

ELIXIR-ITA Bioinformatics Training
Protein Networks and Systems Biology
Bologna, 14-18 December 2015

Practical Session

Network module detection and functional annotation with Cytoscape

Andreas Zanzoni

Inserm TAGC UMR_S1090, Aix-Marseille University
Marseille, France

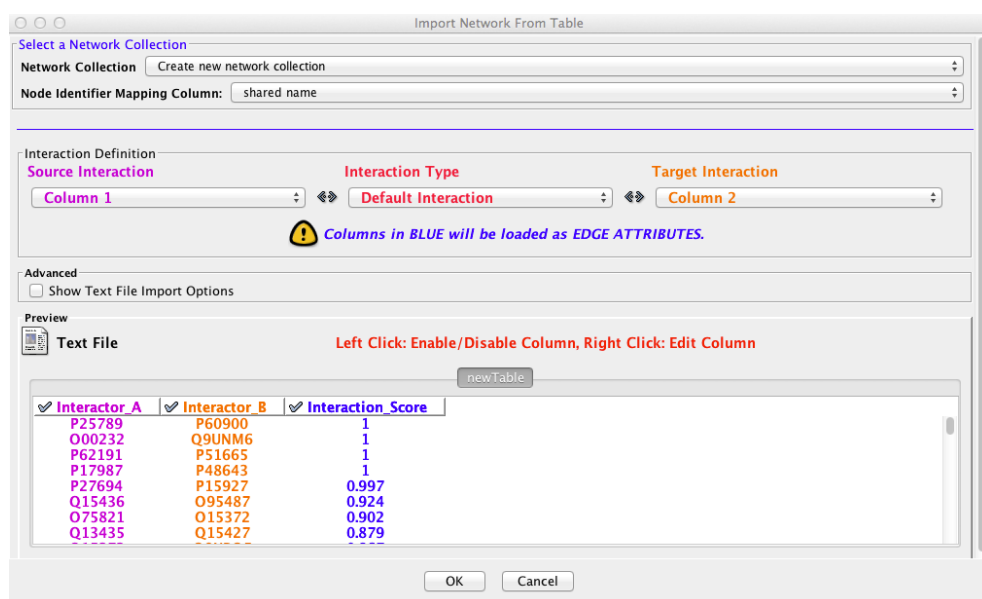
Network module detection and functional annotation with Cytoscape

In this practical session you will detect network modules in a protein-protein interaction network generated by applying an integrative global proteomic profiling approach, based on chromatographic separation of cultured human cell extracts (Havugimana et al., 2012). To do so, you will use the ClusterONE algorithm (Nepusz et al., 2012), implemented as a Cytoscape App, to detect putative macromolecular complexes. We will next use the BiNGO App (Maere et al., 2005) to perform functional enrichment analysis.

1. Download all the files at the provided URL:
 - a. **Havugimana_network.txt**: a tab-delimited file with three columns: Interactor_A, Interactor_B and Interaction_Score. Interactors are reported as UniprotKB accession numbers.
 - b. **Havugimana_xref.txt**: a tab-delimited file with additional identifiers for the proteins present in the network (i.e., UniprotKB identifiers, HGNC gene symbols).
 - c. **ClusterONE_documentation.pdf**: a short reference documentation the ClusterONE algorithm with the description of the running parameters (taken from the ClusterONE website):

http://www.paccanarolab.org/static_content/clusterone/cl1-cytoscape3-1.0.html

2. Open Cytoscape.
3. Check if the **ClusterONE** and **BiNGO** Apps are installed. If this is not the case, installed them through the *App Manager* in the *Apps* drop-down menu.
4. Load the network file (File >> Import >> Network >> from File).
5. Select the interactors and Interaction_Score columns. The latter will be loaded as an interaction attribute.



6. Load the additional identifiers file (File >> Import >> Table >> from File):

Import Columns From Table

Target Table Data

Where to Import Table Data: To a Network Collection

Select a Network Collection

Network Collection: Havugimana_network.txt

Key Column for Network: shared name

Importing Type

Import Data as: Node Table Columns

Advanced

☒ Show Mapping Options ☒ Show Text File Import Options ☒ Case Sensitive

Annotation File to Table Mapping

Select the primary key column in table:

UniprotKB_AC

Text File Import Options

Delimiter: ☒ Tab ☐ Comma ☐ Semicolon ☐ Space ☐ Other

Preview Options: ☐ Show all entries in the file ☒ Show first 100 entries.

Column Names

☒ Transfer first line as column names Start Import Row: 1 Comment Line: ☐

Refresh Preview

Preview

Text File


Left Click: Enable/Disable Column, Right Click: Edit Column

newTable

UniprotKB_AC	UniprotKB_ID	Gene_Symbol
A0AVT1	UBA6_HUMAN	UBA6
A0FGR8	ESYT2_HUMAN	ESYT2
A0MZ66	SHOT1_HUMAN	SHTN1
A1L0T0	ILVBL_HUMAN	ILVBL
A1X283	SPD2B_HUMAN	SH3PXD2B
A2A3N6	PIPSL_HUMAN	PIPSL
A4D2B8	PM2P1_HUMAN	PMS2P1
A5PLL7	TM189_HUMAN	TMEM189

File Size: Unknown

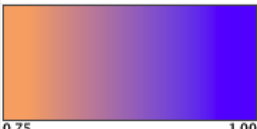
OK Cancel

- Import the data as "Node Table Columns". Verify that all the columns are correctly selected with the ✓ mark.
- By selecting the *Style* tab in the *Control panel* on the left, you can change several visual parameter of the network. For instance, you can map **Gene Symbols** as node label or change size and color of the edges according to the **Interaction Score**. You might also change the *Layout* of the network by clicking on the  button (WARNING: given the size of the network, this action will take several minutes!).

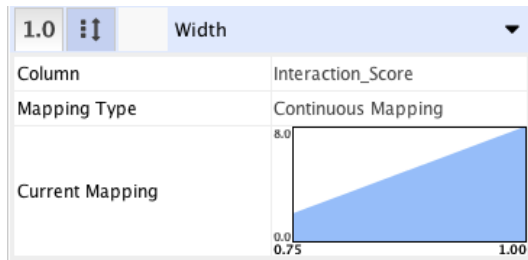
Label

Column	Gene_Symbol
Mapping Type	Passthrough Mapping

Stroke Color (Unselected)

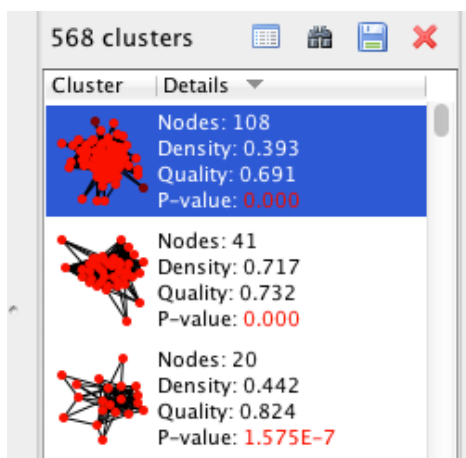
Column	Interaction_Score
Mapping Type	Continuous Mapping
Current Mapping	


0.75 1.00

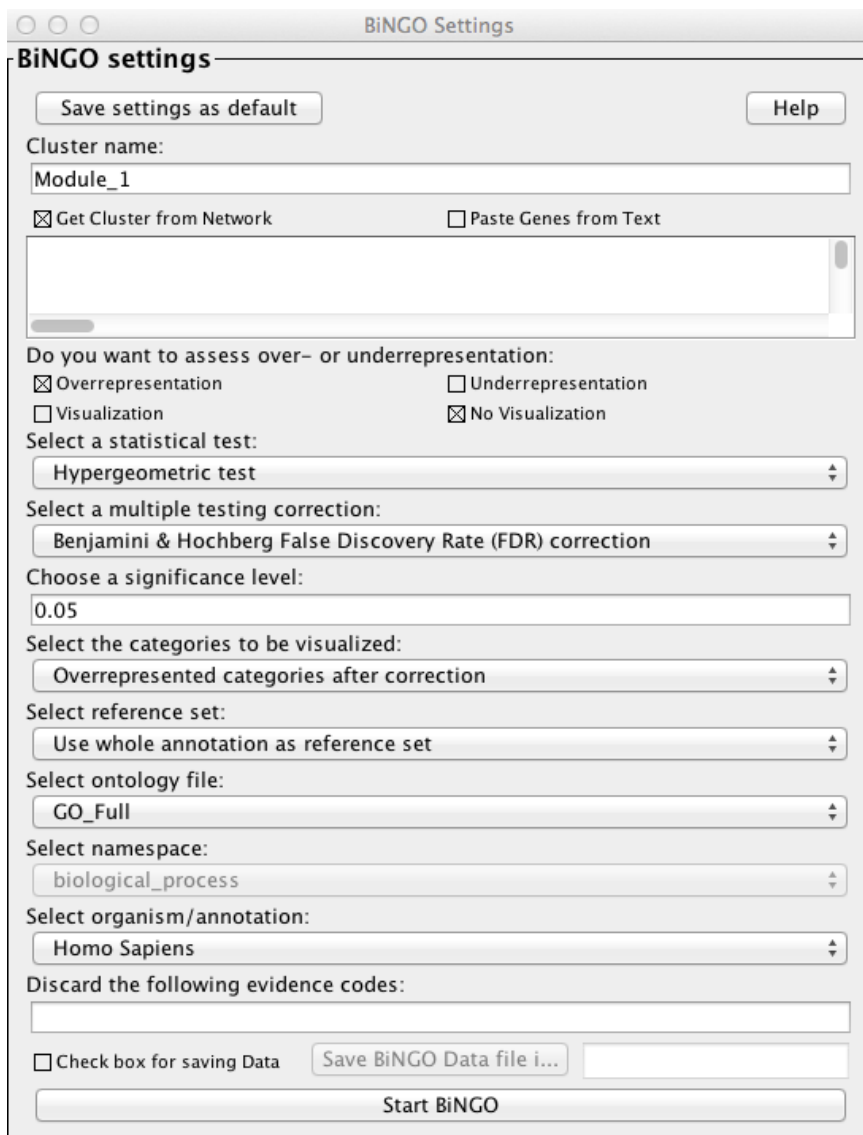


9. For your convenience, make two or three copies of the network through the *Clone Current Network* command (File >> New >> Network).
10. Select one of the cloned network in the Control panel and launch the ClusterONE App from the Apps menu (Apps >> ClusterONE >> Start). A new tab will open the *Control Panel*:

11. Set the **Minimum density** parameter to **0.4** and the **Node penalty** value to **2.9** (these values correspond to the ones used in the Havugimana *et al.* paper). Leave the other parameters unchanged.
12. Click on the *Start* button. The Result panel will appear automatically on the right side of the Cytoscape user interface.



13. Scroll down the list for see all the detected modules. You may switch to the detailed view by clicking on  and then sort for any of the parameters of your choice.
14. Select the biggest module; create a sub-network by typing *ctrl+N* (*cmd+N* on a Mac). Orange diamond nodes are proteins present in more than one module.
15. You will now perform a functional enrichment analysis of the selected module. Select all the nodes of modules (either by *Left click + Drag* or by typing *ctrl+A* / *cmd+A*).
16. Open the **BiNGO App** in the *Apps* menu.
17. Give a *Cluster name* for the module.
18. Check the *No Visualization* option.
19. Select the Ontology you want to assess (e.g., *GO_Full*).
20. Leave all the other parameters unchanged and click on *Start BiNGO*.



BiNGO settings

Save settings as default Help

Cluster name:
Module_1

☒ Get Cluster from Network ☐ Paste Genes from Text

Do you want to assess over- or underrepresentation:
☒ Overrepresentation ☐ Underrepresentation
☐ Visualization ☒ No Visualization

Select a statistical test:
Hypergeometric test

Select a multiple testing correction:
Benjamini & Hochberg False Discovery Rate (FDR) correction

Choose a significance level:
0.05

Select the categories to be visualized:
Overrepresented categories after correction

Select reference set:
Use whole annotation as reference set

Select ontology file:
GO_Full

Select namespace:
biological_process

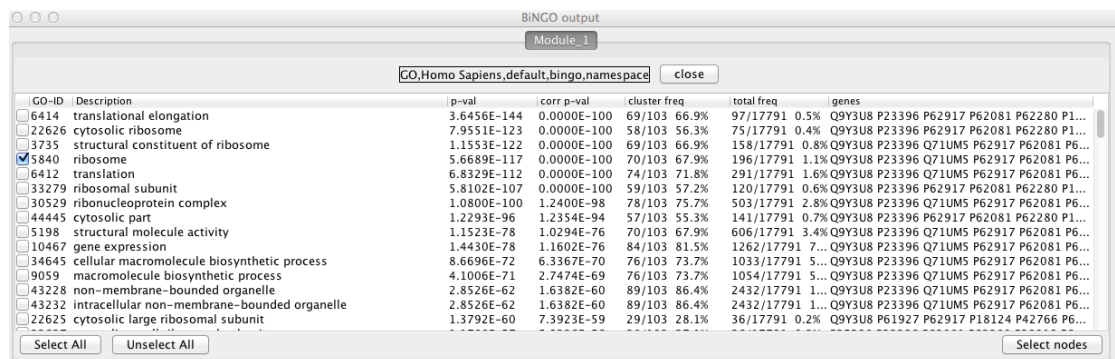
Select organism/annotation:
Homo Sapiens

Discard the following evidence codes:

☐ Check box for saving Data Save BiNGO Data file i...

Start BiNGO

21. The BiNGO output table will appear just below the Network view window.



BiNGO output

Module: 1

GO, Homo Sapiens, default, bingo.namespace

GO-ID	Description	p-val	corr p-val	cluster freq	total freq	genes
<input type="checkbox"/> 6414	translational elongation	3.6456E-144	0.0000E-100	69/103 66.9%	97/17791 0.5%	Q9Y3U8 P23396 P62917 P62081 P62280 P1...
<input type="checkbox"/> 22626	cytosolic ribosome	7.9551E-123	0.0000E-100	58/103 56.3%	75/17791 0.4%	Q9Y3U8 P23396 P62917 P62081 P62280 P1...
<input type="checkbox"/> 3735	structural constituent of ribosome	1.1553E-122	0.0000E-100	69/103 66.9%	158/17791 0.8%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input checked="" type="checkbox"/> 5840	ribosome	5.6689E-117	0.0000E-100	70/103 67.9%	196/17791 1.1%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 6412	translation	6.8329E-112	0.0000E-100	74/103 71.8%	291/17791 1.6%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 33279	ribosomal subunit	5.8102E-107	0.0000E-100	59/103 57.2%	120/17791 0.6%	Q9Y3U8 P23396 P62917 P62081 P62280 P1...
<input type="checkbox"/> 30529	ribonucleoprotein complex	1.0800E-100	1.2400E-98	78/103 75.7%	503/17791 2.8%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 44445	cytosolic part	1.2293E-96	1.2354E-94	57/103 55.3%	141/17791 0.7%	Q9Y3U8 P23396 P62917 P62081 P62280 P1...
<input type="checkbox"/> 5198	structural molecule activity	1.1523E-78	1.0294E-76	70/103 67.9%	606/17791 3.4%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 10467	gene expression	1.4430E-78	1.1602E-76	84/103 81.5%	1262/17791 7.0%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 34645	cellular macromolecule biosynthetic process	8.6696E-72	6.3367E-70	76/103 73.7%	1033/17791 5.8%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 9059	macromolecule biosynthetic process	4.1006E-71	2.7474E-69	76/103 73.7%	1054/17791 5.9%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 43228	non-membrane-bounded organelle	2.8526E-62	1.6382E-60	89/103 86.4%	2432/17791 1.3%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 43232	intracellular non-membrane-bounded organelle	2.8526E-62	1.6382E-60	89/103 86.4%	2432/17791 1.3%	Q9Y3U8 P23396 Q71UM5 P62917 P62081 P6...
<input type="checkbox"/> 22625	cytosolic large ribosomal subunit	1.3792E-60	7.3923E-59	29/103 28.1%	36/17791 0.2%	Q9Y3U8 P61927 P62917 P18124 P42766 P6...

Select All Unselect All Select nodes

22. Select one of the enriched terms (e.g., ribosome) through the corresponding checkbox and then click on the *Select nodes* button. In the Network window, annotated nodes will be highlighted in yellow.

23. Feel free to:

- Repeat the functional enrichment analysis on any module of your choice.
- Select one of the other cloned networks and change the **ClusterONE** running parameters:
 - Use the *Interaction_Score* as edge weight. Modify the *density* threshold or the *node penalty*.
 - Do you get more or less modules compare to the first run? Does the composition of the biggest modules change (Tools >> Merge >> Networks)? And their functional annotations?

References

Havugimana PC et al. A census of human soluble protein complexes. Cell. 2012 Aug 31;150(5):1068-81. doi:10.1016/j.cell.2012.08.011. PubMed PMID: 22939629; PubMed Central PMCID: PMC3477804.

Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005 Aug 15;21(16):3448-9. PubMed PMID: 15972284.

Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods. 2012 Mar 18;9(5):471-2. doi: 10.1038/nmeth.1938. PubMed PMID: 22426491; PubMed Central PMCID: PMC3543700.