

## J-Express Pro 2009 Exercise – Gene Ontology Analysis

### ***Before starting***

You should update a couple of files before you start. These are:

- OBO v.1.0 describing the relation between the GO terms. A new version of this file is released daily by the GO consortium
- A species specific file associating the different genes for an organism to the GO terms


Both these files can be downloaded from <http://www.geneontology.org>. If more than 3 months have past since last time you downloaded these files, you may want to consider updating them.

For this course we have put the files you need on the course home page.

- Download the file called gene\_ontology.obo.zip and unzip it to the: [<J-Express Settings>\resources\GO](#) folder. Overwrite the old file. **Important:** It is important that there is only one .obo file present in this folder. If the files have different names so that the old file is not overwritten, you may risk that J-Express reads the old file instead of the new one.
- Download the file called gene\_association.rgd.zip and unzip it to the [<J-Express Settings>\resources\GO\goassociations](#) folder.

### ***The Exercises:***

#### ***Mapping of terms to your data set***

1. Open the “RatBrainProfiling”.pro file
2. Again select the node in the project tree called “Log(2) Quantile normalized intensity data” and click on the *Gene Ontology Mapping* () button or select it from the “Methods | Supervised analysis” menu
3. Select **gene\_association.rgd** as the *Mapping File*, and set *Data Identifier Column* to **Gene Symbol**.
4. **Uncheck** the **Use Synonyms** checkbox and click **Map DataSet**. What do the colours of the numbers in the GO tree mean?

*We will first explore this window manually a bit. Later, in the next section with heading **Overrepresentation**, this GO tree will be used as a reference to calculate p-values for the different GO groups. So keep it open.*

### *Manual exploration*

5. In the top panel, you see the Gene Ontology represented as a tree like structure, with the three top level terms represented. Open up each of the three different top level terms, one at the time, to inspect the next level of terms.
6. Select a term that looks interesting by name, and read the definition of the term in the lower right panel of the window. Repeat with some other terms.
7. Note that when a GO term is selected, the member genes (genes annotated with this term) appear as a table in the lower left panel of the window.
8. Open a “Gene Graph” window on the same “Log(2)...” dataset, and click “Shadow unselected”. Select a new term in the GO tree browser of the GO window, so that you then can view the gene profiles of the member genes.
9. Dive as far as you can into one of the sub terms of “biological process” opening deeper and deeper levels, too see how many levels you can find.
10. If you have a key word you think would be part of a GO term, you can search using the Locate text area in the middle of the GO window. Try to enter “.\*apoptosis.\*” and press enter. The first GO term containing the word “apoptosis” would then be selected.
11. Press enter again to find the next hit
12. Continue pressing Enter until you find the term called exactly/only “apoptosis”
13. Investigate the neighbourhood of this term (parent, sibling and children terms)
14. Close your Gene Graph window before continuing, but keep the GO window open.

### *Overrepresentation analysis*

*We now move on to the actual overrepresentation analysis, to find the GO terms that have most genes differentially expressed compared to what is expected by chance.*

15. In the project tree, select the SAM node from the Differential Expression exercise done earlier. Open another Gene Ontology Mapping and repeat the above steps (**Mapping of terms...** section) for this dataset.
16. Open a **Gene Graph** and click the **Shadow unselected** button.
17. Move the two GO DAG windows and the Gene Graph window so that you can see them all.
18. In the **GO DAG** of the SAM dataset, select **Statistical comparison to other GO components** from the **View** menu. *A popup window will state: "Comparing to open tree GO DAG Log(2) Quantile normalized intensity data" Click Ok.*
19. A new window now opens called: **GO Comparison result**. The first column contains the names of the different GO terms, the second column displays the p-values, the third and fourth columns shows the number of genes belonging to a particular GO term that was found in the test dataset (here genelist from SAM) and the Reference dataset respectively, i.e. the two GO DAG windows. Click one of the rows towards the top of the result list.

20. Look at the two **GO DAGs** and the Gene Graph window. The selected term will be selected in all these windows. If the number of genes of that particular go term exceeds the maximum number of members in the GO DAG of the SAM dataset, 100 by default, the gene profiles will not be drawn in the Gene Graph. The maximum number of members can be changed in the GO DAG window. You can also see that p-values have now been added to the nodes in the tree of the GO DAG of the SAM dataset.
21. Play around with a different GO terms from the **GO Comparison result** window.
22. Some genes may belong to several gene ontology terms. If some of the gene ontologies that are overrepresented in the gene list overlap, than this may help identify a smaller group of genes that are particularly interesting in the gene list. Create groups of some of the genes belonging to interesting groups. (Hint: Grouping component)
23. Use the group controller to find out whether some of the groups you created overlap. You can also easily check this by opening a Gene Graph window, sort on the column “Groups”, and see if some of the genes are found in both groups.