ToPNet—an application for interactive analysis of expression data and biological networks

Daniel Hanisch¹, Florian Sohler²,* and Ralf Zimmer²

¹Institute for Algorithms and Scientific Computing (SCAI), Fraunhofer Gesellschaft, Schloss Birlinghoven, 53754 Sankt Augustin, Germany and ²Institut für Informatik, Ludwig-Maximilians-Universität München, Amalienstrasse 17, 80333 München, Germany

Received on July 16, 2003; Revised on November 13, 2003; accepted on November 28, 2003
Advance Access publication February 12, 2004

ABSTRACT

Summary: ToPNet is a new tool for the combined visualization and exploration of gene networks and expression data. ToPNet provides various ways of restricting, manipulating and combining biological networks according to annotation data (e.g. Gene Ontology terms) and presents results to the user via different visualization procedures and hyperlinks to the underlying data sources. To easily identify relevant parts of the network, ToPNet provides a method of detecting significant subnetworks with respect to expression measurements. As ToPNet is a pure JAVA application with additional scripting capabilities, it is well-suited as a test-bed for algorithm development and exploratory biological data analysis alike.

Availability: ToPNet is freely available for academic institutions at http://www.biosolveit.de/ToPNet/
Contact: Florian.Sohler@bio.ifi.lmu.de

INTRODUCTION

The analysis of gene expression data poses a great challenge to molecular biology today. With present microarrays, the expression level of several thousands of genes in a biological sample can be estimated simultaneously. A biological expert analyzing this information usually has in-depth knowledge of only a small subset of the measured genes. Moreover, knowledge of interaction partners and associated regulators of significantly expressed genes is essential for interpreting experimental results accurately. This prior knowledge is contained in databases of biological networks (e.g. metabolic, regulatory or interaction networks). Based on such information, the ToPNet program is able to find connected subnetworks that are relevant with respect to the expression measurements and supply the user with database links and literature references of proteins and interactions of interest.

CONCEPTS

ToPNet was designed to handle several networks from multiple sources that can be restricted according to user-specified criteria and manipulated using a simple graph editor. Properties of the networks (e.g. color, size and hyperlinks) can be associated with annotation data (e.g. expression data or functional annotations) via mappings of names or identifiers. Figure 1 shows how these components work together to provide a useful visualization of networks and annotation data.

Representation of networks

Networks are implemented in the form of PETRI Nets. These are essentially bipartite graphs in which one set of nodes (termed places) represents molecules (e.g. proteins or metabolites) and the other set of nodes (called transitions) defines relationships among these molecules. This representation is valuable when complex reactions are considered. For instance, in metabolic networks, places represent metabolites and enzymes. One metabolic reaction is then represented by a transition vertex that is connected via edges to the participating molecule places (cf. Fig. 2). Other network types include regulatory, protein–protein interaction and literature networks.

*To whom correspondence should be addressed.
Annotations for networks: data maps

Data maps handle annotation data in ToPNet by providing standardized information about their content. For example, expression data are often available in tabular format where rows represent genes and columns correspond to specific experimental conditions. The table itself contains, e.g. probability values or fold changes quantifying differential expression. A data map then provides information for a gradient color coding of corresponding genes on a linear or logarithmic scale, for the size property of displayed molecules and annotation with the corresponding value as a tooltip. As another example, terms from the Gene Ontology (GO) (The Gene Ontology Consortium, 2001) can be treated as a data map in ToPNet, thereby associating a set of GO terms with each gene. This data map provides, besides direct annotation of places with terms, the useful possibility of linking directly to the corresponding entries in the GO hierarchy via a Web browser.

Providing the link: mappings

To connect annotation data to network properties, a mapping is essential. As several major gene and protein databases exist and a general nomenclature for protein and gene names is still missing, ToPNet is able to load mappings for different sets of identifiers interactively and visualize the results. In practice, we found that the generation of these mappings constitutes a major obstacle in everyday work and therefore provide standard mappings (which make use of the International Protein Index available at http://www.ebi.ac.uk/IPI/) in our distribution that allow seamless integration of expression data with many types of networks.

Algorithms

The goal of ToPNet is to facilitate the generation and verification of biological hypotheses. Besides the visualization of annotations, ToPNet provides several means of restricting networks to interesting regions with respect to experimental data. First, using only the network information, the user can compute shortest paths among places of interest and explore the neighborhood of genes by building hulls of specified size. Second, a simple query language based on available data maps is implemented. For example, given that probability values and GO annotations are available, an expression like

\[ \text{GO biological process \text{ like } apoptosis \& p-value \leq 0.05} \]

would select all apoptosis-related genes with a significant p-value. Third, two algorithms, termed significant area search and pathway queries, can extract interesting regions of the network based on expression data (Sohler et al., 2004). Both algorithms produce networks that are reported to the user for further exploration with ToPNet. In practice, results often correspond to pathway-like structures undergoing significant regulation according to the measured expression data.

CONCLUSIONS AND FUTURE WORK

ToPNet is a JAVA-based application with a rich set of GUI elements, which makes the above-mentioned concepts and algorithms easily accessible. The visualization of networks and associated properties is based upon the open-source JGraph library (Alder, 2002, http://prdownloads.sourceforge.net/jgraph/paper1.0.6.pdf). ToPNet aims at enabling biologists to generate and validate hypotheses concerning expression data and biological networks. The algorithms in conjunction with the graphical user interface make ToPNet a valuable tool for exploring interactively the meaning of expression experiments in context of biological networks.

ACKNOWLEDGEMENTS

The development of ToPNet was in part supported by Aventis Pharma, Frankfurt (project BEX) and the German ministry for research and education (project BOA).

REFERENCES