

Generating a heatmap of GC content for *D. melanogaster* housekeeping genes

This is the verbose description of the video demonstrating how to generate a heatmap using deepTools Galaxy:

http://youtu.be/M_ie-3sxZJE

The generation of heatmaps requires 2 files and 2 tools:

- **bed-file** of genome regions, e.g. genes
- **bigwig file** of scores, e.g. read coverages, GC content...
- deepTools **computeMatrix** will use those files to calculate the scores from the bigwig-file for the regions from the bed-file
- then, deepTools **heatmapper** will use the output of computeMatrix to generate the image

Data import from Data Library

- go to "Shared Data" (top menu) → "Data Libraries" → "Quick Intro Example"
- select the files "Drosophila melanogaster housekeeping genes" and "GC_content_Dm3.bigwig"
- click on "go" next to "import to current history" on the bottom of the page

Analyses

- select "Analyze Data" in the top menu (left hand side, upper corner), this should bring you back to the main frame. on the right hand side, you should now see the two data sets (if you do not see them, click on the refresh icon on the top of the data set panel)

compute Matrix: calculate the GC content for all the genes

- go to the left hand panel and select "deepTools" → "computeMatrix"
- for the *Regions to plot*, select the housekeeping genes (bed-file), for the *score file*, select the GC content file (bigwig-file)
- keep the default parameters except for the following options:
 - "Distance in bp to which all regions are going to be fitted" should be 2000
 - "Set distance up- and downstream of the given regions" should be set to "yes", then replace 1000 with 1500 in the newly appeared boxes
- hit "Execute"

on the right-hand side, you should see a new data set, possibly in grey or yellow, indicating that the analysis is being performed (will take ca. 5 min)

heatmapper: generate the image for the values calculated by computeMatrix

- once the new data set is shown in green, select "deepTools" → "heatmapper" (left-hand side panel)
- use the newly created data set as the sole input file for heatmapper
- hit execute
- there should be another new data set, this time containing a .png file which you can see by clicking on the eye-shaped button

The image will contain a summary plot on top of the heatmap which is the default setting of heatmapper. You will see that the GC content drops dramatically around the start and end of Drosophila housekeeping genes. If you repeat this analysis with mouse data (can be found in "Shared Data" → "Data Libraries" → "Sample Data"), you will find a reverse pattern with increased GC content at the start of genes.