PLATEFORME
BIOPUCES & SÉQUENÇAGE
INSTITUT DE GÉNÉTIQUE ET DE BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Correlation between RNA-seq and ChIP-seq data

Stéphanie Le Gras
slegras@igbmc.fr

igbmc
institut de génétique et de biologie moléculaire et cellulaire

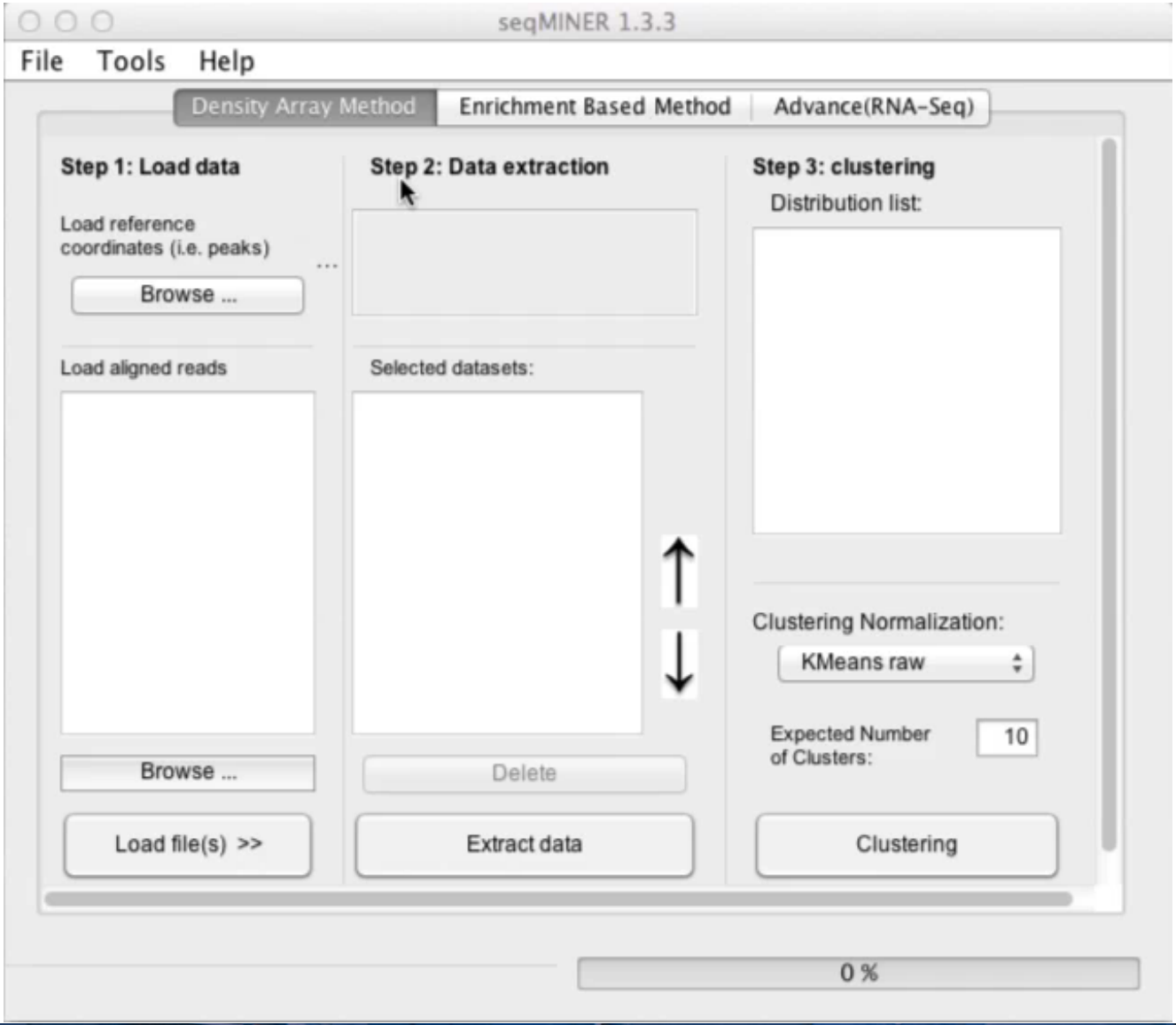
PLATEFORME BIOPUCES & SÉQUENÇAGE
INSTITUT DE GÉNÉTIQUE ET DE BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Guidelines

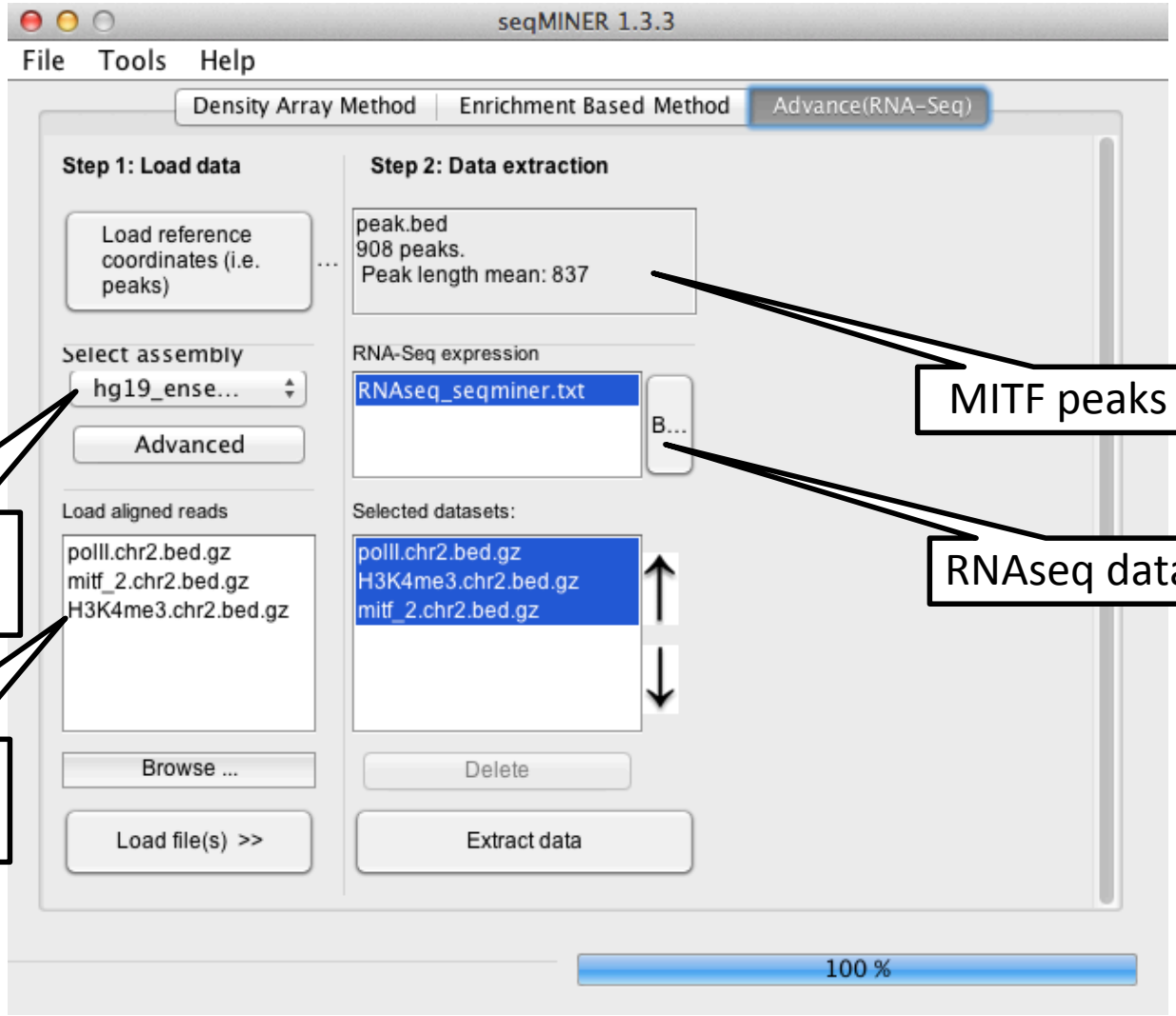
- Clustering of ChIP-seq data and RNA-seq data
- Comparing data with Venn diagram

CLUSTERING

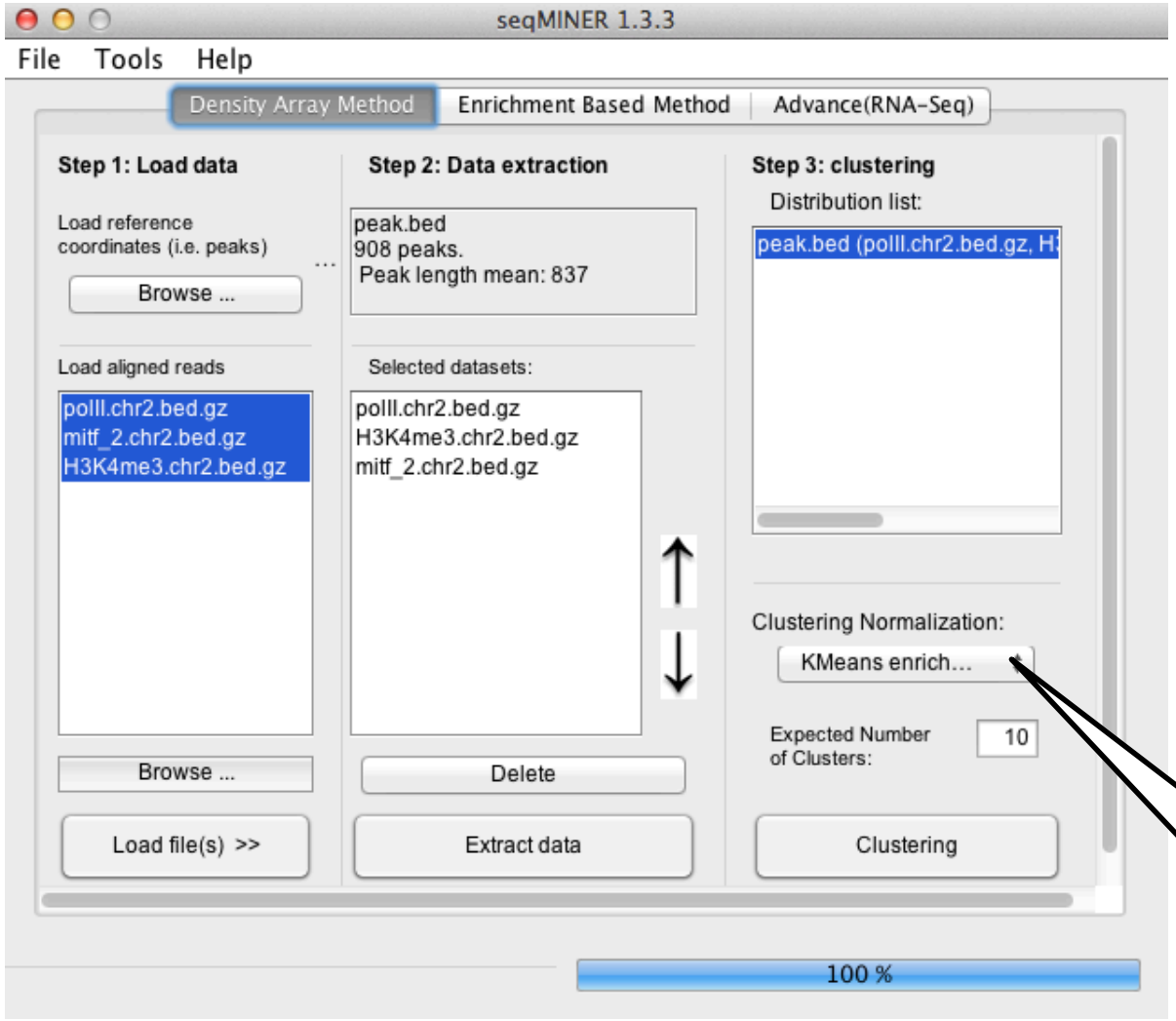
Clustering


seqMINER 1.3.3
File
Tools
Help
Density Array Method
Enrichment Based Method
Advance(RNA-Seq)
Step 1: Load data
Load reference coordinates (i.e. peaks)
Browse ...
Load aligned reads
Browse ...
Load file(s) >>
Step 2: Data extraction
Selected datasets:
Delete
Extract data
Step 3: clustering
Distribution list:
Clustering Normalization:
KMeans raw
Expected Number of Clusters:
10
Clustering
0 %

Clustering

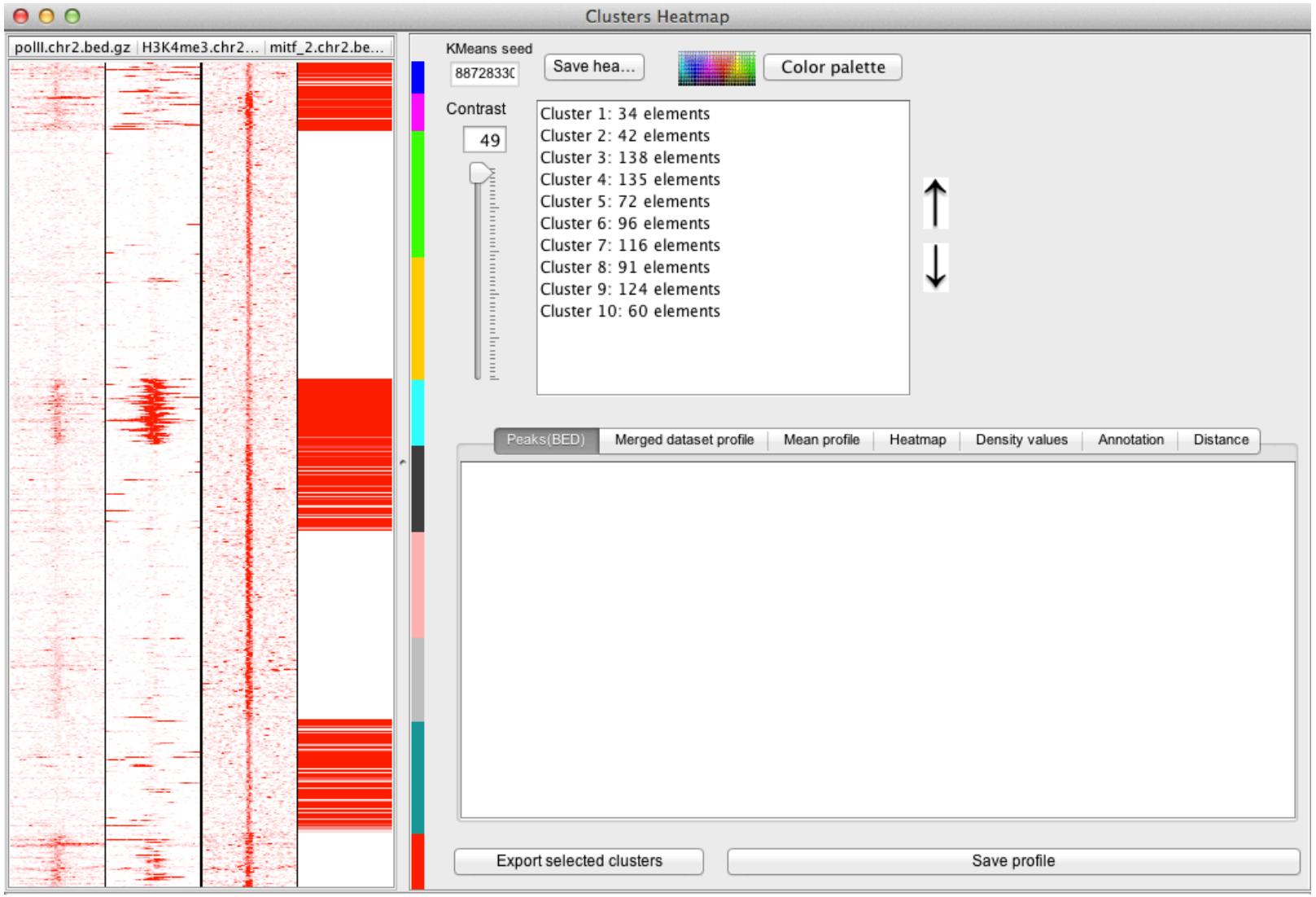

seqMINER 1.3.3
File
Tools
Help
Density Array Method
Enrichment Based Method
Advance(RNA-Seq)
Step 1: Load data
Step 2: Data extraction
Load reference coordinates (i.e. peaks)
peak.bed
908 peaks.
Peak length mean: 837
Select assembly
hg19_ense...
Advanced
RNA-Seq expression
RNAseq_seqminer.txt
B...
Load aligned reads
polII.chr2.bed.gz
mitf_2.chr2.bed.gz
H3K4me3.chr2.bed.gz
Selected datasets:
polII.chr2.bed.gz
H3K4me3.chr2.bed.gz
mitf_2.chr2.bed.gz
Browse ...
Delete
Load file(s) >>
Extract data
100 %
MITF peaks
RNAseq data
Ensembl 75 (hg19
ChIP-seq datasets

Clustering



Normalization: Kmeans enrichment linear

Clustering


Clusters Heatmap
polII.chr2.bed.gz
H3K4me3.chr2...
mitf_2.chr2.be...
KMeans seed
Save hea...
Color palette
Contrast
49
Cluster 1: 34 elements
Cluster 2: 42 elements
Cluster 3: 138 elements
Cluster 4: 135 elements
Cluster 5: 72 elements
Cluster 6: 96 elements
Cluster 7: 116 elements
Cluster 8: 91 elements
Cluster 9: 124 elements
Cluster 10: 60 elements
Peaks(BED)
Merged dataset profile
Mean profile
Heatmap
Density values
Annotation
Distance
Export selected clusters
Save profile

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
      - log2FC >= 1 or log2FC <= -1
    - All chIPseq data were used
  - Use Venny (http://bioinfogp.cnb.csic.es/tools/venny/)