

J-Express Pro Practical – Differential Expression

In this exercise we will first look at two different methods for sorting all genes in a data set according to a differential expression measure: t-test and Rank Product. After comparing these a bit, we will move on to a very simple but intuitive way to assess if an a priori defined set of genes are showing trends of differential expression.

Both of these concepts will be needed as foundation before we get to the topic of integrating different datasets later on.

One sided t-test

1. Open the “CancerPhenotypes.pro” project file.
2. Select the “MA log 2 ratios (per pair)” node from the project tree.
3. Open the “t-score” module: “Methods | t-score”
4. Open a “Gene Graph” window as well, and click “Shadow unselected”.
5. Select the top 10 genes or so, and inspect them in the Gene Graph window.
6. Browse down to the bottom of the list, and select the 10 or so genes with highest positive t-score.
7. Keep the window open for comparison with Rank Product method below.

Rank Product

8. We are now going to analyse the data using yet another method: Rank Product. Select the the same node in the project tree as above (“MA log 2 ratios (per pair”).
9. Select **Rank Product** from the J-Express **Methods → Supervised analysis** menu.
10. We will do a one class analysis, since each measurement is a log ratio between type1 and type2 samples. Select the single group named ALL, telling the method to do a one class analysis on all the samples in the data set, and press “Next”
11. Leave all parameters at default value and press “Run”. The computation might take some time before the results appear in a new window.
12. The Rank Product method computes both positive scores and negative scores for upregulated and downregulated genes respectively. Resize the window to see the different columns better.
13. Investigate some of the top genes by clicking on them, and look at their profile in the “Gene Graph” window you opened earlier.
14. Click on the “Neg Score” header to have the table resorted and look at some downregulated genes as well, now appearing at the top of the list.

15. Resort on the “Pos Score” column, and select the top 10 genes. How does these compare to the t-score top 10 genes? You may alternate between selecting the top 10 genes in the two different method windows to see how the gene profiles change in the “Gene graph” window.
16. Select all genes from the top with a q-value<10% (=0.10), and branch this off in a new data set, and rename it to “Top upregulated Rank Product q<10%”.

Gene Ontology Analysis

Before starting

Before starting a GO analysis, it's important to be aware of which annotation files one are using. The used files are:

- OBO v.1.0 file describing the relation between the GO terms. A daily updated version of this file is available on the net.
- A species specific file associating the different genes for an organism to the GO terms

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 exercise we have already downloaded the association file for the human genes (gene_association.goa_human maintained by EBI):

1. Open a web browser and download the gene_association.goa_human file from the course page (<http://www.bioinfo.no/training/mcb-integrative-09/home>).
2. Move/Save the file into the <jexpressuserdir>/resources/GO/goassociations folder.

<jexpressuserdir> is the directory that J-Express will look for annotation files and other user specific data files. For this course, that directory is: <C:\Program Files\JExpress>


General procedure from <http://www.geneontology.org>: (Not necessary to be done now in the course due to bandwidth limitations, but you could navigate the site a bit to get familiarised with it.)

1. Locate the Ontology files under Downloads and Ontology file downloads. Download the file called OBO v1.0 by right-clicking the link and “save link as”. This file should be saved to <jexpressuserdir>/resources/go
2. **It is important that the existing gene ontology file in the folder named “gene-ontology.obo” is deleted or renamed, and that the file name on the gene**

ontology file you just downloaded is changed to “gene-ontology.obo”. If not, the old version of the gene ontology file will be read by the software!

3. Click on the Annotations link under Download and find *Your Species*, in this case human. Download and unzip the file. Save the unzipped annotation file to <jexpressuserdir>/resources/go/goassociations.

The Exercises:

1. Open the “CancerPhenotype.pro” file if you do not have it open already.
2. Again select the node in the project tree called “MA log 2 ratios (per pair)” and click on the *Gene Ontology Mapping* () button or select it from the “Methods | Supervised analysis” menu
3. Select **gene_association.goa_human** as the *Mapping File*, and set *Data Identifier Column* to **Gene Symbol**.
4. **Uncheck** the **Use Synonyms** checkbox and click **Map DataSet**. What does the coloured numbers in the GO tree mean?

This GO tree will now be used as a reference to calculate p-values for the different GO terms.

1. In the project tree, select the “Top upregulated Rank Product q<10%” data set. Open another Gene Ontology Mapping and again select **gene_association.goa_hyma** as the *Mapping File*, and set *Data Identifier Column* to **Gene Symbol**
2. Open a **Gene Graph** and click the **Shadow unselected** button.
3. In the **GO DAG of Top upregulated Rank Product q<10%**, select **Statistical comparison to other GO components** from the **View** menu. A *popup window will state: “Comparing to open tree GO DAG MA log 2 ratios (per pair)”* Click **Ok**.
4. A new window now opens called: **GO Comparison result**. Click the GO term column at the top of the result list.
5. Look in the **GO DAG Top upregulated Rank Product q<10%** window. The selected term will be selected in this GO tree. (Unless the number of genes in that particular go term exceeds the maximum number of members, 100 by Default, but you can change this in the GO DAG window.) You can also see that p-values have now been added to the nodes in the tree.
6. Look in the gene graph. This will now show the genes belonging to the selected GO term.
7. Try some different GO terms from the **GO Comparison result** window, and investigate nearby terms in the GO DAG tree as well.