

Here is your homework !

Please download data first then follow the requirements step by step to complete homework.

Test Data

ChIP-seq data:

Treat:

```
/f/share/test_data/ChIP-seq/c1.clean_1.fastq.gz
```

```
/f/share/test_data/ChIP-seq/c1.clean_2.fastq.gz
```

Control:

```
/f/share/test_data/ChIP-seq/input_1.fastq.gz
```

```
/f/share/test_data/ChIP-seq/input_2.fastq.gz
```

Homework Requirements

Since you have process raw data, now you can start the downstream analysis of ChIP-seq! After processing, you need to learn and complete these steps including:

以下只是参考，可以使用网上在线分析网站或者自己搭建软件

1. Set up the environment

nf-core/chipseq(recommended) ~几分钟

```
###见https://github.com/nf-core/chipseq
```

2. input csv design

```
###见https://github.com/nf-core/chipseq/blob/master/docs/usage.md(Full design)
```

3. run nextflow pipeline ~16cpus差不多2小时

###见<https://github.com/nf-core/chipseq>

分批进行，设置cpus 4，建议错峰跑

Output like(successfully)

```
-[nf-core/chipseq] Pipeline completed successfully-  
WARN: To render the execution DAG in the required format it is  
for more info.  
Completed at: 13-Jul-2021 12:56:53  
Duration      : 1h 49m 51s  
CPU hours     : 23.2  
Succeeded     : 34
```

4. Quality control

ChIP-seq success? fail?(qc表格，可能失败原因，写入word)

Name	NRF	PBC1	PBC2	NSC	RSC	FRiP	IGV
------	-----	------	------	-----	-----	------	-----

Output like

Name	NRF	PBC1	PBC2	NSC	RSC	FRiP	IGV
27me3-JJ	0.774868	0.761995	3.932041	1.008663	0.3161772	0.0240527	failed

5. Describe the data

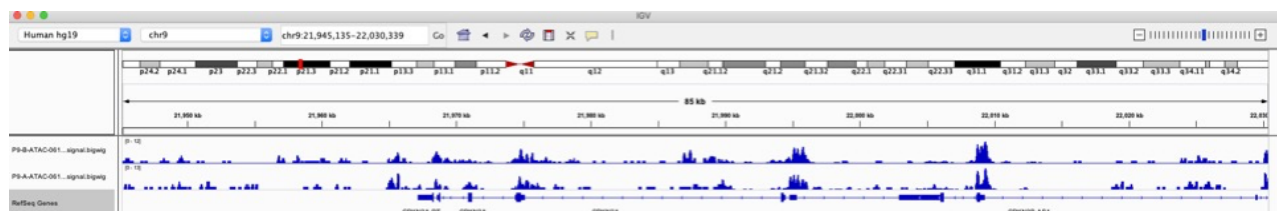
Interpret the generated output file, like narrowpeak/bigwig colnames (输出文件列名含义写入word)

###<https://genome.ucsc.edu/FAQ/FAQformat.html#format12>

6. Visualize signal/peak at MYC(基因)

Bigwig/peak visualize by IGV/WashU (导入IGV的MYC处的截图写入word)

look like



7. Annotate narrowpeak by genetic elements

Annotate narrowpeak region (Exon/Intron/promoter/Intergenic etc)

Homer download

```
###见http://homer.ucsd.edu/homer/introduction/install.html
或者conda 安装
###下载 homer hg19 注释文件
###或者使用R包安装注释peak
ChIPseeker/ChIPpeakAnno
```

annotate peak

```
###见http://homer.ucsd.edu/homer/ngs/annotation.html
```

output like

PeakID	Chr	Start	End	Strand	Peak Score	Focus Ratio	Annotation	Detailed Annotation	Distance to TSS	Nearest Protein
Peak_3130	chr7	56399903	56400084	+	0	0	Intergenic	Intergenic	116075	NR_029420
Peak_3180	chr15	45407493	45407693	+	0	0	intron (NM_207581, intron 1)	intron (NM_207581)	1069	NM_207581
Peak_4950	chr10	132667360	132667735	+	0	0	Intergenic	Intergenic	93384	NR_036180
Peak_4408	chr10	43491110	43491326	+	0	0	Intergenic	MLT1D L	-1793	NR_049836
Peak_1678	chr22	44692407	44692807	+	0	0	exon (NM_001099294, exon 3)	exon (NM_001099294)	16224	NM_001099
Peak_3399	chr4	24977888	24978062	+	0	0	intron (NM_173463, intron 1)	intron (NM_173463)	3851	NM_173463
Peak_3787	chr19	38786575	38786761	+	0	0	Intergenic	HERV140	-8223	NM_033520
Peak_4005	chr3	60719625	60719912	+	0	0	intron (NR_148921, intron 4)	intron (NR_148921)	-116165	NR_128708
Peak_3418	chr1	22177822	22178204	+	0	0	intron (NM_005529, intron 55)	intron (NM_005529)	39255	NM_001013
Peak_3977	chr16	27880664	27880840	+	0	0	intron (NM_001323901, intron 1)	CT-rich L	18726	NM_144675
Peak_3555	chr22	19611552	19611726	+	0	0	Intergenic	MER82 D	-57277	NR_024381

plot pie(饼图插入word)

<http://www.sthda.com/english/wiki/ggplot2-pie-chart-quick-start-guide-r-software-and-data-visualization>

```
#统计peak anotation占比
ggplot2
###见https://www.r-graph-gallery.com/piechart-ggplot2.html
```

output like



8. Plot heatmap at gene TSS region(图片插入word)

Use deeptools computematrix/ plotProfile/ plotHeatmap

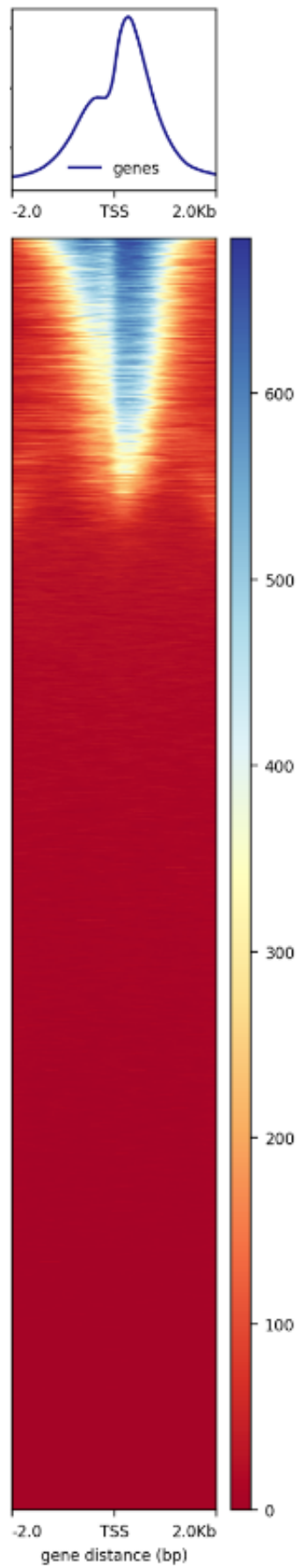
```
##deeptools
computematrix
https://deeptools.readthedocs.io/en/develop/content/tools/computematrix.html
plotHeatmap
https://deeptools.readthedocs.io/en/develop/content/tools/plotHeatmap.html
plotProfile
https://deeptools.readthedocs.io/en/develop/content/tools/plotProfile.html
```

```
gene TSS bed region download
####http://grch37.ensembl.org/biomart/martview/f96c305548e80753a6cff1f2ad0d7b93
选择label下载文件
```

The screenshot shows the Ensembl Biomart interface. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. The main header displays the GRCh37 logo and a search bar for the Stable ID of the Transcript. Below the header, there are tabs for New, Count, and Results, along with buttons for URL, XML, Perl, and Help. The left sidebar shows the Dataset selection (Human genes (GRCh37.p13)) and the Attributes selection (Gene stable ID, Transcript start (bp), Transcript end (bp), Transcription start site (TSS), Gene name). The main content area displays the selected attributes and their corresponding values, including Ensembl and Phenotype information.

Attribute	Value
Gene stable ID	ENSG00000171862
Transcript start (bp)	11868
Transcript end (bp)	12148
Transcription start site (TSS)	11868
Gene name	BRCA1

output like



9. Summary(解释6, 7产生图的含义, 用生物学描述)

Summarize your analysis and try to clarify the biological meaning

Homework Submmit

In the end of this week, the files you need submit are as follow :

1. A word document (*.docx)

The document should contain:

Set up the environment process

output file description

Quality control description

pie

heatmap

IGV MYC signal

Summary

2. Two result files(*fc.signal.bw/*narrowPeak)

Bigwig file

Narrowpeak file

Please compress files into a *zip and submit it on time !