

Motif analysis practical

Goal: Define binding motif(s) for the ChIPed transcription factor and identify potential cofactors

Tools: RSAT Metazoa (<http://rsat.sb-roscoff.fr/index.php>)

1 - Retrieve the peak sequences corresponding to the peak coordinate file (bed)

For the motif analysis, you first need to extract the sequences corresponding to the peaks. There are several ways to do this (as usual...). If you work on a UCSC-supported organism, the easiest is to use RSAT fetch-sequences

1. In the menu on the left, in **NGS - ChIP-seq**, choose **fetch-sequences from UCSC**
2. In the form page, choose the **mm10 - Mouse Dec. 2011 (GRCm38/mm10)** genome
3. Upload the bed file with peak coordinates (remove brackets in file name if any)
4. **Go !**
5. In the **Result file(s)** table, check that the **Fetches sequences (fasta)** link contains sequences
6. In the **Result file(s)** table, check that the **Log file (txt)** link does not contain any error message

2 - Motif discovery

2. In the fetch-sequence output, click on the **peak-motifs** button below the **Result file(s)** table. A new page opens, with a form
3. The default peak-motifs web form only displays the essential options. There are only two mandatory parameters:
 - a. The **Title** box, prefilled with information based on your previous step (**fetch-sequences**). You may want to adapt to your personal tastes.
 - b. The **Peak sequences** that were transferred from the previous step (**fetch-sequences**)
4. We will now modify some of the advanced options in order to fine-tune the analysis according to your data set.
 - Open the **Reduce peak sequences** title, and make sure the **Cut peak sequences: +/-** option is set to 0 (we wish to analyze our full dataset)
 - Open the **Motif Discovery parameters** title, and make sure that **Discover over-represented words**, **Discover words with a positional bias** and **Oligomer lengths 6 and 7** (but not 8) are checked.
 - Under **Compare discovered motifs with databases**, select **Jaspar** in the database list and **JASPAR core nonredundant vertebrates** in the collection list
5. You can indicate your email address in order to receive notification of the task submission and completion. This is particularly useful because the full analysis may take some time for very large datasets.
6. Click **GO**. As soon as the query has been launched, you should receive an email confirming the task submission, and providing a link to the future result page.
7. Upon completion, you will receive another email, again with a link to the result page.