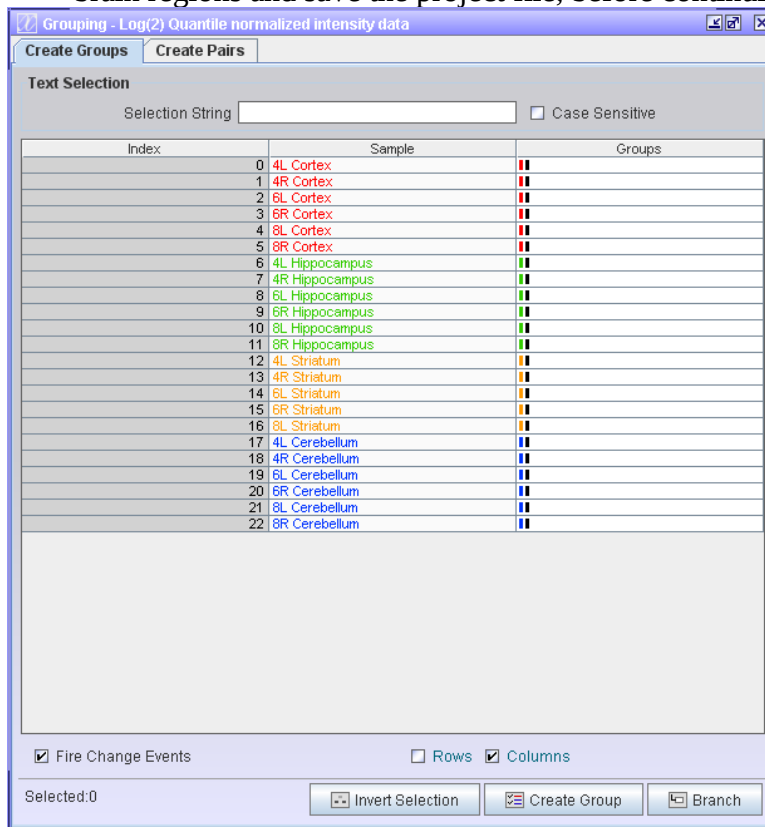



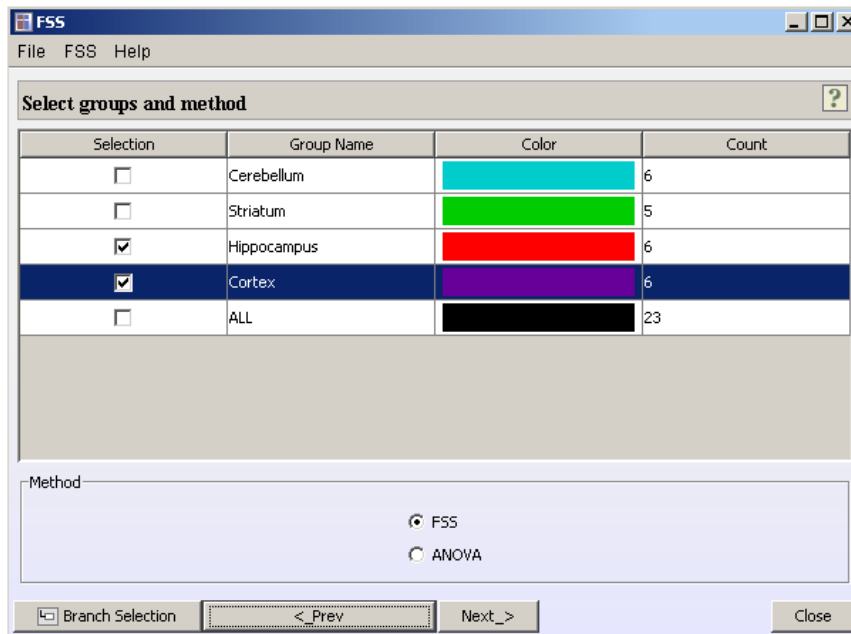
J-Express Pro Practical – Differential Expression

Two sample t-test

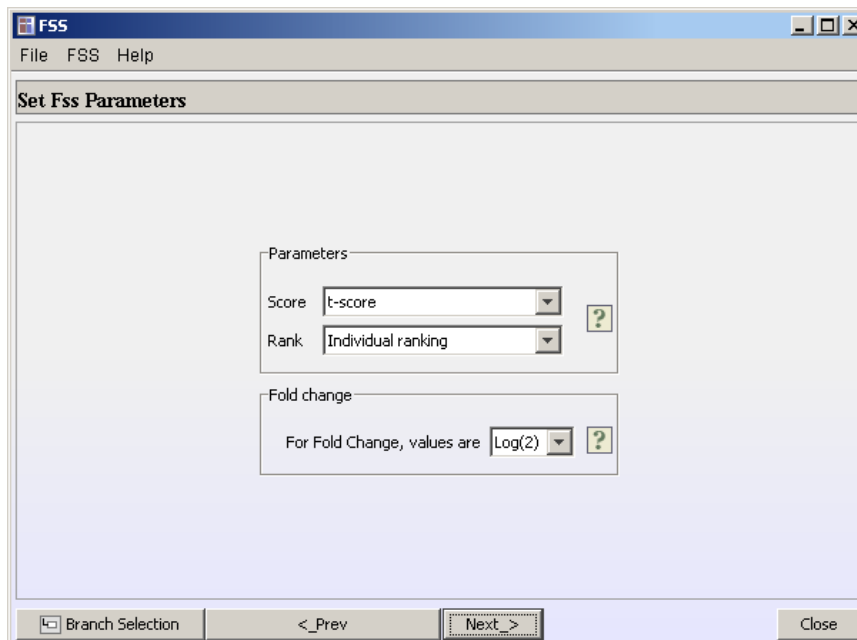
1. Open the “RatBrainProfiling” project file.
2. Select “Log(2) Quantile normalized intensity data” node from the project tree.
3. To find differentially expressed genes between two sample groups we must first define the sample groups. If you saved the file after creating groups yesterday you can continue to the next step. If not, create groups based on the different brain regions and save the project file, before continuing.




4. Make sure the correct data set is selected and click the *Feature Subset Selection* button  or select Supervised analysis | *Feature Subset Selection/ANOVA* from the Methods menu
5. Select the two groups you want to compare.

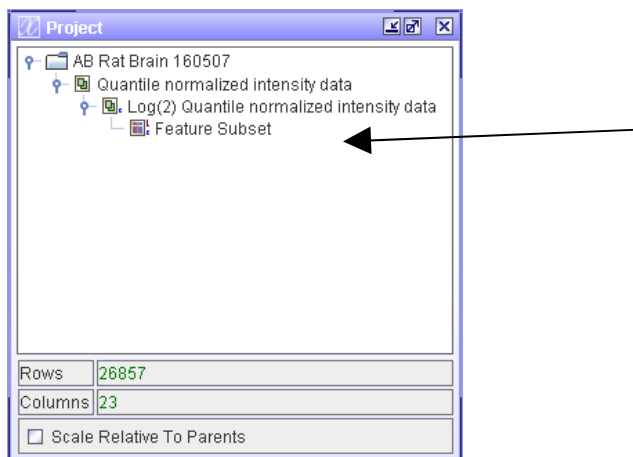


6. Have the FSS method selected and click next



7. Have the t-score and individual ranking selected and click next
8. Open a Gene Graph and click the *Shadow Unselected* button 
9. Move the FSS window and the Gene Graph window so you can see both
10. In the FSS window: The table is sorted according to the Score column.

11. Select some of the rows at the top of the FSS table, and look at the scatter plot. Are the two groups (spots with different colours) well separated? You can see the names of the genes by moving the divider between the plot and the table and by re-sizing the columns (click and hold the column header between two columns).
12. Look at the *Gene Graph* window. If you consider only the groups you selected for this analysis, for which groups are the selected genes up or down regulated?
13. Click on the header of the Score column. The column will now be resorted. Select some of the genes at the top. Again look at the scatter plot and the Gene Graph window. How does these genes compare to the ones you had selected in exercise number 11 and 12?
14. Sort the genes according to Fold change values. What is the difference between sorting the genes according to Scores and sorting them according to Fold change?
15. Press the <Prev button, and select *P-Values* as a score. Press **Next** again.
16. The P-Values are based on the t-test. Select some genes at the top of the table and look at the gene graph. How does this result compare to using t-score? Why do you think the results looks different? (HINT: look at the sign of the numbers in the score column)
17. In the FSS window, select Save Table from the File menu, to save the result to a tab-separated text file. The .txt extension is not added automatically, so find a suitable name for you file and type .txt at the end
18. Make sure the genes are sorted according to Score so the genes with the smallest p-values appear at the top of the ranked list. Select the top 200 up-regulated genes. (To see how many genes you have selected you may have to use the Create Group window, which always shows the number of genes selected on a particular dataset.) If you get a warning click cancel. Click the *Branch Selection* button. In the Project window you will now see a new data set called “Feature Subset”




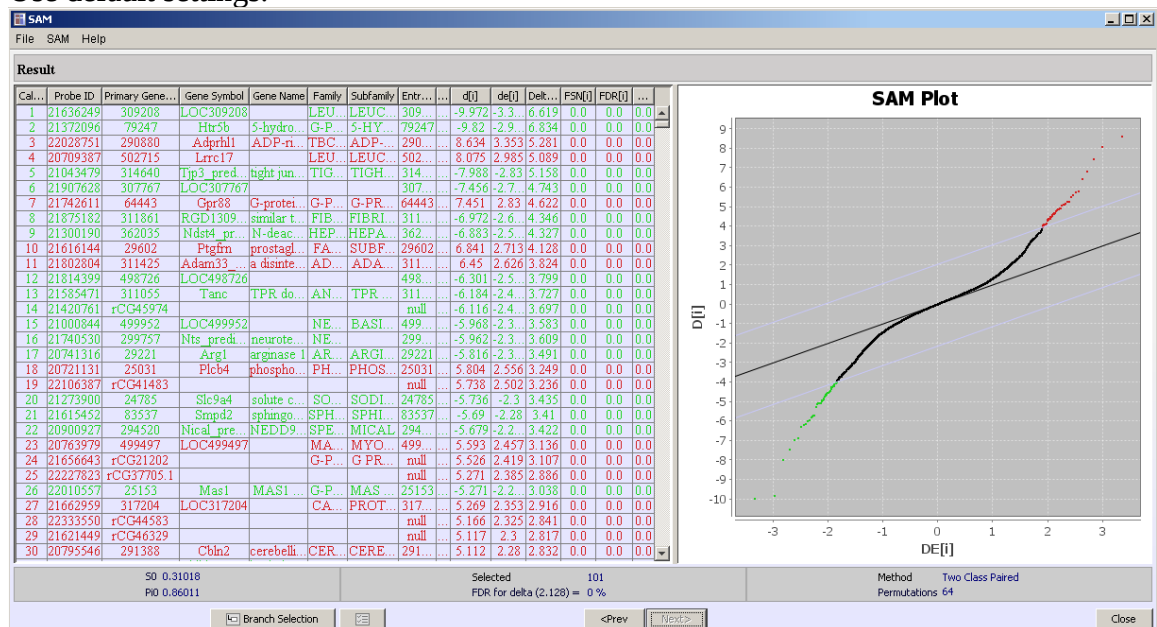
You have now created a subset of the data most of the genes are differentially expressed between two brain regions.

19. Close the FSS window

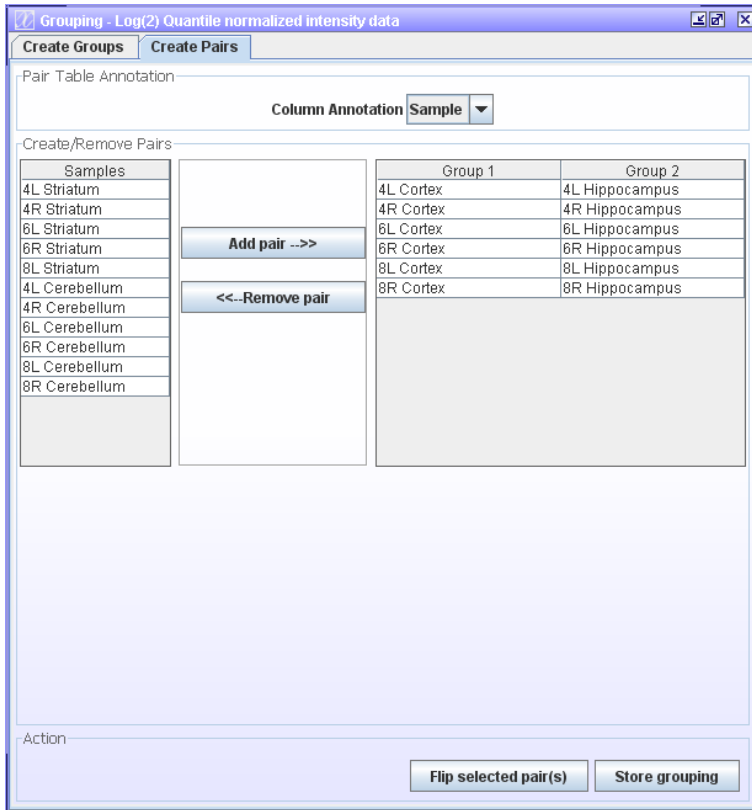
20. Save the project from the J-Express *File* menu

Significance Analysis of Microarrays (SAM)

21. We will now do a similar analysis to the one we just did by using SAM. Make sure you have selected the dataset called “Log(2) Quantile normalized intensity data”, click the Significance Analysis of Microarrays button () on the toolbar or choose Supervised analysis | *Significance analysis of microarrays* from the Methods menu.
22. Do an unpaired analysis of the same two sample groups you compared by FSS. Use default settings.



23. The most important columns to look at in the result table are $q\text{-val}[i]$ and Fold change. It is still good to know what the other columns mean. Press F1 for help on what the other columns mean.
24. What do you think the green and the red colours in the table mean? (HINT: Look at the gene profiles in the Gene Graph.)
25. These columns cannot be sorted. Select the top 200 genes and branch them off. You will now get a new node in the project tree called SAM.
26. Make sure the data set “Log(2) Quantile normalized intensity data” is selected again. It is also possible to do a paired analysis in SAM. Close the first SAM window and open a new one. Select the **Paired** tab.
27. If you created pairs and saved the project file yesterday you can proceed to the next step. If not, create pairs of Cortex and Hippocampus samples from the same rat before continuing.



28. Use default settings for permutations and fold change and click *next*. You should now be familiar with the result table.
29. There are different ways of saving the results from the SAM analysis. We have already looked at how you can branch off some interesting genes to a separate dataset. We will now look at a couple other ways of saving the results: Saving the table to a text file and storing the entire analysis in the project tree.
30. In the SAM window, select Save Table from the File menu, and name the file “results_paired_sam.txt”. This saves the entire table to a tab delimited text file.
31. To save the entire analysis, select Put in project tree from the SAM menu. The analysis will now be available in the project tree with the name SAM Result. It is not a node that contains a normal data set, but you can double click this node to reopen the analysis window.
32. Save the project from the J-Express *File* menu

You have now done a paired and an unpaired analysis using SAM, and looked at different ways of storing the results.

33. In the project tree you can now select the SAM node containing the 200 genes you branched off, and look at the Thumb view window. The thumb view window gives you a quick view of all the genes in a dataset. This window updates as different datasets are selected. Select the Feature Subset dataset from the project

tree and take another look in the thumb view window. Does it look like SAM and T-test ranked the same genes as the top 200 differentially expressed genes?

Rank Product

34. We are now going to analyse the data using yet another method: Rank Product. Select the node in the project tree called “Log(2) Quantile normalized intensity data”.
35. Select Supervised analysis | **Rank Product** from the J-Express Methods menu.
36. Analyse the data by doing unpaired analysis, set the number of permutations to 100.
37. The result table is sorted according to the Pos score column. As you know from the lecture, a good score from Rank Product is a small value. If you select some genes and then look at the Gene Graph again, what can you say about the genes that have a good positive score or a good negative score? (The different genes can be sorted according to a different column by clicking on the column headers, you only need to sort the genes according to Pos score or Neg score.)

There are different ways of saving results from Rank Product as well, so we will now look at how you can do this.

38. First of all you can save the entire analysis by selecting “Store in project” from the Results menu.
39. Sort the table according to Pos score. See that you have the smallest numbers towards the top and that the q-values are close to 0.
40. Select some of the genes listed at the top and click the branch button at the bottom of the window. A new node called Rank Product will appear in the project tree. Notice that it is also possible to create a group of these genes.
41. Sort the table according to Neg score and branch off some of the best scoring genes.
42. There is no functionality for saving the entire table to a text file, but if we wish to do this we can select a row in the table, then press Ctrl – A to select all, Ctrl – C to copy and Ctrl – V to paste in notepad or other text editor. *Note that by copying the results this way, the headers will not be exported.*

The results in J-Express are sorted according to the Pos Score. The genes with good positive score are listed towards the top. The genes with good negative score are listed towards the bottom of the list. When the list is sorted according to Pos Score, the order of the genes with good negative scores may not be optimal, so to get the genes with good negative scores, we have to sort the genes according to the Neg Score column.

43. What is the most significant gene found? What is the q-value of this gene?

Questions:

1. What are the main differences between T-test, SAM and Rank Product?
2. Which statistical value is used to say something about significance in
 - a. T-test ?
 - b. SAM ?
 - c. Rank Product ?
3. Describe in your own words how you understand the different statistical values.