

Poster J-44

GenePro: a Cytoscape Plugin for Advanced Visualization and Analysis of Interaction Networks



Authors:

Jim Vlasblom (*Center for Computational Biology, The Hospital for Sick Children*)
Samuel Wu (*Center for Computational Biology, The Hospital for Sick Children*)
Shuye Pu (*Center for Computational Biology, The Hospital for Sick Children*)
Mark Superina (*Center for Computational Biology, The Hospital for Sick Children*)
Gina Liu (*Center for Computational Biology, The Hospital for Sick Children*)
Chris Orsi (*Center for Computational Biology, The Hospital for Sick Children*)
Shoshana J. Wodak (*Center for Computational Biology, The Hospital for Sick Children, Departments of Biochemistry and Me*)

Short Abstract: Systems biology studies frequently require integrative, interactive analysis and visualization of large scale protein/gene interaction networks. We present GenePro, a plugin to Cytoscape, featuring multi-resolution display of interaction networks at both cluster and protein/gene level. GenePro also allows mapping between different partitioning of the network.

Long Abstract:

Analysis of the genome-scale networks of interacting genes and proteins has become a major focus of systems biology. Currently however, there is a shortage of effective software tools for the interactive analysis and display of such networks. Available tools tend to offer simple graph layout options but lack interactive capabilities for mapping complex properties of genes/proteins onto the graphs at various granularity levels. Cytoscape, a public domain software environment (Shanon et al., 2003), provides a series of basic functionalities to layout and query network graphs, and to visually integrate the network with other types of information. These functionalities are extensible through a straightforward plugin architecture, allowing rapid development of additional computational analyses and features. A number of such plugins have already been reported (<http://www.cytoscape.org/plugins2.php>).

Here we present GenePro, a Cytoscape plugin that provides several integrative and interactive visualization and analysis tools for networks of interacting proteins/genes. In particular, these tools enable visualization and analyses of interaction networks at two levels of resolution. Lower resolution views display clusters of genes/proteins as individual nodes enabling the analysis of relationships between clusters. Higher resolution views allow for more detailed interrogation of individual clusters or a few neighboring clusters by displaying genes/proteins and their mutual interactions within the cluster. With this multi-resolution feature, GenePro greatly facilitates the analysis of protein networks derived from high throughput interactions data and the validation of various methods for parsing these networks into meaningful functional modules. In fact, GenePro has been successfully used in visualization and analysis of multi-protein complexes of the yeast *S. cerevisiae* derived from a protein-protein interaction (PPI) network reported in a recent proteomics study (Krogan et al., 2006).

Visualizing clusters of interacting proteins: The Cytoscape core and other software packages

allow the display of graphs where nodes usually represent individual genes or proteins. The nodes are connected to one another whenever a physical or functional interaction has been observed or inferred. GenePro allows displaying of more complex graphs where nodes are groups of genes/proteins onto which a set of properties of interest can be mapped, and then visualized or queried interactively. Furthermore, the composite nodes within these graphs are connected to one another on the basis of a derived metric which indicates the strength of the link between nodes. For instance, two nodes can be connected if there are interactions between proteins in each cluster, with the thickness of the edge being proportional to the number of such interactions. This representation positions the protein clusters (or complexes) in the context of global interaction network, revealing which of the clusters is highly connected to other clusters and suggesting the extent to which the predicted complexes may share genes in vivo. Some clustering procedures produce clusters that share components (overlapping clusters). For such cases the user can specify edges to represent the number of shared genes between nodes.

To facilitate the validation and analysis of the computed clusters, GenePro offers an additional set of features. Each cluster can be represented as a pie chart where the size and color of wedges represent the fraction of the genes/proteins in each cluster that share identical attributes. Positioning the Mouse over a given wedge displays the number and names of the proteins/genes that share the particular attribute within the complex, and a mouse click over the same wedge highlights proteins in other nodes anywhere in the network sharing the same attribute. This latter feature provides an overview of the extent to which proteins with the same attribute are distributed throughout other clusters.

These visualization features are very general and can be used to map any two grouping of genes/proteins onto one another. They include, but are not limited to, mapping the hand-curated complexes in the Comprehensive Yeast Genome Database catalogue, the functional categories in the Gene Ontology, the subcellular localizations, the level of sequence conservation, or groups of co-regulated genes/proteins onto the network of complexes/clusters and vice versa. Various options for mapping mRNA expression data onto the network of protein/gene clusters are also available, as are various options for interactively querying the displayed information.

Displaying the interacting genes/ proteins in individual clusters: In addition to displaying and querying the network of protein clusters, GenePro allows analyzing the proteins in individual clusters and their pair-wise interactions. A double click on a given cluster displays a new graph, whose nodes are the proteins within the cluster, as well as their nearest neighbors from other clusters (colored differently from the proteins within the cluster) and the arcs are their interactions in the PPI network. Clicking on a link between two genes displays a Table listing the reliability score of the interaction, as well as the raw score of each observation recorded for the interaction in the original proteomics experiments.

Implementation aspects: The data required to activate the features are loaded into GenePro via tab-delimited text files. SIF files that are required by Cytoscape to create the interaction networks are generated automatically. The various features in the GenePro plugin have been developed in Java. (We thank the CIHR for support under the Canada Research Chair program and support by the McLaughlin Center for Molecular Medicine is gratefully acknowledged.)

References

1. Krogan, N. J., Cagney, G., et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, 440(7084): 637-643.

2. Shannon,P., Markiel,A., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11): 2498-2504.