December 2019

An Introduction to Deep-Sequencing Data Analysis
Exercise 6: Functional analysis

Noa Wigoda and Bareket Dassa

Functional analysis using GeneAnalytics™

GeneAnalytics aims to identify potential associations of gene sets with pathways, compounds and Gene Ontology terms

In this part you will explore the functional signatures of the differentially expressed genes of the RNA-seq data from your UTAP analysis.

Note: Before opening the DESeq data file, read the instructions in the "EXCEL_tips"

Download the Excel file "Deseq_all_results.txt". This file is one of the outputs of DESeq analysis of the RNA-seq experiment.

1. Filter the column to select genes which were up-regulated in the LPS phenotype, according to the following thresholds:
   |log2FoldChange| ≥2 (this means that the absolute value of fold change is greater than 4)
   padj ≤ 0.05
   LPS_control.baseMean >= 5 (the average number of reads for all samples is higher than 5).
   You should get a list of 268 upregulated genes.

2. Analyze the up-regulated genes set which you just selected, using **GeneAnalytics™**.
   For this, open the GeneAnalytics website: https://ga.genecards.org/
3. You will need to register to this website **using your Weizmann email address**!
   (for non-Weizmann users, please send us an email and we will create an account for you).
4. Copy and paste the up-regulated genes, choose the relevant organism, and press "Analyze" to continue (although some IDs will not be properly identified).

   Explore the results in the different tabs and answer the following questions:
   Note: you can sort the results by their columns.

Question 1:

A. Which **pathway (SuperPath)** got the maximal score in the up-regulated gene set?

B. How many genes from your dataset matched this pathway?

Question 2:

A. Which "**GO molecular function**" is associated with the maximal number of genes of your dataset?
B. Explain why this GO term did not get the maximal score.

Download the results excel file (at the top-right of the page) and save the file in the **Exercise6 folder in your home directory**.

Repeat the analysis for the **down-regulated** genes set (use the same criteria as above). You should get a list of 72 genes.

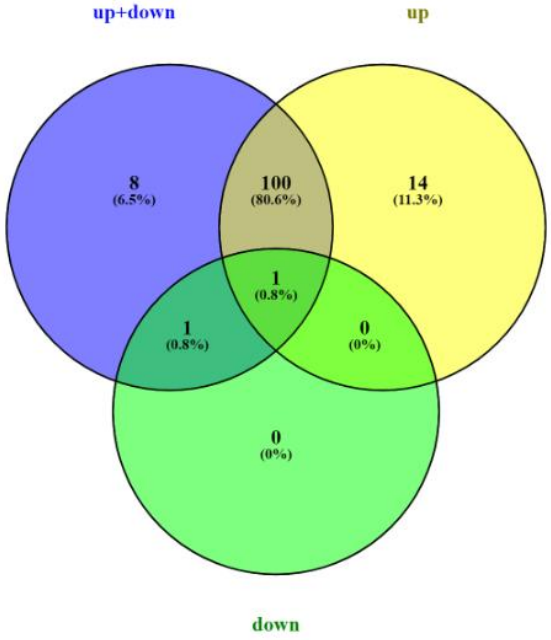
Question 3:

Which pathways were identified with a high score (colored in green) for the down genes?

Question 4:

We ran an additional analysis, using both up and down regulated genes. The overlap between the identified pathways (with high enrichment score in each separate analysis) is shown in the Venn diagram below. Note: this is not the number of genes but the identified pathways.

Explain why 14 pathways identified for up genes are not significant in the analysis of up+down genes?



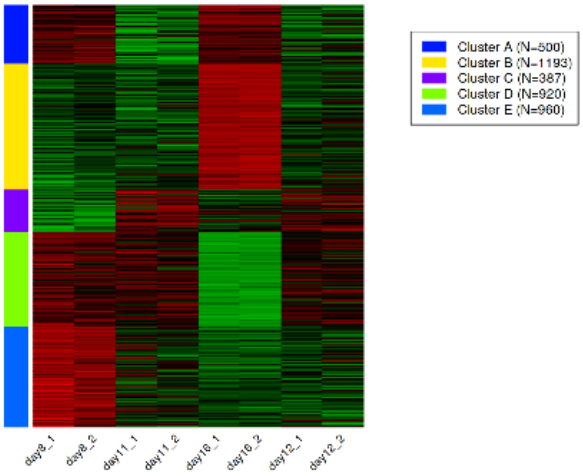
Functional analysis using Metascape

Metascape is a web-based tool, which provides a comprehensive gene list annotation and analysis resource for experimental biologists.

Open the url: http://metascape.org in your browser.

In this section, you will explore and compare the functional annotation of genes from the Arabidopsis experiment that you analyzed in exercise 5. In that exercise, you clustered the differentially expressed genes into 5 clusters using k-means algorithm.

The following table includes the clustered data from which you will prepare the input for the functional analysis. Download the file from the course webpage: Kmeans_ex6.csv
and open it in Excel.



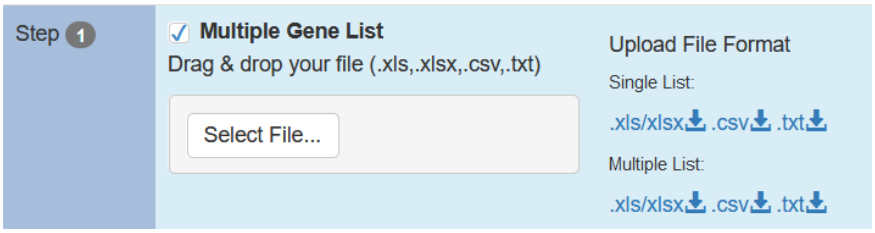
You will have to prepare two input files:

File 1: a table listing only the genes names belonging to each cluster, for example:

cluster-A	cluster-B	cluster-C	cluster-D	cluster-E
AT5G20630.1	AT3G54340.1	CRU3	AT4G22505.1	AT2G14580.1
AT5G26000.1	AT2G33810.1	AT3G20760.1	AT1G08090.1	AT1G64370.1
AT5G25980.2	AT2G39330.1	AT2G43670.1	AT3G19710.1	AT1G32470.1
AT2G05380.1	AT3G28500.1	AT5G55450.1	AT1G16410.1	AT2G32860.2
AT5G28300.1	AT2G20870.1	AT4G28520.1	AT2G43100.1	AT1G01190.1

File 2: Prepare a “background list”, which is a list of all genes, which were detected in the RNA experiment. Use the file “arabidopsis_rld.csv”.

Step 1: Upload the input file that you prepared: make sure you check the “Multiple list”



Step 2: Select the correct organism for the analysis, in both “Input as species” and “Analysis as species”.

Step 3: Press “Express analysis” button and wait for the analysis to end. This may take a few moments. Once the results are ready, press “Report Analysis page”.

Question 5:

A. Download the Gene List Report file and put it in your Exercise6 folder in your home directory. Look at the Report Analysis webpage and answer:

B. Copy the “Heatmap summary” image to your answers document.

C. Name a GO term which is exclusively enriched only in cluster A.

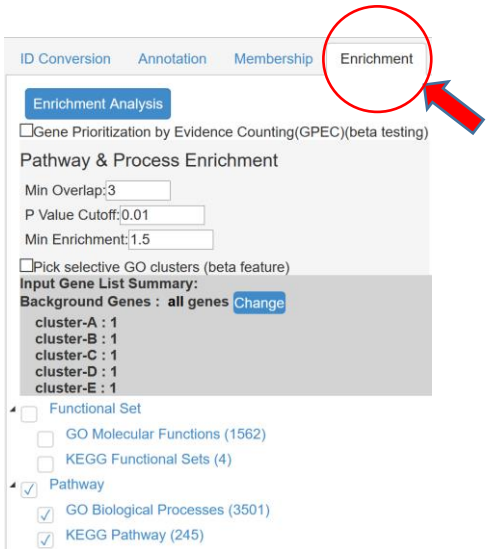
D. In which clusters is the GO term “ion transport” enriched?

As a default, Metascape performs enrichment analysis using all genes of the organism as a background. It is possible to run an analysis with a specific gene list as background.

Start a new analysis in Metascape webpage:

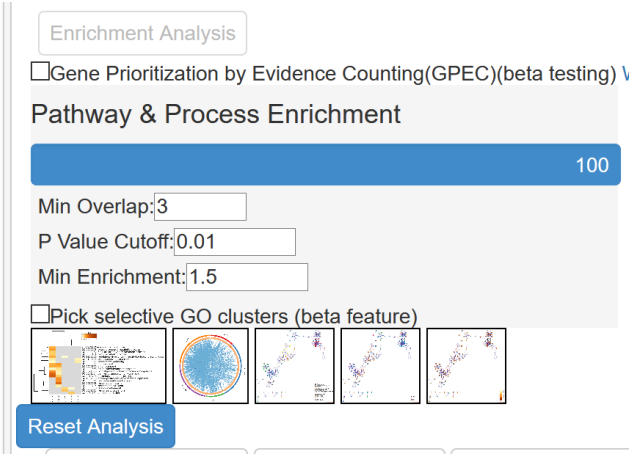
In step 1, Upload the input file (as multiple gene list).
In step 2, Select the relevant organism.
In step 3, choose “**Custom Analysis**”.
A new window will open. Select the “Enrichment” tab:

Copy the gene names from the Arabidopsis file and paste it as the background list.



Press “submit” on that window and wait patiently for the uploading of the background genes. When done, press the “Enrichment Analysis” blue button.

Results will be ready as small image icons, for example:



Save the heatmap of GO terms into your answers document.

Question 6: Compare this heatmap to the one without a customized gene background.

A. Look at cluster A and name which GO terms are different between the analysis with and without the background list.
B. Explain why did you get different results in the analysis with and without the background list?

The end!