




J-Express Pro Practical – Analyse Gene Sets

This practical completely focuses of one module in J-Express, the Gene Set Enrichment Analysis component. Although we will use the GO component to define sets of related genes to analyze.


Default GSEA

1. Download and open the project file called “RatBrainCortex.pro”.
2. Select the data set “Cortex” in the Project window
3. Open “Methods | Supervised Analysis | Gene Set Enrichment Analysis”.
4. You will now be asked if you want to collapse probes to genes. Some genes may be represented by more than one probe on the array. Analysing the data without collapsing the probes can therefore introduce some bias in the statistics. After collapsing probes to genes, we only have one entry for each gene in the data-set. The column in the data-set that specifies which genes the different probes map to is called Primary Gene ID. Set **Gene info column to Primary Gene ID**.
5. When collapsing probes we also create a new profile for the gene. Set the collapse mode to **Max probe**. Max probe means that the probe that have the highest value of all the probes belonging to the same gene is selected to represent the gene from a particular sample. Click **Next**. Notice that a new node called Collapsed to Genes has now been added and selected in the Project tree.
6. Before we continue with GSEA, is there any sort of global analysis you can do on this data-set that can give you an idea about which groups of samples have bigger contrasts? Use this to find out which groups are more different and use these two groups for the GSEA exercise.
7. We will first do an unpaired analysis. Select the two groups to be compared.
8. Leave all other parameters to defaults, and press **Next**
9. Select to use a GO Tree as input gene sets and click on the button  **map dataset to a GO tree**. This will open a GO DAG. Repeat the mapping from the GO exercise.
10. Back to the GSEA window: Use the default minimum and maximum number of genes, and press **Run**
11. First the small and large gene sets are filtered and then a window will pop up letting you know how many gene sets it found within the right size limits. Click Ok.
12. Click on the top gene set and examine the table and plot.
13. Open a “Gene Graph” to see the genes belonging to the selected gene set (remember to also click “Shadow unselected”). You may have to select the gene set again to get it to refresh in the gene graph.
14. Move GSEA, GO Tree, and Gene Graph windows so that you can see them all.
15. **GSEA:** Click some other interesting gene sets and look at the updates in GO Tree and Gene Graph.
16. **GSEA:** By default the “All” button is pressed in the Mark Selection area, try the “Leading Edge” button instead, and click around on different gene sets

again. You should now see less gene profiles. How many genes do you see, and are they looking better/more consistent?

17. Branch off one or more interesting gene sets by selecting them and then clicking the **Branch** button 
18. Save the GSEA component for later use by putting it in the project tree by clicking the **Store result in project tree** button 
19. Change the name of the GSEA component to “GSEA perm_samples”
20. Close GSEA and GO and Gene Graph windows

Weighted scoring

1. Make sure the “Collapsed to Genes” node is selected and open a new GSEA window.
1. Select the same groups as for the first analysis.
2. In the scoring function, change the weighting to **Weighted 1.5**
3. Keep all other settings the same as before
4. Compare these FDR values to the ones you got in the previous analysis. To do this right-click the GSEA component that you saved in the project tree earlier and select Open (or double-click on it).
5. Are the FDR values the same?
6. Save the GSEA component for later use by putting it in the project tree by clicking the **Store result in project tree** button 
7. Change the name of the GSEA component to “GSEA weighted 1.5”
8. Close the GSEA and GO windows.

Changing different parameters

1. You now know how to run GSEA. As you have seen choosing different parameters for the analysis can affect the results. We will now play with some more and for each case try to explain how it affects the result:
 1. Use gene permutations instead of sample permutations
 2. Change the scoring function to SAM
 3. Also try doing a paired SAM analysis

Use external gene set file, use larger groups

1. Download the [c2.all.v2.5.symbols.gmt](#) file from the course homepage.
2. Do a GSEA analysis using the newly downloaded gene sets. In addition to or instead of using the GO tree as a basis for gene sets, select **File** as Gene Set Source and locate the .gmt file that we just down loaded.
3. Set Data Identifier column to Gene Symbol
4. Use the Gene Set Filters to control the size of the groups that are used.
5. Examine the results.
6. The genesets you just used were downloaded from the online resource MsigDB. **Optional:** Go to <http://www.broad.mit.edu/gsea/msigdb/> and download another set of gene sets (.gmt file) and run GSEA with this file. You have to register to use this resource.
7. Close the GSEA window.