Correlation of RNA-seq and ChIP-seq data

Stéphanie Le Gras
(slegras@igbmc.fr)

Exercise 1

- We want to know how many up-regulated genes contain a peak for MITF
  - Compare Gene names of the chIPseq data (annotation step) and the RNAseq data (up-regulated genes).
    - Use the file siMitfvssiLuc.up.annot.txt (annotated with BioMart)
    - All chIPseq peaks are used (annotated with Homer)
  - Use Venny (https://bioinfogp.cnb.csic.es/tools/venny/)

Exercise 2

- Use seqMINER to visualize at the same time chIP-seq data along with RNA-seq data. RNAseq data have to be formatted before using them into seqMINER
  - seqMINER accept a tab separated file formatted like:
    - Gene ID <tab> Expression Values
    - Expression values are normalized read counts divided by gene length in Kb
  - This file can generated using the Galaxy tool **RNAseqDataAnnotation** and other **Galaxy tools** like this (but we won't do it as it takes too much time):
  1. Create a new History named "**Correlation**"
  2. Use the tool **RNAseqDataAnnotation** in Galaxy. This tool takes the output of SARTools as input file and annotate it with genes and do gene length normalisation.
     1. **Select file** : the file siMitfvssiLuc.complete.txt you generated with SARTools
     2. **Select the species for your data**: Homo Sapiens
     3. **Ensembl version?** Version 95

Exercise 2

3. The tool take time to run.

4. We are going to create a two columns file from this file. The first column contains Ensembl Gene Ids (column Ensembl_gene_id) and the second column is « norm siMitf4 normalized and divided by gene length in kb ».

- Use the tool **cut** to extract this two columns from the file Data_normalization_annotation.txt
- The range of data in column « norm siMitf4 normalized and divided by gene length in kb » is too large. We are going to divide all values by 50. Use the tool **Compute an expression on every row** to achieve this.
- Use **cut** to create a final file with two columns :
  - Ensembl Gene Ids
  - « norm siMitf4 normalized and divided by gene length in kb » / 50
- Download the two columns file.

Exercise 2

- 5. Use seqMINER to visualize at the same time chIP-seq data along with RNA-seq data
    - Use the Advance (RNAseq) tab to upload a 2-columns table: 1st column contains Ensembl Gene IDs and 2nd column contains normalized read counts of MITF divided by gene length in Kb. Use the file **RNAseq_data_ready_for_seqMINER** of the directory **correlation**.
    - Use MITF peak summits (second MACS2 run) as reference coordinates. Download the file from GalaxEast or use the file **MITF_peak_summits.bed** or directory **correlation**.
    - Be careful, make sure that in Options>Gene profile, Gene profile analysis is not activated before clicking on « **Extract data** ».