

Correlation between RNA-seq and ChIP-seq data

Stéphanie Le Gras
slegras@igbmc.fr

Guidelines

- Clustering of ChIP-seq data and RNA-seq data
- Comparing data with Venn diagram

CLUSTERING

Clustering

The screenshot displays the seqMINER 1.3.3 software interface, which is organized into three main steps for data processing and clustering. The interface includes a menu bar (File, Tools, Help) and three tabs: Density Array Method, Enrichment Based Method, and Advance(RNA-Seq). The current view shows the following components:

- Step 1: Load data**: Contains a text input for "Load reference coordinates (i.e. peaks) ..." with a "Browse ..." button, and a larger empty box for "Load aligned reads" with a "Browse ..." button and a "Load file(s) >>" button.
- Step 2: Data extraction**: Features a large empty box for "Selected datasets:" and a "Delete" button. A vertical double-headed arrow is positioned between this step and Step 3.
- Step 3: clustering**: Includes a "Distribution list:" section with a large empty box, a "Clustering Normalization:" dropdown menu set to "KMeans raw", and an "Expected Number of Clusters:" input field set to "10". A "Clustering" button is located at the bottom of this section.

A progress bar at the bottom of the window indicates 0% completion.

Clustering

The screenshot shows the seqMINER 1.3.3 software interface. The window title is "seqMINER 1.3.3" and the menu bar includes "File", "Tools", and "Help". There are three tabs: "Density Array Method", "Enrichment Based Method", and "Advance(RNA-Seq)".

Step 1: Load data

- Buttons: "Load reference coordinates (i.e. peaks)", "Advanced", "Browse ...", "Load file(s) >>"
- Dropdown: "Select assembly" with "hg19_ense..." selected.
- List "Load aligned reads":
 - poll.chr2.bed.gz
 - mitf_2.chr2.bed.gz
 - H3K4me3.chr2.bed.gz

Step 2: Data extraction

- Text: "peak.bed", "908 peaks.", "Peak length mean: 837"
- Text: "RNA-Seq expression" with "RNAseq_seqminer.txt" selected and a "B..." button.
- List "Selected datasets":
 - poll.chr2.bed.gz
 - H3K4me3.chr2.bed.gz
 - mitf_2.chr2.bed.gz
- Buttons: "Delete", "Extract data"

A progress bar at the bottom indicates "100 %".

Ensembl 75
(hg19)

ChIP-seq
datasets

MITF peaks

RNAseq data

Clustering

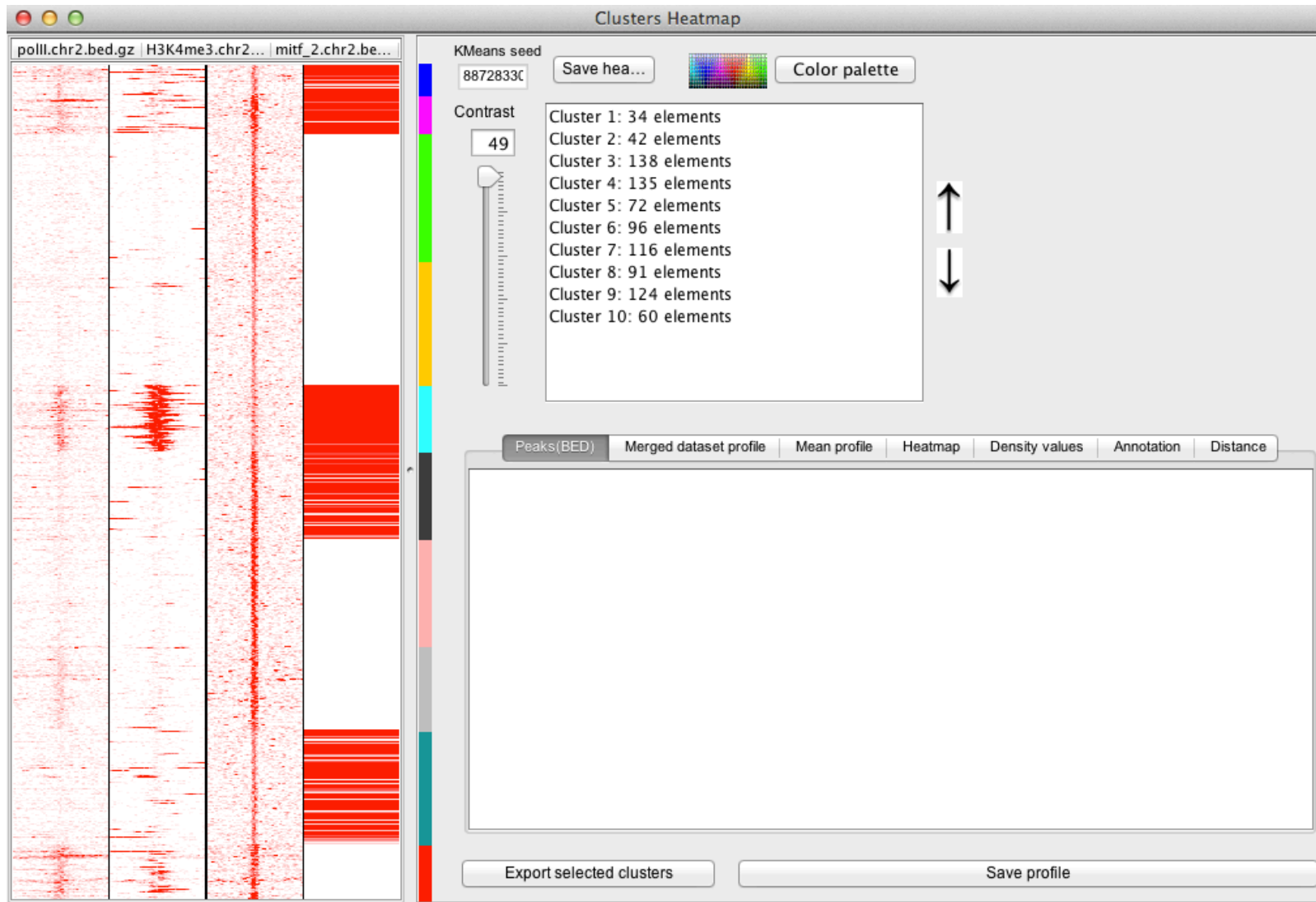
The screenshot displays the seqMINER 1.3.3 software interface, which is organized into three main steps:

- Step 1: Load data**
 - Load reference coordinates (i.e. peaks): Includes a "Browse ..." button.
 - Load aligned reads: A list of files including `poll.chr2.bed.gz`, `mitf_2.chr2.bed.gz`, and `H3K4me3.chr2.bed.gz`. Includes a "Browse ..." button and a "Load file(s) >>" button.
- Step 2: Data extraction**
 - Input: `peak.bed` (908 peaks, Peak length mean: 837).
 - Selected datasets: `poll.chr2.bed.gz`, `H3K4me3.chr2.bed.gz`, and `mitf_2.chr2.bed.gz`. Includes a "Delete" button.
 - Buttons: "Extract data" and "Load file(s) >>" (shared with Step 1).
- Step 3: clustering**
 - Distribution list: Shows `peak.bed (poll.chr2.bed.gz, H...`.
 - Clustering Normalization: A dropdown menu currently set to "KMeans enrich...".
 - Expected Number of Clusters: A text input field containing the value "10".
 - Buttons: "Clustering" and "Load file(s) >>" (shared with Step 1).

Navigation arrows (up and down) are located between Step 2 and Step 3. A progress bar at the bottom indicates 100% completion.

Normalization:
Kmeans
enrichment
linear

Clustering



COMPARING DATA WITH VENN DIAGRAMS

Exercise 1

- We want to know how many deregulated genes contain a peak for MITF
 - Compare Ensembl Gene IDs of the chIPseq data (annotation step) and the RNAseq data (filtered excel table).
 - RNAseq data filter:
 - Adjusted p-value ≤ 0.05
 - $\log_2FC \geq 1$ or $\log_2FC \leq -1$
 - All chIPseq data were used
 - Use Venny (<http://bioinfogp.cnb.csic.es/tools/venny/>)