Systems biology

GenePro: a cytoscape plug-in for advanced visualization and analysis of interaction networks

James Vlasblom[1], Samuel Wu[1], Shuye Pu[1,*], Mark Superina[1], Gina Liu[1], Chris Orsi[1] and Shoshana J. Wodak[1,2,3,*]

[1]Structural Biology and Biochemistry Program, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada, [2]Department of Biochemistry and [3]Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

Motivation: Analyzing the networks of interactions between genes and proteins has become a central theme in systems biology. Versatile software tools for interactively displaying and analyzing these networks are therefore very much in demand. The public-domain open software environment Cytoscape has been developed with the goal of facilitating the design and development of such software tools by the scientific community.

Results: We present GenePro, a plugin to Cytoscape featuring a set of versatile tools that greatly facilitates the visualization and analysis of protein networks derived from high-throughput interactions data and the validation of various methods for parsing these networks into meaningful functional modules.

Availability: The GenePro plugin is available at the website http://genepro.ccb.sickkids.ca

contact: shuyepu@sickkids.ca

1 INTRODUCTION

Analyzing the networks of interacting genes and proteins has become a major focus of present day biology. Currently however, there is a shortage of effective software tools for the interactive analysis and display of such networks. Most available tools tend to offer simple graph layout options (Breitkreutz *et al.*, 2003; Goldovsky *et al.*, 2005) but lack interactive capabilities for mapping complex properties of genes/proteins onto the graphs.

Few existing tools allow visualization of protein–protein interaction (PPI) networks at various granularity levels. CNplot (Batada, 2004) displays pre-clustered PPI networks, but does not support further analysis of individual cluster members as well as the interactions among them. VisANT (Hu *et al.*, 2005) uses a bottom-up approach that allows grouping of proteins in an existing PPI network into meta-nodes; however, input of the grouping is not straightforward, and the visual representation, manipulation and storage of the meta-nodes is not suitable for large-scale analysis of network modules (such as all multiprotein complexes in the proteome of yeast *Saccharomyces cerevisiae*).

Here we present GenePro, a Cytoscape (Shanon *et al.*, 2003) plugin that provides several integrative and interactive visualization and analysis tools for PPI networks. UnlikeVisANT, we use a

cluster-centric top-down approach. Given a PPI network and a predefined clustering of its proteins, GenePro first provides a low-resolution view by displaying the protein clusters as individual nodes that are interconnected through interacting or shared proteins. For higher resolution views, individual clusters can be expanded to allow for more detailed analyses of clusters in terms of their component proteins and their interactions in the context of the full PPI network. With this cluster-centric multi-resolution feature, GenePro greatly facilitates the analysis of PPI networks derived from high-throughput interactions data and the validation of various methods for parsing these networks into meaningful functional modules. Various options for mapping gene expression data and other relevant properties onto the network of protein/gene clusters are a particularly attractive feature of GenePro.

2 VISUALIZING CLUSTERS OF INTERACTING PROTEINS

Proteome-scale PPI networks with thousands of proteins and tens of thousands of interactions cannot be readily displayed in a meaningful way using conventional graph display tools. GenePro makes it possible by using composite nodes that represent 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 contains relevant information about the strength of the link between nodes.

The particular example illustrated in Figure 1a shows a graph in which individual nodes represent multi-protein complexes (clusters) in yeast derived using clustering procedures from a PPI network reported in a recent comprehensive, high-throughput pull-down analysis (Krogan *et al.*, 2006). The clusters/nodes are linked to one another whenever there are at least two observed 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 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 members *in vivo*. Some clustering procedures produce clusters that share components (overlapping clusters). For such cases the user can specify edges between nodes to represent the number of shared proteins.

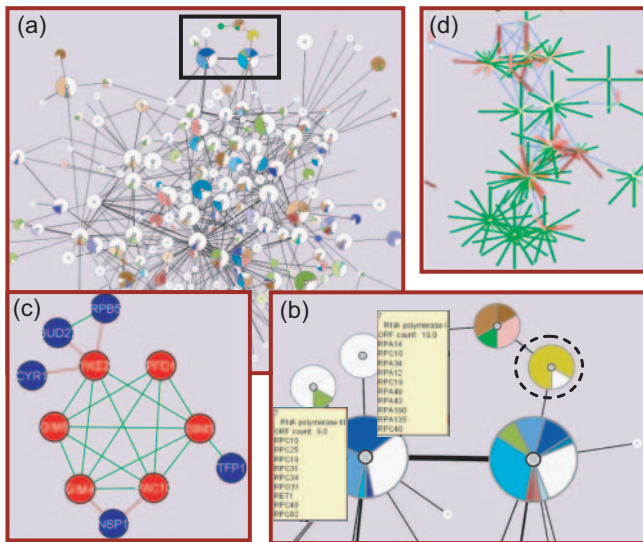---

*To whom correspondence should be addressed.

Fig. 1. Illustration of features in GenePro, a Cytoscape plugin for advanced visualization and analysis of protein complex networks. (**a**) Protein cluster network of *S.cerevisiae*, with each node representing one cluster. (**b**) Illustrating nodes in (a) as pie charts, with each wedge of the pie representing the fraction of genes with a common feature, here the known CYGD complex to which they belong; the actual genes included in this fraction are listed. (**c**) A new graph representing interactions between individual genes belonging to the node/clusters highlighted by a dotted circle in (b). (**d**) mRNA levels for individual genes in some complexes are shown as red spikes (up-regulation) and green spikes (down-regulation), the length and color shade of the spikes illustrate the magnitude of the expression levels.

To facilitate the validation and analysis of the computed clusters, GenePro offers an additional set of features. Figure 1b displays a close up view of a portion of the network of clusters displayed in Figure 1a. In this view each node, still representing a cluster of proteins, appears as a pie chart where the size and color of the wedges represent the fraction of the proteins in the cluster that share the same property. One such property is the membership in a complex from the Comprehensive Yeast Genome Database (CYGD) complexes catalogue (Güldener *et al*., 2005), which is displayed in Figure 1a and b in order to illustrate the overlap of the computed clusters with the annotated complexes. Positioning the mouse over a given wedge displays the number and names of the proteins that belong to a given CYGD complex (e.g. RNA polymerase I complex) within the cluster, and a mouse click over the same wedge highlights proteins in other nodes anywhere in the network that belong to the same complex. This latter feature (data not shown) provides an overview of the extent to which proteins from the same CYGD complex are distributed throughout other clusters.

These visualization features are very general and can be used to map any two groupings of proteins onto one another. In addition to complexes/clusters, such groupings may include (1) functional categories, (2) subcellular localizations, (3) level of sequence conservation and (4) groups of co-regulated genes/proteins (Simonis *et al*., 2004). And all these can be mapped onto the network of complexes/clusters and vice versa. Additional options for mapping mRNA expression data (Spellman *et al*., 1998; Gasch *et al*., 2000; Hugues *et al*., 2000) onto individual gene in the network

of protein/gene clusters are also available (Fig. 1d), as are various options for interactively querying the displayed information.

3 VISUALIZING INDIVIDUAL CLUSTERS IN THE CONTEXT OF THE INTERACTION NETWORK

In addition to displaying and querying networks of protein clusters, GenePro allows analyzing the proteins in individual clusters and their pairwise interactions. A double click on a given cluster displays a subgraph (Fig. 1c), whose nodes are the proteins within the cluster (red nodes), as well as their nearest neighbors from other clusters (blue nodes) and the arcs are their interactions in the PPI network. Clicking on an arc 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.

4 IMPLEMENTATION ASPECTS

The data required to activate the features described in this note are loaded into GenePro via tab-delimited text files. SIF files that are required by Cytoscape to create the interaction networks are generated automatically as needed. The various features in the GenePro plug-in have been developed in Java. In addition to a user manual, example datasets, tutorial movies and a tutorial document are also available at http://genepro.ccb.sickkids.ca.

REFERENCES

Batada,N.N. (2004) CNplot: visualizing pre-clustered networks. *Bioinformatics*, **20**, 1455–6.

Breitkreutz,B.J. *et al*. (2003) Osprey: a network visualization system. *Genome Biol.*, **4**, R22.

Gasch,A.P. *et al*. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, **11**, 4241–57.

Goldovsky,L. *et al*. (2005) BioLayout(Java): versatile network visualisation of structural and functional relationships. *Appl. Bioinform.*, **4**, 71–74.

Güldener,U. *et al*. (2005) CYGD: the comprehensive yeast genome database. *Nucleic Acids Res.*, **33**, D364–D368.

Hu,H. *et al*. (2005) VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res.*, **33**, W352–W357.

Hughes,T.R. *et al*. (2000) Functional discovery via a compendium of expression profiles. *Cell*, **102**, 109–26.

Krogan,N.J. *et al*. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, **440**, 637–643.

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

Simonis,N. *et al*. (2004) Transcriptional regulation of protein complexes in yeast. *Genome Biol.*, **5**, R33.

Spellman,P.T. *et al*. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–97.