Module 6 Lab: Motif discovery with MEME-ChIP

Michael M. Hoffman[1,2,3]

April 16, 2015

[1] Princess Margaret Cancer Centre, Toronto Medical Discovery Tower, 101 College St, Toronto, ON M5G 1L7, Canada.
[2] Department of Medical Biophysics, University of Toronto, Toronto Medical Discovery Tower, 101 College St, Toronto, ON M5G 1L7, Canada.
[3] Department of Computer Science, University of Toronto, 10 King's College Rd, Toronto, ON M5S 3G4, Canada.

Reading for this laboratory: Ma, Noble, and Bailey, "Motif-based analysis of large nucleotide data sets using MEME-ChIP", *Nat Protoc* 9:1428.

This laboratory contains instructions to carry out a MEME-ChIP analysis (marked like "1."). There are also questions to help you interpret the process and results (marked like like "**Q1.**").

Preparing for and executing MEME-ChIP

1. Create a Galaxy account.
   - Visit Galaxy (https://usegalaxy.org/).
   - From the top menu, select "User" and then "Register".
   - Follow the instructions to create your account.
2. Download ENCODE c-Myc ChIP-seq peaks in A549 cells.
   - Adapt your activities from the instructions in Ma *et al.*, Box 5. But do not set "ENCODE Data Freeze."
3. Upload the narrowPeak file to Galaxy.
4. Import the Galaxy workflow "Create MEME-ChIP input input FASTA file (500bp summit regions) from ENCODE narrowPeak file."
5. Run the Galaxy workflow.
6. Save the created FASTA file.
7. Submit your FASTA file to MEME-ChIP.
   - Adapt your activities from the instructions in Ma *et al.*, Procedure (p. 1437).
8. It may take a few hours for your MEME-ChIP job to complete.

   Q1. How could you make this complete more quickly?

Examining MEME-ChIP results

9. Find a link on the Wiki to a previously completed MEME-ChIP report on another set of ChIP-seq peaks, created using a novel DNA-binding factor.
10. Examine the discovered motifs.

    Q2. Do any of the motifs have likely biological relevance? Why?

    Q3. Are any of the motifs likely to be irrelevant? Why?

    Q4. What known transcription factor is the novel DNA-binding factor similar to?

Q5. Which discovery programs (MEME, DREME, or Centrimo) were important in determining this?

11. Research the most similar transcription factor by using Google to search for it in GeneCards. Your search query is the name of the transcription factor with the modifier site:genecards.org.

    Q6. The name of the transcription factor describes its consensus motif. How does this consensus motif match your discovered motif?

12. Find the most relevant motif and use it to search the whole human genome with FIMO.

    1) Click the "Show *X* More" link.
    2) Find the first row where the "Discovery/Enrichment Program" is MEME and visit its report.
    3) Click the "Submit" pushbutton.
    4) Select "FIMO".
    5) Click the "Send" pushbutton.
    6) Set database "Category" to "ENSEMBL Genomes".
    7) Set "Database" to "homo sapiens (peptide and nucleotide)".
    8) Click "Start search".

13. This will also take some time to complete.

Examining FIMO results

14. Find a link on the Wiki to a previously completed FIMO report on a motif you discovered in the previous section.

15. Examine the motif occurrences.

    Q7. Which of these motif occurrences represent biologically relevant TFBSs?

    Q8. What could you do to increase the proportion of relevant binding sites?