This pipeline process raw CUT&RUN sequening reads from mouse to generate BAM, bedGraph and bigwig files.

Processing of Raw reads

Steps to be followed to run the pipeline.

1. Create Project folder

```
$ mkdir -p Name_of_project_folder/data/
$ mkdir -p Name_of_project_folder/results/
$ mkdir -p Name_of_project_folder/figures/
$ mkdir -p Name_of_project_folder/scripts/
```

2. Create sampleSheet.csv *Note: the content of sampleSheet.csv should not contain space* e.g.

```
SampleID,Read1.fastq.gz,Read2.fastq.gz
7wk_MECP2_WT_LR_CTX_H3K27ac_ab4729_TC240531_A1,TC240531_A1_S1_R1_001.fastq.
gz,TC240531_A1_S1_R2_001.fastq.gz
7wk_MECP2_WT_LR_CTX_H3K4me1_ab8895_TC240531_A2,TC240531_A2_S2_R1_001.fastq.
gz,TC240531_A2_S2_R2_001.fastq.gz
```

save the sampleSheet.csv in `Name_of_project_folder/data/`

3. Download the mm10 bowtie2 index

Note: skip this step if you already have the reuqired index files of the genome

```
$ mkdir -p Name_of_project_folder/data/bowtie2_mm10_index
$ wget https://genome-idx.s3.amazonaws.com/bt/mm10.zip -P
Name_of_project_folder/data/bowtie2_mm10_index
$ unzip Name_of_project_folder/data/bowtie2_mm10_index/mm10.zip -d
Name_of_project_folder/data/bowtie2_mm10_index/
```

\

4. Run the pipeline with following argument as inputs

```
$ perl process_CNR.pl \
<result dir> \ # the path of result directory
<raw file directory> \ # the path of directory containing raw files
<bowtie2_index directory> \ # the path of directory containing bowtie2
index
<sampleSheet.csv> # the path of directory with files (e.g sampleSheet.csv)
```

e.g.

```
$ perl process_CNR.pl \
/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder
/results/240531/ \
/project/OBI/Neuroinformatics_Core/Stroud_lab/shared/D3aCNR_forGyan/ \
/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder
/data/bowtie2_mm10_index/ \ \
/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder
/data/sampleSheet.csv
```

5. Generate Qualimap report in tabular format for all the samples

```
$ Python GenerateQualimap_Report.py -d <Qualimap directory>
```

e.g.

```
$ Python GenerateQualimap_Report.py -d
/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder
/results/240531/qc_qualimap
```

open below file to to see the alignment statistics
`/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder/results/240531/qc_qualimap/Qualimap_report.tsv`

Details of output folder :

1. `Name_of_project_folder/results/240531/bam` This directory contains the final duplicate removed alinged coordinate sorted bam file and its index.
2. `Name_of_project_folder/results/240531/bamCoverage` This directory contains RPKM normalized bedGraph and bigwig files with 10bp bin size generated by bamCoverage program of

deeptools

3. `Name_of_project_folder/results/240531/fastqc_before_trimmomatic` This directory contain FASTQC amd multiqc output of original fastq files.
4. `Name_of_project_folder/results/240531/fastqc` This directory contain FASTQC and multiqc output of trimmed fastq files.
5. `Name_of_project_folder/results/240531/logs` This directory contain slrum script and log file for each sample
6. `Name_of_project_folder/results/240531/qc_qualimap` This directory contain output directory from Qualimap program which reports alignment statistics.

Peak calling (macs2)

1. Prepare tab separated sampleSheet.tsv The targetID and controlID should be smapleID used in the sampleSheet.csv file.

```
peakCallingID   targetID    controlID
MECP2_vs_IgG
240531TC_A1_7wk_MECP2_WT_LR_CTX_H3K27ac_ab4729,240531TC_A1_7wk_MECP2_WT_LR_
CTX_H3K27ac_ab4729
240531TC_A1_7wk_MECP2_WT_LR_CTX_H3K27ac_ab4729,240531TC_A1_7wk_MECP2_WT_LR_
CTX_H3K27ac_ab4729
```

*save the file in `Name_of_project_folder/data/macs2_sampleSheet.tsv`

2. Run the peak calling script

```
$ perl macs2_callPeak.pl \
<result directory> \ # the path of result directory
<pvalue or qvalue> \ # pvalue - to find peaks with p-value cut-off of 1e-3;
qvalue - to find peaks with qvalue cut-off of 0.05
<macs2_sampleSheet.tsv> # path with name of macs2_sampleSheet.tsv
```

e.g

```
$ perl macs2_callPeak.pl \
/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder
/results/240531/ \
pvalue \
/work/OBI/Neuroinformatics_Core/s225347/CNR_pipeline/Name_of_project_folder
/data/macs2_sampleSheet.tsv \
```

\

Details of macs2 output directory :

1. `Name_of_project_folder/results/240531/macs2_callPeak/MECP2_vs_IgG/`
   This folder will be created after running `macs2_callPeak.pl` and it will contain following files.
   - a. MECP2_vs_IgG_model.r
   - b. MECP2_vs_IgG_peaks.narrowPeak
   - c. MECP2_vs_IgG_peaks.xls
   - d. MECP2_vs_IgG_summits.bed
   - e. MECP2_vs_Igg_peaks.narrowPeak.noBlacklist.bed
   - f. MECP2_vs_IgG_summits_noBlacklist.bed