atBioNet Get Started

Introduction

This application is developed at the National Center for Toxicological Research (NCTR), a U.S. Food and Drug Administration (FDA) center. It can be used by following the steps below:

1. User copy/pastes their protein/gene data in the ID panel by right mouse click or CTRL + V. Currently the following formats are supported:
    - EntrezGeneID (default)
    - GENEBANKACC
    - GENENAME
    - UNIGENEID
    - SWISSPROT_NAME
    - SWISSPROT_ACC
    - IPI
    - AffyID
2. Click "Run".
3. Protein/gene IDs are matched with our database to find their interactions in the atBioNet PPI (protein-protein interaction) database.
4. Cluster analysis is performed by using SCAN.
5. Results are summarized and the top six modules are graphically displayed. Proteins/genes inputted by the user will appear as square nodes while additional network nodes will appear as circular nodes.

Leukemia data [1] and Lupus data [2] are provided as examples. You can use the example data to get familiar with the application.

The Structural Clustering Algorithm for Networks (SCAN) is developed by Professor Xiaowei Xu [3]. Some modifications by NCTR are implemented. The graphic exploration system (GUESS) is visualization software modified from http://sourceforge.net/projects/guess/.

Graph

When the "draw 6 modules" option is selected, their positions on screen are as follows:

1 (highest ranking module)	2	3
4	5	6

Summary Report

Shows how many nodes and edges were added to the input nodes.

Module/Pathway Summary

Shows pathway summary for each module. Detailed pathway analysis can be accessed by double clicking on a row. The results can be saved as .txt file using the right mouse click.

SCAN Results

Displays results from the SCAN program. Module IDs of (-1) are outliers and that of (-2) are hubs as defined in graphic network. The results can be saved as .txt file using the right mouse click.

Menu Items

File->Open: Open data file.
File->Save As: Save data file.
File->Calculate Statistics: Calculate Degree Rank, Page Rank, HITS and BETWEENNESS. You can view them in InfoWindow.
File->Save Image: Save the graph as an image file. Available formats include .JPG, .PNG, etc.
File->Save GDF…: Save to GDF graphic file.
File->Exit: Exit program.

Options->Show Options: Open option dialog (see below for detailed explanation).

Graph->Draw 6 Modules (checkbox): Indicate the current graph mode. Clicking here will switch between draw the first 6 modules or the full graph.
Graph->InfoWindow: Open information window. It will display information when the user mouse over a node or an edge.
Graph->background color: Change background color.
Graph->center: Center graph to default view.

Help->Online Help: Display this page.
Help->About: Displays details of current version of atBioNet.

Node addition options

1. *Use only input nodes: uses user inputted protein/gene only:* No additional nodes are added.
2. *Add all directly connected nodes:* any protein/gene directly connected with the inputted protein/gene from our database are added (default).
3. *Add only nodes directly connected to at least two input nodes (more stringent):* add related protein/gene from our database only if the added protein/gene directly connected with at least two of the user input protein/gene. This option will add much less nodes than option 2 above.

Edge addition options

1. *Add only connections containing at least one input node:* each connection is a protein pair. This option requires that any added protein pairs must contain at least one input protein/gene (default).
2. *Add all connections (including added nodes):* this will include connections between nodes added from our database. Note this will only affect results if you have selected option 2 or 3 in the node addition options.

Database options

1. *Human Database (default).*
2. *K2 Human Subset Database:* A smaller and more robust human PPI database obtained by the integration of seven original databases using K-votes approach, which is proposed by Martha et al [4].
3. *All Species Database:* Use PPI from all species including human, rat, and mouse.

SCAN options

1. *Sort module by*
   - *Input Nodes:* Number of input nodes (default).
   - *Nodes:* Number of nodes or size of the module.
   - *Modularity:* M. J. Newman's modularity score [5].
2. *Modified SCAN method (less outliers):* Check this box to add outliers/hubs to the nearest directly connected module; if checked, modules will contain more nodes and less outliers.
3. *Use the largest connected component as input for SCAN:* The network established by using the input proteins/genes may contain several isolated connected components. If checked, only the largest connected component will be used as the input for SCAN. Proteins/genes are connected if they have a direct connected path.
4. *Sub-scan Method*
   - *Scan only:* No sub-scan is performed.
   - *Scan with sub-scan.* After scan, the program will continue to scan the largest module if the size of the largest module is more than 3 times of the size of the second largest module. This is to ensure that the size of the largest module is controlled and to ensure relatively even distribution of size of modules. The program will continue try to sub-scan up to 3 times or until the criterion is met (default).
   - *Largest<=50/100/200:* The program will do sub-scan if the size of the largest module is more than 50/100/200.
5. *Search Intervals (10/50/100/200):* The program will search in the specified steps in the interval (0,1) to find the largest modularity score. The larger number will give more accurate results but with slower speed (default=10).

References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA *et al*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, 286(5439):531-537.
2. Arasappan D, Tong W, Mummaneni P, Fang H, Amur S: Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. *BMC Med* 2011, 9:65.
3. Xu X, Yuruk N, Feng Z, Schweiger T: SCAN: a structural clustering algorithm for networks. In: *In Proceedings of the 13th ACM SIGKDD international conference on Knowledge Discovery and Data Mining.* San Jose, California, USA; 2007: 824-833.
4. Martha V-S, Liu Z, Guo L, Su Z, Ye Y, Fang H, Ding D, Tong W, Xu X: Constructing a robust protein-protein interaction network by integrating multiple public databases. *BMC Bioinformatics* 2011, **12**(Suppl 10):S7.
5. M. Newman: Fast algorithm for detecting community structure in networks. *Phys. Rev. E 69*, art. No. 066133, 2004.