Electrical and Computer Engineering Publications   Electrical and Computer Engineering

2008

OmicsViz: Cytoscape plug-in for visualizing omics data across species

Tian Xia
Iowa State University

Julie A. Dickerson
Iowa State University, julied@iastate.edu

OmicsViz: Cytoscape plug-in for visualizing omics data across species

Abstract

Motivation: OmicsViz is a Cytoscape plug-in for mapping and visualizing large-scale omics datasets across species, including those with many-to-many mappings between homologs. This allows users to map their data onto pathways of related model organisms. Mapping schemas across species or different experimental protocols allow users to comparatively analyze the omics data. The data can also be visualized in parallel-coordinate plots.

Disciplines

Bioinformatics | Systems and Communications

Systems biology

OmicsViz: Cytoscape plug-in for visualizing omics data across species

Tian Xia[1] and Julie A. Dickerson[1,2,*]

[1]Program of Bioinformatics and Computational Biology and [2]Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50010, USA

ABSTRACT

Motivation: OmicsViz is a Cytoscape plug-in for mapping and visualizing large-scale omics datasets across species, including those with many-to-many mappings between homologs. This allows users to map their data onto pathways of related model organisms. Mapping schemas across species or different experimental protocols allow users to comparatively analyze the omics data. The data can also be visualized in parallel-coordinate plots.

Availability: The latest version of OmicsViz with documentation, tutorial and jar files can be downloaded from http://metnet.vrac.iastate.edu/MetNet_fcmodeler.htm

Contact: julied@iastate.edu

1 INTRODUCTION

High-throughput technologies provide large-scale datasets which describe biointeractions at various levels of detail, such as genes, protein and metabolites. However, integrating multiple independent datasets and mapping such datasets across species and gaining biological insight are challenging. One of the main obstacles for easy use of this data with networks is multiple naming conventions for experimental data and network elements and many-to-many mappings when applying BLAST homology to map across species.

Cytoscape (Shannon *et al.*, 2003) is an open-source software platform for network modeling in biology. OmicsViz is a Cytoscape plug-in that associates experimental datasets from transcriptomic, proteomic and metabolomic datasets with biological network models in Cytoscape according to node attribute information. OmicsViz can uniquely handle many-to-many mappings between nodes in a network and experimental datasets.

2 METHODS AND IMPLEMENTATION

Node attributes in Cytoscape characterize elements with specific name IDs and diverse biological information, such as subcellular location, node type, etc. Edges reflect the interactions or other relations between the nodes.

2.1 Omics data import and mapping

OmicsViz maps labeled measurements in the input experimental data to target nodes if the names are identical. However, due to inconsistencies in nomenclature, different names may be used to describe experimental quantities and network nodes, which actually refer to the same biological component in one species. One common example is that microarray datasets use gene probe IDs but metabolic pathways use gene locus names. OmicsViz provides a mapping file which translates between locus IDs and probe IDs to load transcriptomic data into the network model.

A unique feature of OmicsViz is that it deals with problems caused by many-to-many mappings between nodes due to homologs or uncertainty as to which node was matched or which compartment it occurs in. The matched homologs are represented by newly created nodes and connected to target nodes. Functionally, they are grouped into HomologGroups (implemented using CyGroups) with the target nodes. The cross-species transcriptomic data, therefore, can be loaded into target species network. Combining further visualization and comparative analysis, one can easily explore biological insights for both species based on prior knowledge and omics experimental observations. The imported data can be stored as a Cytoscape pvals file.

The mapping rules between the labels in the data file, X, and the node names in the active graph, Y, $F(X) \rightarrow Y$ where $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_M\}$ can be summarized as:

(1) If $F(x_i)$ maps to a unique value of y_j (i.e. a one-to-one mapping), then the data associated with x_i is mapped as attributes of node, y_j.

(2) If $F(x_i)$ maps to multiple values of Y, then the data are mapped to the set of nodes, $\{y \mid y = F(x_i)\}$.

(3) If multiple values of X, defined as the set, X_i, map to a set of nodes in the graph, i.e. $\cup_{x \in X_i} \{y \mid y = F(x)\}$, then additional 'homolog' nodes and edges connected to these nodes are added to the graph. The attributes of the newly created nodes are the same as the parent node.

OmicsViz uses data files in formats ranging from comma separated value files to Excel spreadsheets to take advantage of data from PLEXdb (Shen *et al.*, 2005), GEO, and other microarray databases. OmicsViz can map multiple datasets to networks at the same time. This helps users to compare results across datasets. The plug-in dynamically updates the CyNode attribute panel and lists all selected datasets.

2.2 Data visualization

OmicsViz provides line chart visualization component for loaded omics datasets using the JFreeChart library Gilbert and Morgner, 2007 (Fig. 1). The experimental data associated with nodes in the network can be displayed on a parallel coordinate plot. The parallel coordinates of line chart delineate the variability of omics data with respect to different treatments and time points. The plot window can be controlled by users by adjusting parameters, such as line colors, axis ranges and labels.

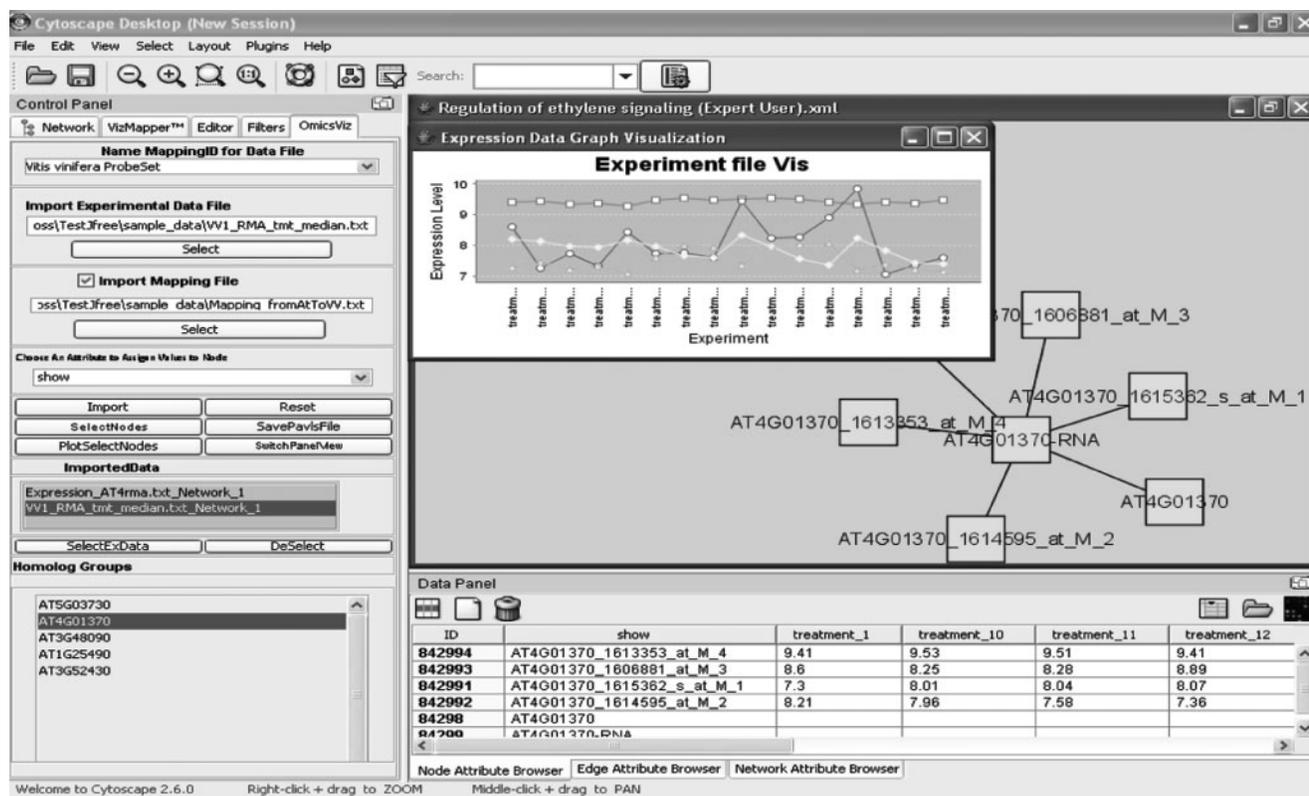*To whom correspondence should be addressed.

Fig. 1. OmicsViz plugin control panel screenshot. The highlighted yellow nodes show mapped Vitis homologs onto an *Arabidopsis* gene locus. The parallel coordinate plot of the expression data is shown in the upper right corner.

2.3 Usage example

OmicsViz is also able to import transcriptomic data into Cytoscape network platform across species to facilitate further comparative survey. For example, the *Arabidopsis* genome has been completely sequenced and the annotation is well developed. Grape (*Vitis vinifera*), has only recently been sequenced and most genes are electronically inferred. The reciprocal better BLAST hits method (Stuart *et al.*, 2003) can produce whole-genome scale homolog mappings between species. This forms a basis for the mapping of omics data from a less-studied organism to the network model developed for the plant model organism, *Arabidopsis*. The better BLAST hit method gives a set of possible homologs between species. This example uses TBLASTN for each *Arabidopsis* protein sequence against the *Vitis* consensus sequences on the Affymetrix Vitis GeneChip to find the top five grape hits and BLASTX to find the top five hits from *Arabidopsis* for each *Vitis* sequence. If *Arabidopsis* gene A and *Vitis* probeset B are reciprocal hits, A and B then are grouped as a possible homolog pair. This gives a many-to-many mapping relationship. Mappings for many plant species can be found in the sample data files available with the OmicsViz plugin. The microarray data is from experiment VV2 in PLEXdb (www.plexdb.org). OmicsViz also manages multiple datasets which are associated with specific networks. It enables users to compare datasets of interest and dynamically update the CyNode attribute panel and switch between corresponding datasets.

REFERENCES

Gilbert,D. and Morgner,T. (2007) JFreeChart, a free java class library for generating charts. Available at http://www.jfree.org/jfreechart/.

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

Shen,L. *et al.* (2005) BarleyBase–an expression profiling database for plant genomics. *Nucleic Acids Res.*, **33**, D614–D618.

Stuart,J.M. *et al.* (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.