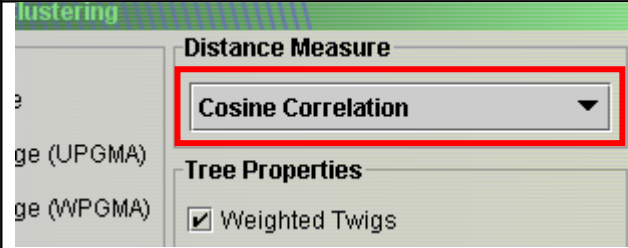
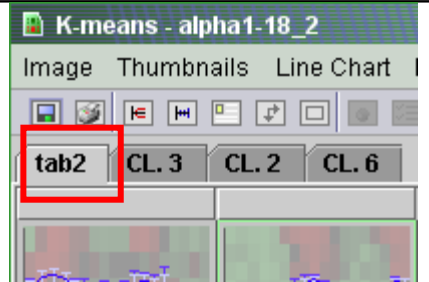


J-Express Practical – the basics

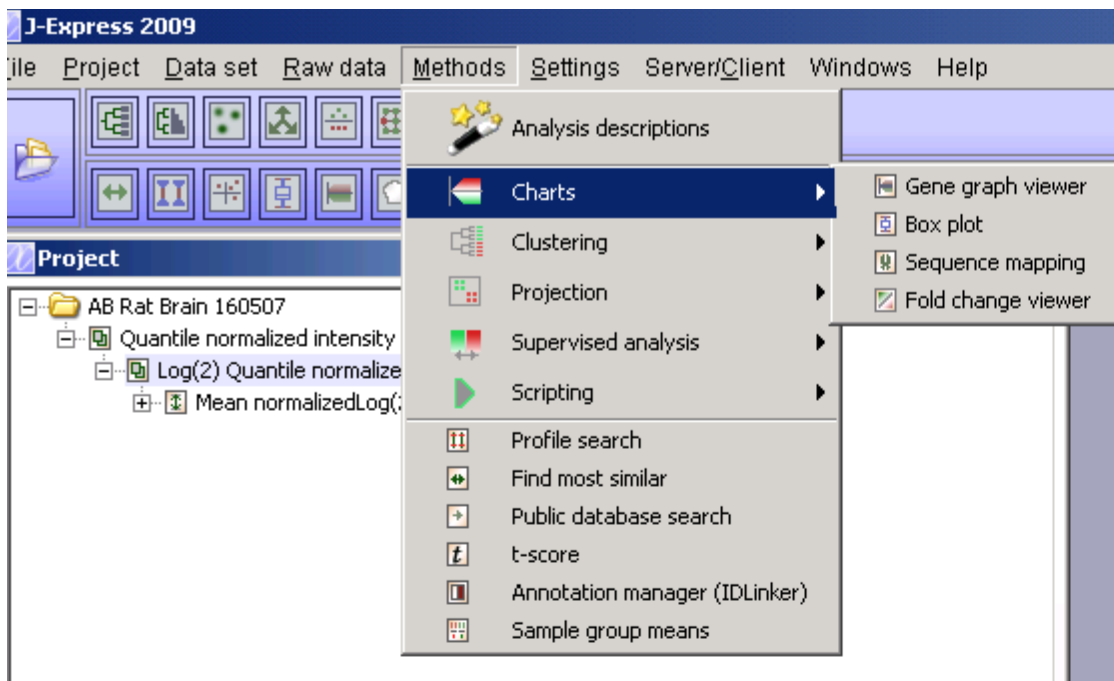
Nomenclature

This first section only explains a bit about the organisation of J-Express and how the exercises refers to different parts of J-Express when describing what you should do in the exercises. The actual exercise starts in the next section.

 <p>This is a Combobox. Click it to change the value.</p>	 <p>This is a tab</p>
--	---

All buttons have a tooltip. Hold the mouse pointer over a button to see what it does.

The notation “Methods | Charts | Fold Change Viewer” in the text means that you should find the “Methods” menu in the J-Express main window’s menu bar, and select the “Charts” item and finally the “Fold Change Viewer” item on this submenu :



Please note that some of the sub windows of J-Express also have a menu bar with drop-down menus. The same notation as above will be used, and in addition specifying which window by the name in the title bar :



Example: Please select “View | Statistical comparison to other GO components” in the “GO DAG Log(2) Quantile normalized intensity data” window.

As you can see from the screenshots above, J-Express have several panels of buttons with icons just below the menu bar at the top of the main window. These are shortcuts to functionality available from the menus. The exercises will give you the option of using either the menus or the buttons.

A sub module of J-Express, typically an analysis method that opens up a new window for interacting with and displaying the data in a new perspective, is commonly called a component.


Getting started

Several files that we will need during the course have been uploaded to the course webpage at the following web-address: <http://www.bioinfo.no/training/mcb-integrative-09/home>

1. Download the J-Express data file CancerPhenotypes.pro to your Desktop.
2. Follow the licence instructions at the course web page to unlock all features of J-Express. (If needed, might be installed already, check if all license components already have been marked green)
3. Start J-Express from the Windows “Start menu | All Programs | J-Express”
4. In J-Express do “File | Load Project”
5. Find your way to the CancerPhenotypes.pro file that you downloaded and press “Open”



Looking at a data set

Using the dataset we've just loaded, we will get more familiar with how J-Express organizes data. In J-Express a dataset is a 2D table of numbers. Depending on how the data have been processed, the numbers might be intensities, ratios or log-ratios. The data we will start out with is normalized intensities.

1. Select the dataset named “MA 8 paired samples quant norm intensities” from the Project window, and read summary information about the data set in the bottom half of the project window.
2. Click the *View Data Set*  button or select “Data Set | View data set” from the main menu bar.
3. Browse a bit around in the opened window, what type of annotation is present, how many data columns does this data set have?
4. And how many genes/rows are there in this data set? (Look around in other windows visible to you as well)
5. Close the data set window (named “Properties for MA 8 paired samples quant norm intensities”).
6. Also inspect the other data set at the root level (“PR 4 paired samples mean norm intensities”) in a similar way.

A second view on the numbers – Search and sort module

Another very useful component to display the actual numbers along with the annotations is the Search and Sort component.



1. Select the MA data set as above, and open the component either by doing “Data Set | Search and Sort” or by clicking the  button.
2. At the bottom of the Search and sort window there is a coloured combo box. Click on the drop down and select the middle colour theme which is red to the right, then changes to orange, yellow, green and white towards the left.
3. Browse the table, what do you think the colours mean?
4. Type in VEGF in the “Search Phrase” area, choose “Advanced options” and select to search “All annotation columns” before pressing the **Search** button. How many genes were found? (You should be able to see this without counting the number of rows)
To search only specific annotation columns select “Columns (comma delimited)” and specify which column(s) to search in the text field behind. Select to only search column 3 (Gene Symbol) before pressing **Search**. How many genes were found now?
5. *In order to get back to search the whole dataset click the  button.*
6. Let's now turn to sorting. By clicking on the header of a column in the “Search and Sort” table, the whole table will be sorted according to this column. Clicking the label again switches the direction of the sorting. Try it out a bit on different columns.

7. Sort the dataset by the Gene Symbol column, and see if you easily can find any genes starting with the letters “ADSL”
8. Now choose to sort on the column for the “Pat1Type1” sample. Make the highest numbers appear at the top, and scroll all the way up to see what it is. And now you’ll also see how the colour shaded background behind the values work :)
9. Reverse the sorting order of the “Pat1Type1” column, and find the lowest value as well (or alternatively scroll all the way to the bottom)


Keep your “Search and Sort” window open for the next exercise as well.

Graphical view


An image tells more than a thousand numbers, so displaying gene expression levels across many samples as a graphical gene graph is very intuitive.

1. Make sure that the node “MA 8 paired samples quant norm intensities” is selected in the Project tree.
2. Open “Methods | Charts | Gene Graph Viewer” or use the  button.
3. Rescale window size by dragging window borders.
4. Rescale the column width to see some of the annotation better
5. Click on the **Shadow unselected**  button in the Gene Graph window
6. Select a single gene by clicking on a row, and look for its graph in the right part of the window. The scale on the y-axis is decided by the genes with the highest expression. This result in some of the genes with lower expression being squashed down, and may not be that visible. Try some other ones.
7. Select more than one gene by holding down the “Ctrl” key and clicking several rows at random.
8. Try selecting the top 30 genes or so (there are at least two standard “windows/ excel” like ways of doing this)
9. Rearrange your windows so that you see both the “Search and Sort” and the “Gene Graph” window.
10. Find the gene with GENE_ID hCG2040304 using the “Search and Sort” window. Note that the Gene Graph is updated showing the profile of the gene searched for in search and sort.
11. What is the approximate difference in gene expression between the Type1 and Type2 samples of each of the 4 different patients? (First: What do you think the values on the y-axis represent? Don’t peek before you’ve thought hard: linear intensities, the alternative is log transformed values, and they would be much lower values, we’ll see more of that later).
12. In which cancer type is this gene most upregulated? Instead of interpreting the sign of a score for which there are no clear definitive rules of what the signs mean, the gene graph always give you a very clear and intuitive answer instead.

Project tree window

In the beginning of the exercise, you were told to open a specific data node in the project tree window, and opened several modules to work on this dataset. If a node has the icon  in front of it, the node has children nodes below it in the tree. These children nodes represent either 1) a subset of the parent data 2) a processed dataset starting from the parent data set.

We will now create a children node in the project tree:

1. Click the  icon to unfold the tree and see any children nodes.
2. Unfold further until you find the “MA log 2 ratios (per pair)” data set and select this.
3. Open the “Search and Sort” module on this data set.
4. Find all genes annotated with “ZINC FINGER PROTEIN” as Panther Annotation and mark all found genes.
5. Click the “Branch selection” button after you have found the genes. Remember the icon on this button, this appears many places in J-Express and will allow you to create a new data set of the current selection of genes or samples.
6. Note that a new data set node appeared in the Project window.
7. Rename the new data set to “Zinc finger proteins” by first selecting it with one click, then do a second click on the name (note: a double-click will not work) and type in the new name, and press enter.
8. Save your project.



Color tables

Some components in J-Express are using colors to visualize the data instead of showing the real numbers. By using colors it is much easier for the eye to spot interesting patterns. *We have already played with color tables in the search and sort window, but that color table was only specific to that particular window. We will now look at this in a more global context to J-Express.*


1. At the bottom left of the J-Express desktop you should find a drop down containing the “Default color table”. Select a different color theme from the drop down and then mouse over the Thumb view window. The color should now change to the new theme.
2. Play with different colors. Notice that some of the color themes have a white line in the middle and only use parts of the scale, while others use the entire scale. Why do you think that is? *Hint: look at the scale of the y-axis in the Thumb view window.*

Selections

Understanding how selections work in J-Express “is the key to everything”, so we will stress that point once more.

3. Open “Methods | Charts | Gene Graph Viewer” (or use the  button) on the data set called “PR 4 paired samples mean norm log intensities”.
4. Rescale window size by dragging window borders.
5. Rescale the column width to see some of the annotation better
6. Click on the **Shadow unselected**  button in the Gene Graph window
7. Select a single gene by clicking on a row, and look for its graph in the right part of the window
8. Try some other genes. As you might remember from the Basics exercise, it could be hard to find genes with a visible difference above the bottom of the graph area. In $\log(2)$ transformed data, genes/proteins from the whole range of different intensities are better displayed, but keep in mind that the y-scale now is $\log(2)$ based.
9. Keep your “Gene Graph” window open before advancing to the next steps.


A very important aspect of selections in J-Express is that a selection done in one window will be highlighted in every other window displaying the same data set from the project tree. To illustrate this we will open another component Hierarchical Clustering on the same data set:

13. Open “Methods | Clustering | Hierarchical Clustering” (make sure that the same data set is still high-lighted in the Project window) or alternatively use the  button.
14. Select “Average Linkage” and “Pearson correlation” as options, leave the rest at default values, and click “Ok”.
15. A new window with the clustering heat map and dendrograms (clustering trees) will appear, readjust the windows size if necessary, and position it so that the “Gene Graph” window also is visible.
16. Tick off the “grid” option, and increase the brick height to 3.
17. Hoover over some of the subtrees in the dendrogram to the left of the heatmap (point at the junctions: \dashv) and observe the change in the “Gene Graph” window. Try to find a subtree with some interesting profiles displayed in the “Gene Graph” window.

Branching sub data sets


Whenever you have made a selection somewhere in J-Express, you can branch out the selected genes/samples in a separate sub data set. This will appear in the Project window as sub nodes of the parent data set it was extracted from. Many components will have a separate “Branch” button for this, since it’s a very common operation from many analysis components, as the one you used earlier in Search and Sort.

In addition there is a special component called “Create groups” that allows you to branch the active selection of a data set, regardless of which component that this selection was made in (and regardless of whether that component has a **Branch** button or not). Since “Hierarchical Clustering” does not have this button, we will try to use the Create Groups component instead to branch of a copy of the cluster/subtree you found interesting above.

6. Make sure “PR 4 paired samples mean norm intensities” still is selected in the Project window.
7. Open the “Data Set | Create Groups” component, or press the  button.
8. You should see “Selected: 189” or similar in the lower left corner of the Grouping window appearing (something different than “Selected: 0”). If not, close the window, make sure that the cluster is selected in the “Hierarchical Clustering” window (and visible in your Gene Graph window), select the proper data set (“PR...”) in the Project window, and open the “Create Groups” component again.
What happens if a different set was highlighted before opening the “Create Groups” component and why?
9. Branch your selected genes by pressing the **Branch** button in the “Create Groups” component (the title bar of the window reads: “Grouping - ...”).
10. Move to the Project window, locate your new data set and check how many rows it has.
11. Rename your new data to something like “Interesting cluster” or similar.
12. Close the “Grouping - ...” window.

Merging data sets

Sometimes it is convenient to merge several data sets into one larger data set. Be aware that this put restrictions on what type of analysis you can do on this merged data set later. For later use in the course, we will need a merged data set of the MA and PR log ratio data sets, where corresponding genes and proteins are aligned with their data values in the same row.

1. Open the “Merge datasets” component: “Data Set | Merge | Merge datasets”
2. A new window appears with two tabs. Resize the window if needed to see the whole content.
3. In the first tab we are asked to specify the data sets to be merged. In the Project window, first highlight the “MA log 2 ratios (per pair) uniq genes” data set, and then in the “Merge datasets” window, click the  button to import this as the first data set under “Merge source 1”.
4. Then highlight the “PR log 2 ratios (per pair) uniq proteins” data set and import this as the second data set in the same manner.
5. The “Data set identifier” drop-down box specifies which annotation column for the datasets should be used to map the relationship between the rows of the two data sets. Change this to Gene Symbol for both data sets.
6. Click the “Next” button at the bottom, or select the “Step 2...” tab at the top.

7. In the second tab we will specify some options for the merging: make sure “Allow genes not present in all data sets” is selected, and unselect the “Quantile normalize all merged data” option.
8. Then click the “Create dataset” button at the bottom. A new dataset will now appear in the Project window at the root level.
9. Rename the new data set to “Merged MA and PR data”. This is the data set we want to visualise on a KEGG pathway at the end of the module.
10. Save your project: “File | Save project”.