

J-Express Pro Practical – Explorative approaches

In this exercise we will be working on two files, the already well known RatBrainProfiling.pro from earlier exercises, her mangler det litt på setningen... trenger man en fil i tillegg til RatBrainProfiling? The exercise will focus on explorative approaches to perform both quality control of your data, and find global biological trends if present.


Preparations

In the same way as a $\log(2)$ transformation makes a much larger range of intensities visible in gene graphs, performing a high level mean normalization of your data will often improve your clustering visualizations.

1. Open the RatBrainProfiling.pro file if you haven't done so already.
2. Select the “Log (2) Quantile” data set.
3. Open “Data Set | Create Sub Data Set”
4. Tick the option “High Level Mean Normalization” and press “Ok”.
5. Notice that you got a new data node. Rename this to “Mean normalized” only.
6. Check the view of your new data set in the Thumbview window.

Hierarchical clustering of samples


A natural first step in order to take a global look of your data set, is to cluster your samples and see how they group together.


1. Open “Methods | Clustering | Hierarchical Clustering With Distance Matrix” (or use the  button) on the data set ”Mean normalized”.
2. Set *Cluster way* to be Columns, use defaults for everything else. Press “Ok” button to proceed.
3. How do the samples cluster together ?
4. Save your project.

Close your window, and we will have a similar global look using another component.

Correspondance Analysis

This analysis give you a plot of both genes and samples in the same window, with genes and samples related to each other plotted in the same direction from the origo. The distance from origo also tells us about how different a sample or a gene is from the “overall average” in the center of the plot.

1. Open “Methods | Projection | Correspondance Analysis” (make sure that the “Mean normalised” node is still high-lighted in the Project window) or alternatively use the  button.
2. Point on some of the spots representing genes furthest away from origo. A tooltip pops up with information on these genes.

3. We'll now zoom in a bit on the samples only. Right click the plot somewhere on the white background.
4. In the Plot properties window, at the lower left, reset the spot size to 0, press "Ok" and then "Close".
5. Resize the window to make it bigger
6. Different things may happen when you select some samples in this window. To zoom in on some samples (square symbols), click on the "Framecontents To PCA"  button before dragging a frame around the samples (click and drag the mouse around the samples you want to zoom in on).
7. Next turn off the tooltip by clicking on "Show tooltip" from the CA menu.
8. Which groups seem to be the most similar to each other from this plot?
9. What effect (sample groups) seems to be driving the global trends we see?
10. The CA menu also contain options to show or hide sample names, as well as some other features. Play with them to find out what they do.


Close your "CA – Mean normalized" window.

Clustering on genes


Until now we have focused on clustering of samples, now we turn to clustering of genes. We will concentrate on two different methods, Self Organising Maps (SOM) and Hierarchical clustering, although J-Express offers other options as well.

Self organising maps – SOM

This is a very powerful clustering method with many nice features. We can use our full dataset with over 26.000 rows representing genes.

1. First close your CA and Group Controller windows, and select the "Mean normalized" dataset in the J-Express Project tree.
2. Click the Self-Organizing Map button  or select "Methods | Clustering | Self-Organizing Map".
3. Select 25 clusters and click *OK*.

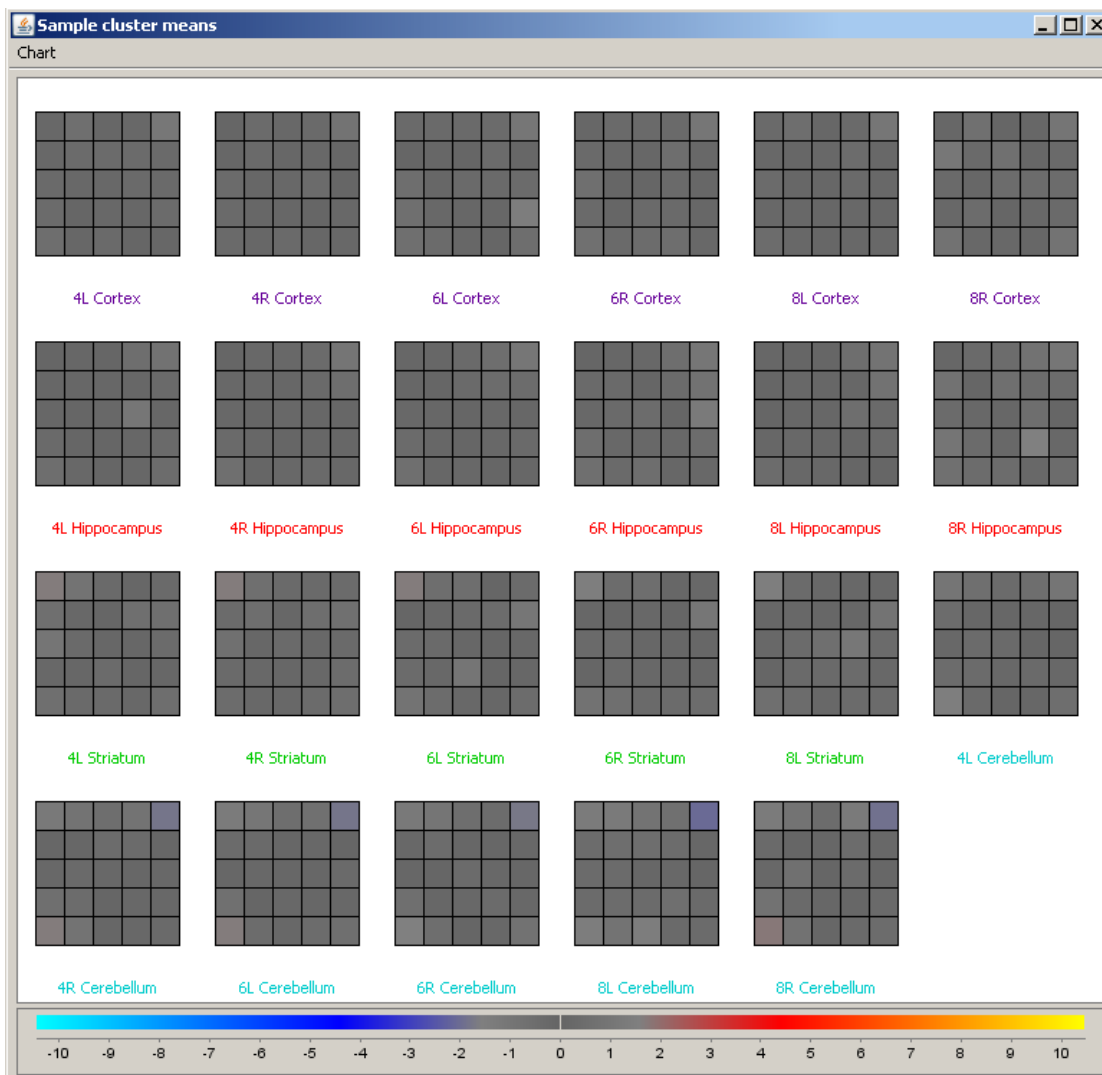
The genes are now divided into clusters according to how similar their gene profiles are.

4. The 5 by 5 table of clusters you now see is ordered so that similar clusters are closer to each other than dissimilar clusters. Can you see this pattern?
5. Click one of the clusters that you think look interesting. This opens a Gene Graph displaying the genes in the particular cluster you selected.
6. Select all rows in the list and click the *Create Group(s)*  button in the Gene Graph window

7. Call the new group SOM and give it a blue color. We will look for this group later on using another method, to see if that method finds the same cluster of genes.

Before exploring a new clustering method, we will take a look at a different representation of all the clusters from an SOM analysis.

8. Back in the SOM window, click on the Thumbnails menu and then click the “All Mean Value Cell View”.




This is a classical SOM view and can be used to see how similar samples are in regard to mean SOM cell expression. Each sub figure displays how a particular sample is expressed in the different clusters


Close the SOM window before proceeding

Hierarchical clustering

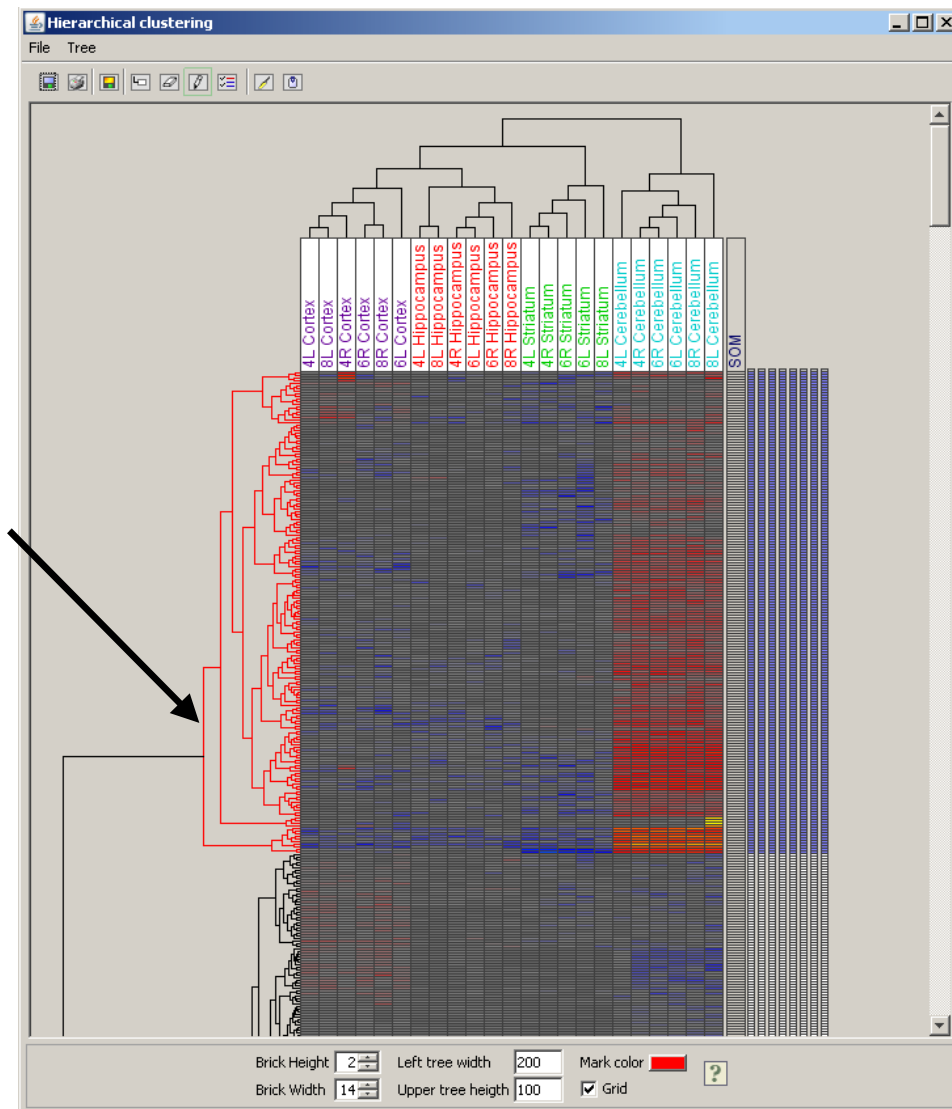
This method is not as efficient as SOM, and will require more memory than we have available to complete an analysis of the full data set with over 26.000 rows/genes. We will therefore use a common trick to trim the data set to a smaller size. When doing clustering, all genes with very little variance, that is having very flat gene profiles, are not giving us much information. We will thus filter these away in a high level filtering step first:


1. Select the “Mean normalized” data set in the Project Window
2. Open the filtering component : “Data Set | Filter Data Set” or use the  button.
3. Click the first method in this window: “Minimum Standard Deviation” and enter 1.0 as the limit. Any genes with less then 1.0 in standard deviation across it’s measurement in the gene profile will be removed.
4. Press the “Try filter” button. Look in the small graphs at the top right (You may have to resize the filtering window). Here you can see which profiles will be removed and which that will remain in the dataset. (Make sure that the “use result charts” option is active below the charts).
5. After testing the filter, press “Create data set” to perform the actual filtering
6. Rename your newly created data set in the Project window to “Minimum std deviation 1.0”
7. Close your filtering window.

This data set is the one we will use onwards in our hierarchical clustering exercise.

1. Select the “Minimum std deviation 1.0” data set in the Project window.
2. Open the Hierarchical clustering using the  button, or select “Methods | Clustering | Hierarchical Clustering”.
3. Press “Ok” to accept the default parameters.
4. First take a look at the dendrogram (clustering tree) of the samples at the top. If you see some of the groups you made earlier in the top of the dendrogram you can turn these off by opening the Group Controler and uncheck the active groups, but keep the SOM group on. The dendrogram will be updated when you click somewhere in the clustering window, making it active again..
5. You can make a selection of the genes in a sub-tree by making the mouse hover over the branch points in the dendrogram. This selection will also be

updated in other components. To see the gene profiles of the selected genes take a look at the Thumb view window or open a Gene Graph and use Shadow unselected.



6. By clicking on a branch point the corresponding sub-tree will be marked in the color set as Mark color. Click the colored square to change the color.
7. Browse the dendrogram of genes. Try to locate a similar cluster to the one you selected as your interesting cluster in your SOM analysis
8. Create a group of the cluster you've decided to compare with your SOM cluster by first clicking the *Create Group of Dendrogram*  button in the *Hierarchical clustering* window and then clicking on the branch point. Click the group color square and select a green color. Call the group Hierarchical. Press "Ok".
9. Open the "Group Controller". How large is the overlap between the groups you created from SOM and Hierarcial clustering?

