

Exercise 9

ChIP-Seq data analysis using MACS.

Ewa Szczurek
MIM UW

November 30, 2015

Exercise 1. We will follow the Nature Protocol with several modifications. We start by preparing the data. For the class we need to limit the data to chr1. You can download the outcome of this step from http://www.mimuw.edu.pl/~szczurek/TSG2/09_lab/

1. Download University of Washington H3K4me3 data from <http://cistrome.dfci.harvard.edu/MACSNatureProtocol/>
2. Limit the bam files to contain alignment to chr1 only

```
./samtools index ../../UW_K562_H3K4me3.bam
./samtools index ../../UW_K562_H3K4me3_Control.bam

./samtools view -b ../../UW_K562_H3K4me3.bam chr1>H3K4me3chr1.bam
./samtools view -b ../../UW_K562_H3K4me3_Control.bam chr1>H3K4me3chr1_control.bam

./samtools view -h -F 4 -b H3K4me3chr1.bam > H3K4me3chr1mapped.bam
./samtools view -h -F 4 -b H3K4me3chr1_control.bam > H3K4me3chr1mappedControl.bam
```

Exercise 2. Running MACS and interpreting the results.

1. Inspect the macs flowchart in http://www.mimuw.edu.pl/~szczurek/TSG2/09_lab/
2. Run macs

```
macs14 -t H3K4me3chr1mapped.bam -c H3K4me3chr1mappedControl.bam -g hs -n H3K4me3
```

3. How much time did it take to run?
4. How many peaks are called?
5. inspect the *H3K4me3_{peaks.xls}* file. How are the FDR values computed?
6. For an example row

```
chr1 910810 912585 1776 968 52 242.54 21.44 2.75
```

what is the p-value for that peak? For 100 peaks called in the ChIP sample with that p-value, what would be the number of peaks in the control sample with the same p-value?

7. Run

```
Rscript H3K4me3_model.r
```

The output image illustrates the distribution of reads on positive and negative strands in the model peaks and the estimated fragment size.

8. Compare the output plot to Figure 2 in the protocol. Where do the differences come from?

Exercise 3. Visualization with IGV

1. Load the bed files into IGV and inspect.
2. Where do the H3K4me3 usually peak?

Exercise 4. Run MACS for the full dataset

```
macs14 -t UW_K562_H3K4me3.bam -c UW_K562_H3K4me3_Control.bam -g hs -n H3K4me3Full
```

1. How much time did it take to run?
2. What is the estimated d ?
3. How many peaks are called?

Homework 1. No homework as of today!