Peak calling

Using MACS2

For both the day 0 and day 3 of differentiation into adipocytes, two files are available

- input, as control
- histone modification H3K4

MACS2 is going to use both files to normalize the read counts and perform the peak calling.

Retrieve the BAM files with all chromosomes

```
cd ~/chip-seq
mkdir bams
cd bams
ln -s /work/users/aginolhac/chip-seq/data/*.bam .
```

Perform peak calling

```
macs2 callpeak -t TC1-H3K4-ST2-D0.GRCm38.p3.q30.bam \
               -c TC1-I-ST2-D0.GRCm38.p3.q30.bam \
               -f BAM -g mm -n TC1-ST2-H3K4-D0 -B -q 0.01 --outdir TC1-ST2-H3K4-D0 &
macs2 callpeak -t TC1-H3K4-A-D3.GRCm38.p3.q30.bam \
               -c TC1-I-A-D3.GRCm38.p3.q30.bam \
               -f BAM -g mm -n TC1-A-H3K4-D3 -B -q 0.01 --outdir TC1-A-H3K4-D3
```

check model inferred by MACS2

first load R as a module and execute R script.

```
module load lang/R
Rscript TC1-A-H3K4-D3/TC1-A-H3K4-D3_model.r
Rscript TC1-ST2-H3K4-D0/TC1-ST2-H3K4-D0_model.r
```

fetch the pdf produced.

sort per chromosomes and coordinates

```
find TC* -name '*.bdg' | parallel "sort -k1,1 -k2,2n {} > {.}.sort.bdg"
```

convert to bigwig

in order to get smaller files

```
find TC* -name '*sort.bdg' | parallel -j 1 "/work/users/aginolhac/chip-seq/bedGraphToBigWig {} /work/users/
```

Fetch the files and display them in IGV

Perform peak calling with broad option

```
macs2 callpeak -t TC1-H3K4-ST2-D0.GRCm38.p3.q30.bam \
               -c TC1-I-ST2-D0.GRCm38.p3.q30.bam \
               -f BAM --broad -g mm -n TC1-ST2-H3K4-D0-broad -B -q 0.01 --outdir TC1-ST2-H3K4-D0-broad &
macs2 callpeak -t TC1-H3K4-A-D3.GRCm38.p3.q30.bam \
               -c TC1-I-A-D3.GRCm38.p3.q30.bam \
               -f BAM --broad -g mm -n TC1-A-H3K4-D3-broad -B -q 0.01 --outdir TC1-A-H3K4-D3-broad
```

GREAT analysis

The website GREAT allows to paste bed regions of enriched regions.

predict functions of cis-regulatory regions

Using the `TC1-A-H3K4_peaks.narrowPeak` file produced by MACS2.

This file has the different fields:

1. chromosome
2. start
3. end
4. peak name
5. integer score for display
6. strand
7. fold-change
8. -log10pvalue
9. -log10qvalue
10. relative summit position to peak start

Let's format the file as a 3 fields BED file and focus on more significant peaks filtering on *q-values*.

```
awk '$9>40' TC1-A-H3K4_peaks.narrowPeak | cut -f 1-3 | sed 's/^/chr/' > TC1-A-H3K4_peaks.bed
```

then

- load the BED in GREAT
- for the relevant genome, `mm10`
- association rule: single nearest genome

Differential peak calling

ODIN allows to compare two conditions associated with their own controls.

A command line looks like

```
rgt-ODIN  --input-1=../TC1-I-ST2-D0.GRCm38.p3.q30.bam \
          --input-2=../TC1-I-A-D3.GRCm38.p3.q30.bam \
          -m -n TC1-I-A-D0vsD15 -v \
          TC1-H3K4-ST2-D0.GRCm38.p3.q30.bam TC1-H3K4-A-D3.GRCm38.p3.q30.bam \
          ../references/GRCm38.p3.fasta ../references/GRCm38.p3.chom.sizes
```

Previous

Next