Mapping and visualization of ChIP-seq data

(answers to questions)

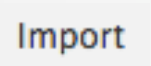
Stéphanie Le Gras
(slegras@igbmc.fr)

Exercise 1: Upload the data in a Genome Browser

- 1.
  - Go to UCSC website (http://genome.ucsc.edu/)
  - Click on My Data > Custom Tracks
  - Clade: Mammals, genome: human, assembly: hg38
  - Browse and select the first wig file and click on upload
  - Once the first dataset is upload, click on Add custom tracks
  - Browse and select the second wig file and click on upload
  - Click on go
  - Change the way data are displayed in the genome browser :
    - Scroll down to the Custom Tracks section
    - Change from dense to full for the two tracks
    - Click on refresh

Exercise 2: mapping statistics

- 2.
  - Click on the button  an select "create new"
  - Click on the history name "Unnamed history", erase "Unnamed history", enter "ChIP-seq data analysis" and press enter
- 3.
  - Click on Shared Data (top menu) and select "Data Libraries"
  - Click on "CNRS training" > "ChIPseq" > "mapping"
  - Select mitf.bam and ctrl.bam datasets (tick boxes beside dataset names)
  - Click on the button 

  - Select history: ChIP-seq data analysis
  - Click on 

  - Go back to the main page by clicking on "Analyzed data" (top menu)

Exercise 2: mapping statistics

- 4
  - Search for “flagstat” in the search field (tool panel)
  - Click on the name of the tool
  - Click on  to select multiple datasets
  - Select all 2 datasets
  - Click on 

Sample name	No. of raw reads	No. of aligned reads
MITF	31,334,257	23,124,393
Ctrl	29,433,042	19,949,607

Exercise 3: duplicate reads estimate

- 1.
  - Search for "markdup" in the search field (tool panel)
  - Click on the name of the tool
  - Click on  to select multiple datasets
  - Select the 2 bam files
  - Select validation stringency: Silent
  - Open the datasets "MarkDuplicates on data : MarkDuplicate metrics"

Sample name	No. of raw reads	No. of aligned reads	No. of duplicate reads
MITF	31,334,257	23,124,393	16,901,318
Ctrl	29,433,042	19,949,607	15,151,227

Exercise 4: Visualization of the data

- 1.
  - ANKRD30BL -> no peak
  - CFAP221 -> no peak
  - DBI -> peak