:$ ls
analyses data regulatory_elements_starts regulatory_elements_starts.sigmoid_colon.tsv regulatory_elements_starts.stomach.tsv

```

```

ae19e0cf92bc:$ mv regulatory_elements_starts.*tsv ./regulatory_elements_starts
ae19e0cf92bc:$ ls
analyses  data  regulatory_elements_starts
ae19e0cf92bc:$ ls regulatory_elements_starts/
regulatory_elements_starts.sigmoid_colon.tsv  regulatory_elements_starts.stomach.tsv
ae19e0cf92bc:$ pwd
/home/pc/master/deadlines_assignments_papers_lectures/epigenomics/epigenomics_first_task_19_03_2024/epigenomics_uvic/regulatory_elements
ae19e0cf92bc:$ mkdir gene_starts
ae19e0cf92bc:$ awk 'BEGIN{FS=OFS="\t"}/chr1/{if ($6=="+") {start=$2} else {start=$3}; print $4, start}' ../ChIP-seq/annotation/-
genome.v24.protein.coding.gene.body.bed | head -10
ENSG00000186092.4    69090
ENSG00000279928.1    182392
ENSG00000279457.3    200322
ENSG00000278566.1    451678
ENSG00000273547.1    686654
ENSG00000187634.10   924879
ENSG00000188976.10   959309
ENSG00000187961.13   960586
ENSG00000187583.10   966496
ENSG00000187642.9    982093
ae19e0cf92bc:$ awk 'BEGIN{OFS="\t"}/chr1/{if ($6=="+") {start=$2} else {start=$3}; print $4, start}' ../ChIP-seq/annotation/-
genome.v24.protein.coding.gene.body.bed > gene_starts/gene.starts.tsv
#
#
#
#IT COULD BE A LITTLE RISKY TO PROCESS A FILE WITH SIMPLY /chr1/ WHICH IS IMPLICIT FORM FOR $0 ~ /chr1/ BUT WE HAVE ALREADY PROCESSED THIS BED FILE WITH BEDTOOL
#THAT PROBABLY ENSURES THAT THE FIRST FIELD ONLY CONTAINS STRINGS REGARDING CHROMOSOMES, A LESS RISKY APPROACH WOULD BE USING $1 ~ /chr1/
#BUT THE HEAD OUTPUT HAS BEEN ALREADY COMPARED WITH GREP IN THE PIPELINE (NOT SHOWN HERE)
#
#
#
#PART 5, TASK 5: COMPLETING THE PYTHON SCRIPT
ae19e0cf92bc:$ wget -P ../bin/ https://public-docs.crg.es/rguigo/Data/bborsari/UVIC/epigenomics_course/get.distance.py
--2024-06-05 14:16:21-- https://public-docs.crg.es/rguigo/Data/bborsari/UVIC/epigenomics_course/get.distance.py
Resolving public-docs.crg.es (public-docs.crg.es)... 84.88.66.216
Connecting to public-docs.crg.es (public-docs.crg.es)|84.88.66.216|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1149 (1.1K) [text/plain]
Saving to: '../bin/get.distance.py'

get.distance.py          100%[=====>]  1.12K  --.-KB/s    in
0s

2024-06-05 14:16:24 (15.0 MB/s) - '../bin/get.distance.py' saved [1149/1149]

#
#
#
#AFTER ADDING / MODIFYING THE FOLLOWING LINES IN THE SCRIPT:
#
#

```

```

#
nano ../bin/get.distance.py
#
#
#
for line in open_input.readlines():          # for each line in the input file
    gene, y = line.strip().split('\t')        # split the line into two columns based on a tab
    position = int(y)                         # define a variable called position that correspond to the integer of the start of the gene
    distance = abs(position - enhancer_start)  # compute the absolute value of the difference between position and enhancer_start

    if distance < x:                          # if this absolute value is lower than x
        x = distance                          # this value will now be your current x
        selectedGene = gene                  # save gene as selectedGene
        selectedGeneStart = position         # save position as selectedGeneStart

print "\t".join([selectedGene, str(selectedGeneStart), str(x)])
ae19e0cf92bc:$ python ../bin/get.distance.py --input ../gene_starts/gene.starts.tsv --start 980000
ENSG00000167693.16 979770 230
#
#
#
#THE RESULT IS NOT THE DESIRED ONE, WHICH WOULD BE THE FOLLOWING:
python ../bin/get.distance.py --input gene.starts.tsv --start 980000
ENSG00000187642.9 982093 2093
#
#
#
#PART 5, TASK 6: FOR EACH REGULATORY ELEMENT CONTAINED IN THE FILE REGULATORY.ELEMENTS.STARTS.TSV, RETRIEVE THE CLOSEST GENE AND THE DISTANCE TO THE CLOSEST
GENE #USING THE PYTHON SCRIPT
#
#
#
ae19e0cf92bc:$ ls
ATAC-seq ChIP-seq bin docker files.txt handsOn_images install.dependencies.txt regulatory_elements test
ae19e0cf92bc:$ cd regulatory_elements/
ae19e0cf92bc:$ ls
analyses data gene_starts regulatory_elements_starts
ae19e0cf92bc:$ mkdir regulatory_elements_distances
#
#
#
ae19e0cf92bc:$ for tissue in sigmoid_colon stomach;
> do
> cat regulatory_elements_starts/regulatory_elements_starts."$tissue".tsv | while read element start ;
> do
> python ../bin/get.distance.py --input gene_starts/gene.starts.tsv --start "$start" ;
> done > regulatory_elements_distances/regulatoryElements.genes.distances."$tissue".tsv
> done
#
#
#

```

```
#PART 5, TASK 7: USE R TO COMPUTE THE MEAN AND THE MEDIAN OF THE DISTANCES STORED IN REGULATORYELEMENTS.GENES.DISTANCES.TSV
```

```
#  
#  
#  
ae19e0cf92bc:$ ls  
analyses data gene_starts regulatory_elements_distances regulatory_elements_starts  
ae19e0cf92bc:$ head regulatory_elements_distances/regulatoryElements.genes.distances.stomach.tsv  
ENSG000000180448.10 1065922 1760  
ENSG000000180448.10 1065922 2594  
ENSG000000162009.8 1078780 713  
ENSG000000131591.17 1116361 8436  
ENSG000000131591.17 1116361 8736  
ENSG000000273554.4 1147868 11462  
ENSG000000273554.4 1147868 3772  
ENSG000000273554.4 1147868 2075  
ENSG000000215182.8 1157952 68  
ENSG000000179403.11 1434860 1653  
ae19e0cf92bc:$ head regulatory_elements_distances/regulatoryElements.genes.distances.sigmoid_colon.tsv  
ENSG000000177666.15 818901 1605  
ENSG000000141433.12 904942 677  
ENSG000000187634.10 924879 1200  
ENSG000000131591.17 1116361 5727  
ENSG000000278782.1 1157686 306  
ENSG000000186891.13 1206691 1382  
ENSG000000162576.16 1361777 282  
ENSG000000205116.3 1426127 4126  
ENSG000000179403.11 1434860 1578  
ENSG000000179403.11 1434860 529  
ae19e0cf92bc:$ R
```

R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

```
> stomach_data <- read.table("/home/pc/master/deadlines_assignments_papers_lectures/epigenomics/epigenomics_first_task_19_03_2024/epigenomics_uvic/-  
regulatory_elements/regulatory_elements_distances/regulatoryElements.genes.distances.stomach.tsv", header = FALSE, sep = "\t")  
> sigmoid_colon_data <- read.table("/home/pc/master/deadlines_assignments_papers_lectures/epigenomics/epigenomics_first_task_19_03_2024/epigenomics_uvic/-  
regulatory_elements/regulatory_elements_distances/regulatoryElements.genes.distances.sigmoid_colon.tsv", header = FALSE, sep = "\t")  
> stomach_mean <- mean(stomach_data$V3) ; stomach_median <- median(stomach_data$V3)  
> sigmoid_colon_mean <- mean(sigmoid_colon_data$V3) ; sigmoid_colon_median <- median(sigmoid_colon_data$V3)
```

```
> cat("Stomach Data - Mean:", stomach_mean, "Median:" , stomach_median, "\n")
Stomach Data - Mean: 13250.63 Median: 5531
> cat("Sigmoid Colon Data - Mean:", sigmoid_colon_mean, "Median:" , sigmoid_colon_median, "\n")
Sigmoid Colon Data - Mean: 16138.57 Median: 5854
>
Save workspace image? [y/n/c]: y
```

```
-----
END_OF_THE_EPIGENOMICS_TASK-----
```


