Handson 3:

The goal is to figure out the potential motif of protein A. The RAP/CLIP-Seq results of protein A was performed. And the peak binding summit of RAP/CLIP-Seq results from two replicates were provided as in the table below (based on MACS1.4 peak caller).

|  |  |  |
| --- | --- | --- |
|  | Peak on Watson Strand | Peak on Crick Strand |
| Replicate1 | R1\_plus\_summits.bed | R1\_minus\_summits.bed |
| Replicate2 | R2\_plus\_summits.bed | R2\_minus\_summits.bed |

The reference genome is : max\_circle.fasta

The annotated transcripts on this genome is provided: max\_circle.gff

Note: **the peak location are stranded specific. The protein is considered to bind to short sequence motif, <=10nt.**

What should be included in the writeup:

1. How you come up with the 1 or more motif based on the data
2. [optional] If you use more than one tools, try to summarize the information and figure out which one motif and more reliable?
3. [optional] Based on your test, what do you think are the key parameters in the tools you use for motif prediction.
4. [optional] check what is the percent of the sites with the predicted motif is bound by protein A in the genome or transcriptome.

Reference:

1. <http://meme-suite.org/>
2. <http://homer.ucsd.edu/homer/motif/>
3. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. ***Nature Protocols*** 7, 1551–1568 (2012) doi:10.1038/nprot.2012.088
4. <http://ccmbweb.ccv.brown.edu/gibbs/gibbs.html>
5. <http://fraenkel.mit.edu/webmotifs-md-programs.html>
6. <http://159.149.160.51/modtools/>