

When the `--kmeans` option is chosen and more than 0 clusters are specified, heatmapper will run the **k-means** clustering algorithm. In this example, we wanted to divide *Drosophila melanogaster* genes into two clusters. As you can see above, the algorithm nicely identified two groups - one with mostly those genes with lots of Pol II at the promoter region (top) from those genes without Pol II at the promoter (bottom).

Please note that the clustering will only work if the initial BED-file used with `computeMatrix` contained only *one* group of genes.

The genes belonging to each cluster can be obtained by via `--outFileSortedRegions` on the command line and "advanced output options in Galaxy". On the command line, this will result in a BED file where the groups are separated by a hash tag. In Galaxy, you will obtain individual data sets per cluster.

To have a better control on the clustering it is recommended to load the matrix raw data into **specialized software like cluster3 or R**. You can obtain the matrix via the option `--outFileNameMatrix` on the command line and by the "advanced output options" in Galaxy. The order of the rows is the same as in the output of the `--outFileSortedRegions` BED file.