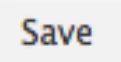
# Analysis of ChIP-seq peaks (answers to questions)

Stéphanie Le Gras
(slegras@igbmc.fr)

# Exercise 1: peak annotation

- 1.
  - Search for “homer annot” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Homer peaks OR BED format: MITF peaks - narrow peaks dataset (2nd run of Macs2)
    - Genome version: hg38
  - Click on 

- 2.
  - The Homer annotatePeaks tool generates two datasets: a log file and a tabular file containing annotated peaks.
  - Click on the  of the dataset which contain annotated peaks.
  - Click on the Datatype tab
  - Select **tabular** in the drop down list “New Type:”
  - Click on 

# Exercise 1: peak annotation

- 3.
  - Search for “histogra” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Dataset: tabular file which contains annotated peaks
    - Numerical column for x axis: column: 10
    - Plot title: Frequency of peaks relative to TSS
    - Label for x axis: Distance to TSS
  - Click on 

- 4.a.
  - Search for “Cut” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Cut columns: c8
    - Delimited by: Tab
    - From: tabular file which contains annotated peaks
  - Click on 

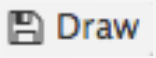
# Exercise 1: peak annotation

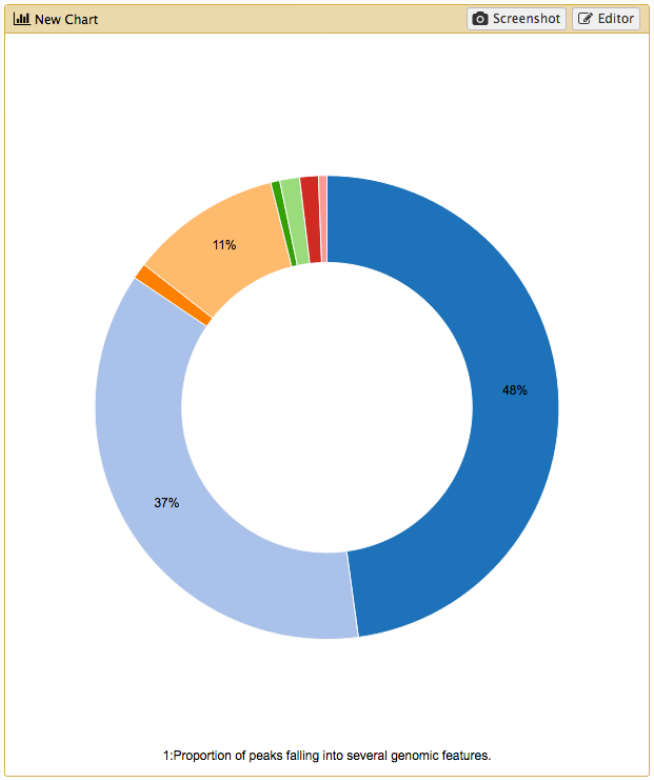
- 4.b.
  - Search for “Convert” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Convert all: whitespaces
    - in Dataset: resulting dataset after 4.a
  - Click on 

- 4.c.
  - Search for “Remove” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Remove first: 1
    - From: resulting dataset after 4.b
  - Click on 

- 4.d.
  - Search for “Count” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - from dataset: resulting dataset after 4.c
    - Count occurrences of values in column(s): column: 1
    - How should the results be sorted?: With the most common values first
  - Click on 

# Exercise 1: peak annotation

- 4.e.
  - Expand the box of the dataset generated in 4.d, click on  and select Charts
  - Double click on Pie charts
  - Fill the form:
    - Provide a label: Proportion of peaks falling into several genomic features.
    - Labels: Column: 2
    - Values: Column: 1
  - Click on 

# Exercise 2: *de novo* motif discovery

- 1.a
  - Search for “Sort” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Sort Dataset: dataset with peak summits
    - on column: Column: 5
    - with flavor: Numerical sort
    - everything in: Descending order
  - Click on 

- 1.b
  - Search for “select first” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Select first: 800
    - From: dataset generated in 1.a
  - Click on 

# Exercise 2: *de novo* motif discovery

- 2.a
  - Import the file which contains chromosome lengths
  - Click on Shared Data (top menu) and select “Data Libraries”
  - Click on “Chromosome length”
  - Select the dataset named hg38.len (tick boxes beside dataset names)
  - Click on the button “To history”
  - Select history: ChIP-seq data analysis
  - Click on “Import”
  - Go back to the main page by clicking on “Analyzed data” (top menu)
- Run slopBed
  - BED/VCF/GFF file: MACS14_in_Galaxy_summits.bed
  - Genome file: hg19.len
  - Choose what you want to do: Increase the BED/VCF/GFF entry by the same number of base pairs in each direction. (default)
  - Number of base pairs: 100
  - Click on 

# Exercise 2: *de novo* motif discovery

- 3.
  - Search for “extract” in the search field (tool panel)
  - Click on the name of the tool
  - Set parameters:
    - Fetch sequences for intervals in: the dataset generated in 2.c
    - Interpret features when possible: No
    - Click on 

- 4.
  - Expand the box of the dataset generated in 3 and click on  to download the file
- 5.
  - Go to MEME-chIP website and run the tool with the fasta file you’ve just downloaded and with default parameters.

# Exercise 3: Clustering

- 1.
  - Select clusters 2, 3, 4, 6, 9 and click on Export Selected clusters
  - Import the file previously exported as reference coordinates. Click on browse, go to the directory which contains the file and click on open.
  - Click on Extract data
  - Click on Clustering