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
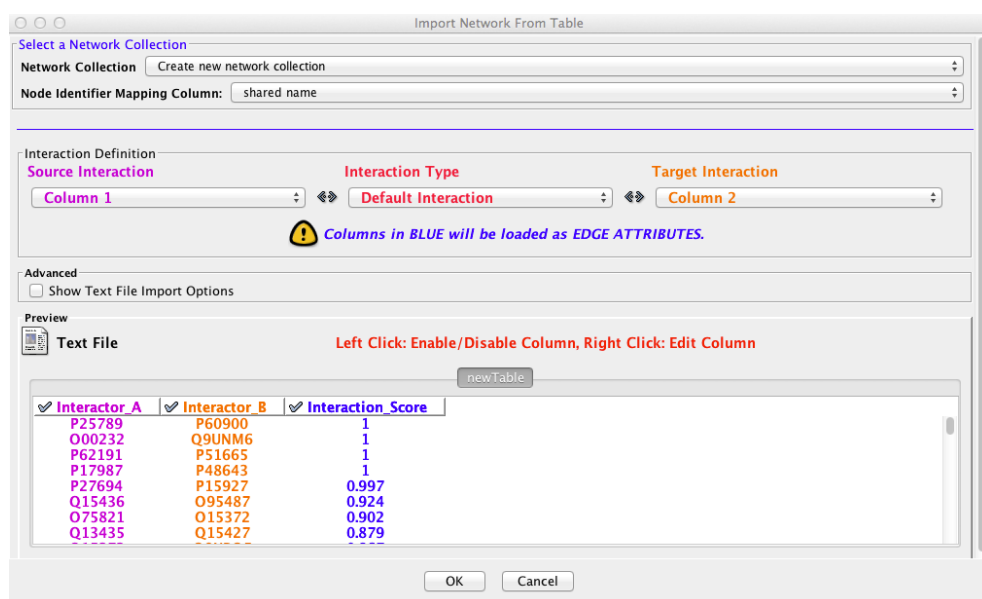Marseille, France

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 provived 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.

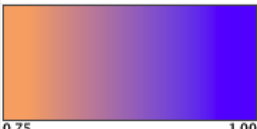6. Load the additional indentifiers file (File >> Import >> Table >> from File):

7. Import the data as "Node Table Columns". Verify that all the columns are correctly selected with the ✔ mark.

8. 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!).
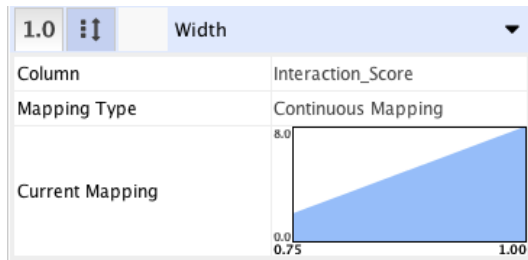
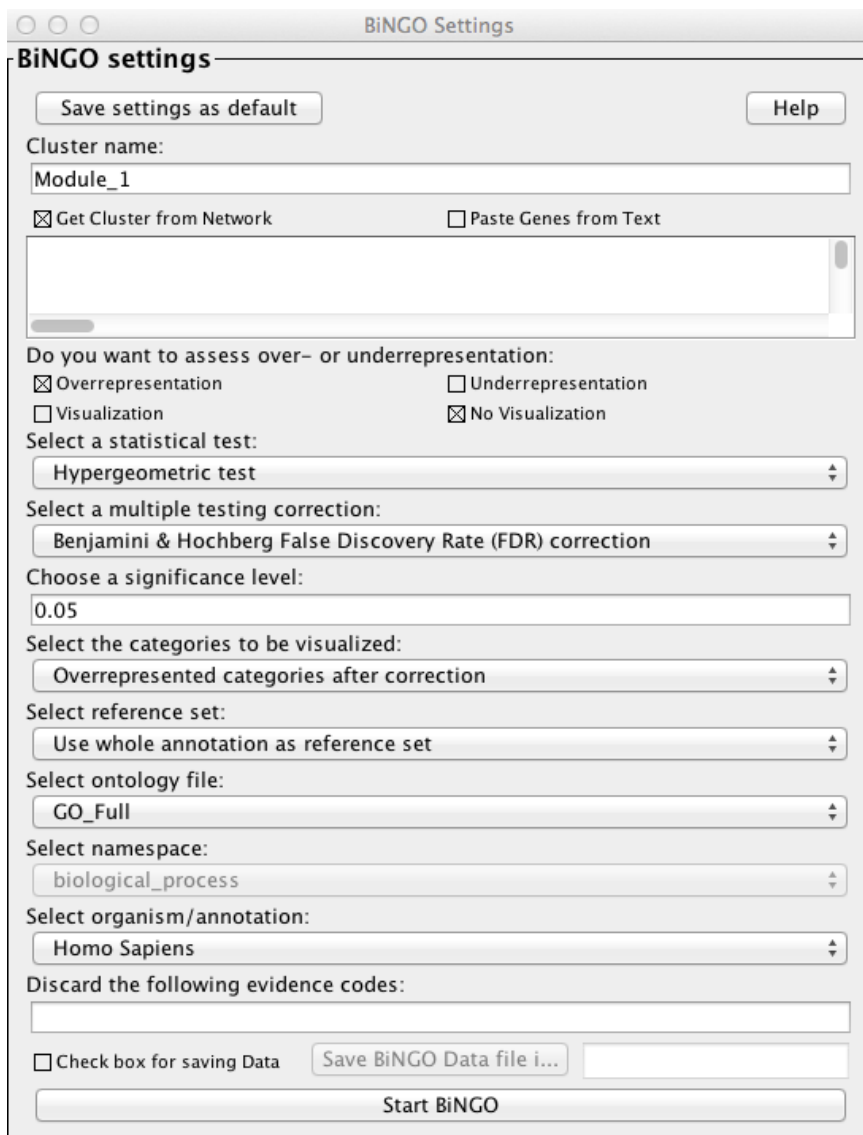9. For your convenience, make two or three copies of the network through the *Clone Current Network* command (File >> New >> Network).

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.
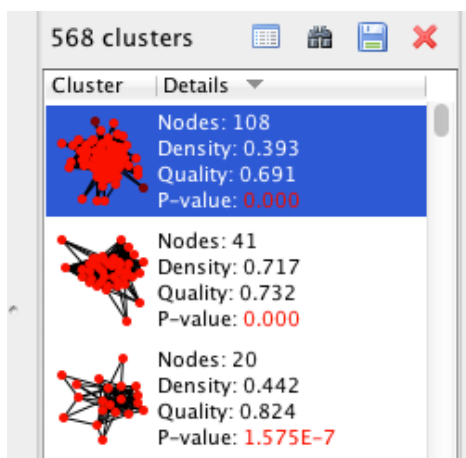
13. Scroll down the list for see all the detected modules. You may switch to the detailed view by clicking on ⊞ and then sort for any of the parameters of your choice.
14. Select the biggest module; create a sub-network by typing *ctrl+N* (*cmd+N* on a Mac). Orange diamond nodes are proteins present 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

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   ↕
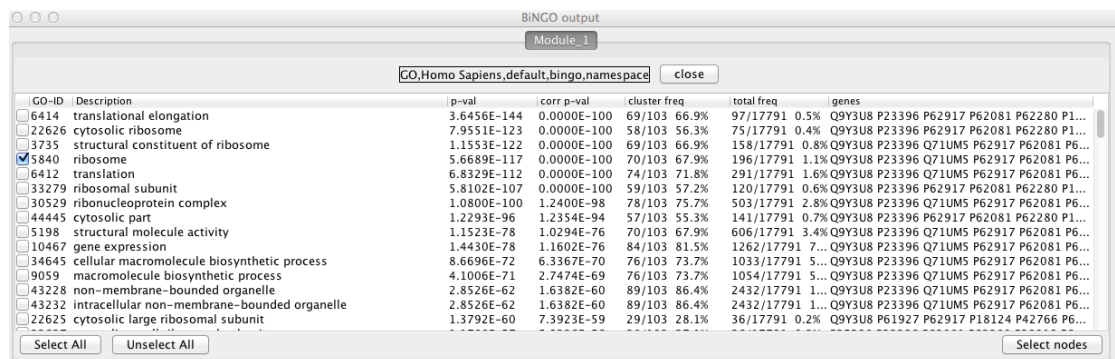
Select namespace:
biological_process   ↕

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.

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:
    a. Repeat the functional enrichment analysis on any module of your choice.
    b. Select one of the other cloned networks and change the **ClusterONE** running parameters:
        i. Use the *Interaction_Score* as edge weight. Modify the *density threshold* or the *node penalty*.
        ii. 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.