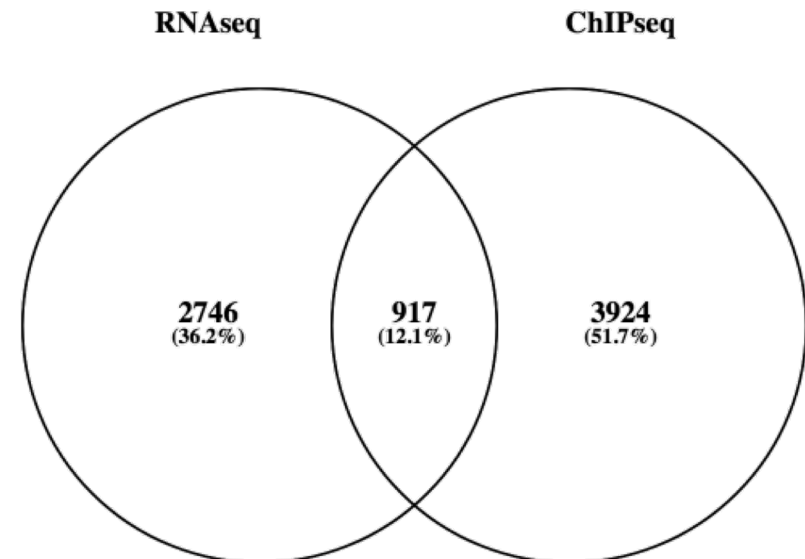
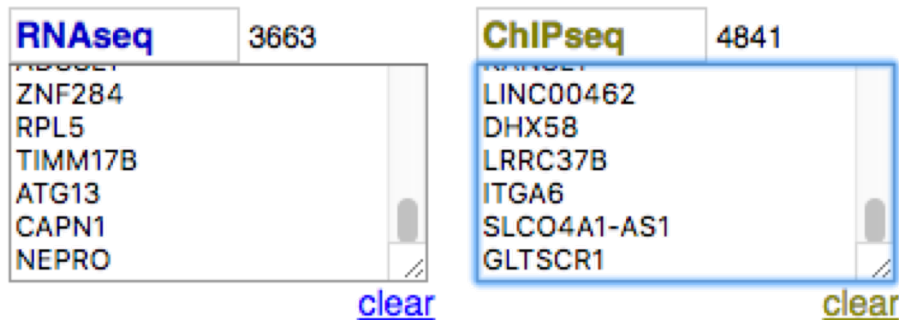


Correlation of RNA-seq and ChIP-seq data (answer to questions)

Stéphanie Le Gras
(slegras@igbmc.fr)


Exercise 1

- 1.
 - Retrieve Gene symbols of up-regulated genes (use the file siMitfvssiLuc.up.txt you annotated with BioMart)
 - Download the annotated peaks (dataset generated with HOMER). Use the Gene Name column.







Exercise 2

1.

- Click on the button  and select “create new”
- Click on the history name “Unnamed history”, erase “Unnamed history”, enter “Correlation” and press enter

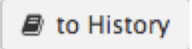
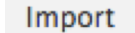
2.

- Click on the button  and select “Saved Histories”
- In the table of histories, click on the history named “RNA-seq data analysis”
- Datasets from this history appears on the history panel
- Click on the button  of the dataset named « SARTools DESeq2 tables »
- Click right on the file named « siMitfvssiLuc.complete.txt » and download it.
- Switch to your history named « Correlation »
 - Click on the button  and select “Saved Histories”
 - In the table of histories, click on the history named “Correlation”
- Import the file to Galaxy by clicking on the button  and drag and drop the file from your computer to the Galaxy window. Type: tabular, Genome: hg38.

Exercise 2

- Search for “**RNAseqDataAnnotation**” in the search field (tool panel)
- Click on the name of the tool
- Fill in the parameters:
 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


3.

- Click on Shared Data (top menu) and select “Data Libraries”
- Click on “NGS data analysis training ” > “Correlation”
- Select Data_normalization_annotation.txt (tick the box beside dataset name)
- Click on the button  to History
- Select history: Correlation
- Click on 
- Go back to the main page by clicking on “Analyzed data” (top menu)

4.

- Search for “**cut**” in the search field (tool panel)
- Click on the name of the tool “Cut columns from a table »
- Fill in the parameters:
 - **Cut columns**: C1, C13
 - **From**: Data_normalization_annotation.txt

Exercise 2

- Search for “**compute**” in the search field (tool panel)
- Click on the name of the tool “Compute an expression on every row »
- Fill in the parameters:
 - **Add expression: $c2/50$**
 - **as a new column to:** Cut on data 1
- Search for “**cut**” in the search field (tool panel)
- Click on the name of the tool “Cut columns from a table »
- Fill in the parameters:
 - **Cut columns: C1, C3**
 - **From:** Compute on data 2
- Download the dataset
 - Click on the name of the dataset: Cut on data 3
 - Click on  to download the file

Exercise 2

- 2.
 - Download MITF peaks (Output of MACS2 narrow peaks) -> Use it as reference coordinates in seqMINER

The screenshot displays the seqMINER web interface with the following components:

- Step 1: Load data**
 - Load reference coordinates (i.e. peaks): A text input field with a dropdown arrow.
 - Select assembly: A dropdown menu showing "hg38_ense...".
 - Advanced: A button.
 - Load aligned reads: A list of files including "H3K4me3.sort.bam", "mitf.sort.bam", and "poIII.sort.bam".
 - Browse ...: A button.
 - Load file(s) >>: A button.
- Step 2: Data extraction**
 - Galaxy25-[MACS2_callpeak_or 7352 peaks. Peak length mean: 325: A text input field with a dropdown arrow.
 - RNA-Seq expression: A dropdown menu showing "Galaxy4-[RNAseq_data_re".
 - B...: A button.
 - Selected datasets: A list of files including "mitf.sort.bam", "H3K4me3.sort.bam", and "poIII.sort.bam".
 - Delete: A button.
 - Extract data: A button.

Annotations with orange callout boxes:

- "Ensembl 95 (hg38)" points to the "Select assembly" dropdown.
- "ChIP-seq datasets" points to the "Load aligned reads" list.
- "MITF peaks" points to the "Galaxy25-[MACS2_callpeak_or 7352 peaks..." field.
- "RNAseq data" points to the "Galaxy4-[RNAseq_data_re" dropdown.

At the bottom of the interface, there is a "100%" progress indicator.

Exercise 2

- Go to Density Array Method (top tabs)
- Click on Extract data
- Click on Clustering

Exercise 2

The interface displays three heatmaps: **mitf.sort.bam**, **H3K4me3.sort...**, and **poll.sort.bam**. A vertical color bar on the left indicates 10 clusters. The **Contrast** slider is set to 20. The **KMeans seed** is 97787647. The **Color palette** is visible. The cluster analysis results are as follows:

Cluster	Number of elements
Cluster 1	837 elements
Cluster 2	3044 elements
Cluster 3	1368 elements
Cluster 4	380 elements
Cluster 5	304 elements
Cluster 6	401 elements
Cluster 7	211 elements
Cluster 8	112 elements
Cluster 9	569 elements
Cluster 10	126 elements

Navigation options include **Peaks(BED)**, **Merged dataset profile**, **Mean profile**, **Heatmap**, and **Density values**. Action buttons at the bottom are **Export selected clusters** and **Save profile**.