


## J-Express Practical – Rank Product Meta analysis

In this final J-Express practical we will identify which Pathways that seems to have support from both the PR and the MA dataset, and you will later use the identified pathways to visualize both data sets in the coming ProMeTra sessions.

### Import ranks as data sets



First we need to import the ranked results for the Pathway analysis done separately on the MA and PR datasets, as data in their own right. We will use J-Express' "Load tabular data" functionality to achieve this.

1. Open "File | Load tabular data"
2. Click the "Choose file" and open the "EnrichedPathwaysInMAupregulated.txt" file you saved earlier.
3. Click the "Select row annotation" button
4. Click in the following columns so they get highlighted: Gene Set, Size, Nom P-value, FDR (%)
5. Next, click the "Set row annotation header" button
6. Click in the first row so that it gets highlighted
7. Next, click the "Set sample annotation" button
8. Click in the top left cell reading "Rank"
9. Next, click the "Select data" button
10. Select the first cell below the "Rank" column header.
11. Browse down to the bottom of spreadsheet, and press "Shift" and click the last cell of the first column ("Rank") that contains data. All values between should then be highlighted.
12. Press the "Ok" button at the lower left, and you will now have a new data set called "EnrichedPathwaysInMAupregulated.txt" at the root level in your Project window.
13. Select the new dataset and do "Data set | View data set" or press the  button.
14. Rename the "Gene Set" column header to "KEGG PW", and prefix all other columns with MA (e.g. "Size" to "MA Size") including the "Rank" data column. Simply double click the column headers and edit them directly.
15. Close the "Properties for ..." window you just edited the data set in, by using the X-button at the top right.
16. Repeat the "Load tabular data" exercise 1-15 above for the "EnrichedPathwaysInPRupregulated.txt" file. **Note:** In 14. replace MA with PR.

### Merge Pathway datasets into one before RP analysis

Next we need to prepare the common data set of the Pathways with their ranks from both the MA and PR analysis:

1. Use the "Data set | Merge | Merge data sets" function that we used in the first exercise.

2. Select the “EnrichedPathwaysInMAupregulated.txt” as Merge source 1.
3. and the “EnrichedPathwaysInPRupregulated.txt” as Merge source 2.
4. Set “Dataset identifier” to “KEGG PW” for both sources.
5. Remember to unselect “Quantile normalize merged data” in Step 2 and keep “Allow genes not present in all datasets” checked/selected.
6. Rename the merged dataset to “Pathway ranks combined”
7. After you have merged the data:
  - In order to get the ranks in correct relative order (making 1 the most significant value) you will run a script on the dataset negating all ranks (-1 is then the highest value).
  - The script also replaces 0 values with a large negative rank to have them sorted to the bottom of the rank lists.
  - Download the “InvertLogRatiosAndReplace0values.pyt” script file from the course web page and save it in the <J-Express>\resources\scripts folder
8. Open the script module “Methods | Scripting | Python scripting” or press .
9. In the “J-Express scripting” window, do “File | Load Script”, navigate to the <J-Express>\resources\scripts folder and double-click the “InvertLogRatiosAndReplace0values.pyt” file.
10. Select the “Pathway ranks combined” dataset in the Project window and press the  button in the “J-Express scripting” window.
11. You will then get a new child data set in the Project window. Rename this to “Final Pathway ranks combined”

## Perform the Rank Product analysis

Now we can do the final step to see which pathways if any that have low ranks in both data sets, and thus receives support from both. This will try out to define your own .gmt file, by taking the top differential expressed proteins in the PR data and evaluate these sets in the MA data.



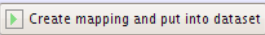
P

1. Do a Rank Product analysis on the “Final Pathway ranks combined” data set.
2. Open “Data set | Search and Sort” on the same dataset, and reorganize your windows so that you can see them both.
3. Select some of the top results in the Rank Product window, and inspect their ranks and statistics in the Search and Sort window (the columns prefixed MA and PR gives the stats from the individual data analyses).
4. Do you find some Pathways that seem to draw support from both the MA and the PR data ?
5. Keep your Rank Product window open, and close the “Search and Sort” window.

## Add annotation to data

To proceed with the pathways, we'll need their KEGG id in addition to their names. We will therefore add these as annotations to the “Final Pathway ranks combined”

data set as well, using the very handy “Annotation Manager” component. It can be used to import additional information from any tab separated file.

1. Select the “Final Pathway ranks combined” data set in the Project window.
2. Open “Data set | Annotation manager (Idlinker)” or press 
3. Select the “Add annotation – manual” tab.
4. Open the “pwtable.gmt” file you downloaded earlier by clicking the  button in the top “Annotation file” panel.
5. Next, select “KEGG PW” as “Data set key column”
6. Next, click the “Set key column” button and select the first column with KEGG PW named (turns blue).
7. Next, click the “Select columns to import” button and select the second column containing the hsa-ids (turns green)
8. Finally click 
9. The KEGG hsa-ids are now in place, and you can reopen “Data set | Search and Sort” on the “Final Pathway ranks combined” data set. Select pathways in your Rank Product window and inspect ranks and statistics again in the “Search and Sort” window. Identify a couple of pathways with id hsaxxxxx with xxxxx < 01000.

### Optional additional exercises

In the previous exercise we looked into enriched upregulated genes and proteins in the same pathways. You could also do the same exercise with down-regulated genes and proteins, or just as well up-regulated genes and down-regulated proteins (or the other way around).

You have all the necessary files prepared from the earlier exercises to do any of this combinations, go ahead if you have some extra time.