Systems biology

VistaClara: an expression browser plug-in for Cytoscape

Robert Kincaid[1],[*], Allan Kuchinsky[1] and Michael Creech[2]

[1]Agilent Laboratories, 5301 Stevens Creek Blvd., Santa Clara, CA 95051 and [2]Blue Oak Software, 1734 Austin Ave., Los Altos, CA 94024, USA

ABSTRACT

Summary: VistaClara is a plug-in for Cytoscape which provides a more flexible means to visualize gene and protein expression within a network context. An extended attribute browser is provided in the form of a graphical and interactive permutation matrix that resembles the heat map displays popular in gene-expression analysis. This extended browser permits a variety of display options and interactions not currently available in Cytoscape.

Availability: http://chianti.ucsd.edu/cyto_web/plugins/index.php

Contact: robert_kincaid@agilent.com

1 INTRODUCTION

Cytoscape (Shannon *et al*., 2003) is a freely available open source platform for the analysis of biological networks which is widely used by system biologists. Cytoscape provides facilities for incorporating node and edge attributes and mapping such data onto the network via visual encodings (color, shape, etc.). For protein- or gene-expression studies, the expression data for a given condition or treatment would correspond to a single attribute measured across a number of biomolecules. More specifically, this would typically correspond to a single micorarray experiment, or possibly a single mass spectrometry run. The default capabilities of Cytoscape are essentially limited to mapping only a single such experimental condition at a time (Cline *et al*., 2007).

VistaClara was originally written as a stand-alone tool for interactively analyzing gene expression data from multiple experimental conditions. We soon realized VistaClara's usefulness as an expression browser for network analysis and prototyped a number of useful coordinated interactions between the tabular style of VistaClara and biological networks using our own proprietary network analysis tools. This investigation (Vailaya *et al*., 2005) led to a natural desire to extend Cytoscape with the same functionality. Hence, we have re-implemented and extended VistaClara as an attribute browser plug-in for Cytoscape (Fig. 1).

2 IMPLEMENTATION

The design of VistaClara is based on an approach informed by research in information visualization. In this article, we describe briefly the main features of the software. The reader can refer to the paper describing the original stand-alone version (Kincaid, 2004) for additional details.

2.1 Permutation matrix with overview

VistaClara starts with the traditional heat map visualization commonly used to display gene expression data, and extends this to a fully interactive permutation matrix supporting column and row rearrangement as described by Bertin (Bertin, 1981).

Since sorting by a single row or column is often ineffective for analyzing expression data, VistaClara also allows sorting rows using measures of similarity between entire rows of expression data. A given row of interest is chosen, and the remaining rows are ordered by similarity (Euclidean or Pearson) to the chosen row. Similarity sorts can be performed almost as quickly as a standard sort, thereby retaining the benefits of a highly interactive row permutation operation while revealing more significant correlations.

While heat map views are commonly used to view ratio-based expression data as a graphical matrix, there is good evidence that for small differences color intensity is difficult to visually resolve. Bertin and others have advocated size as a preferred, more visually comparable representation. We optionally replace the typical heat map rectangles with filled circles (called 'ink blobs'), whose diameters are in proportion to the represented values. If the diameter goes beyond a critical threshold, the cell in which the circle appears is filled. This change in shape is an additional highly visible cue that the data has exceeded this threshold.
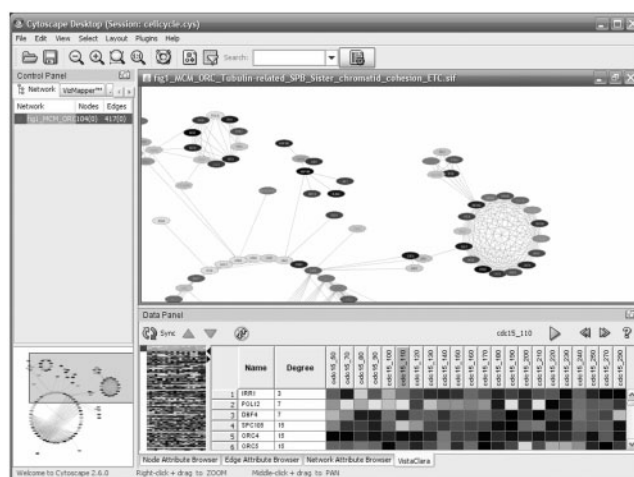

Fig. 1. Cytoscape 2.6 showing the location and appearance of the VistaClara panel. Here, the table view shows a heat map display.

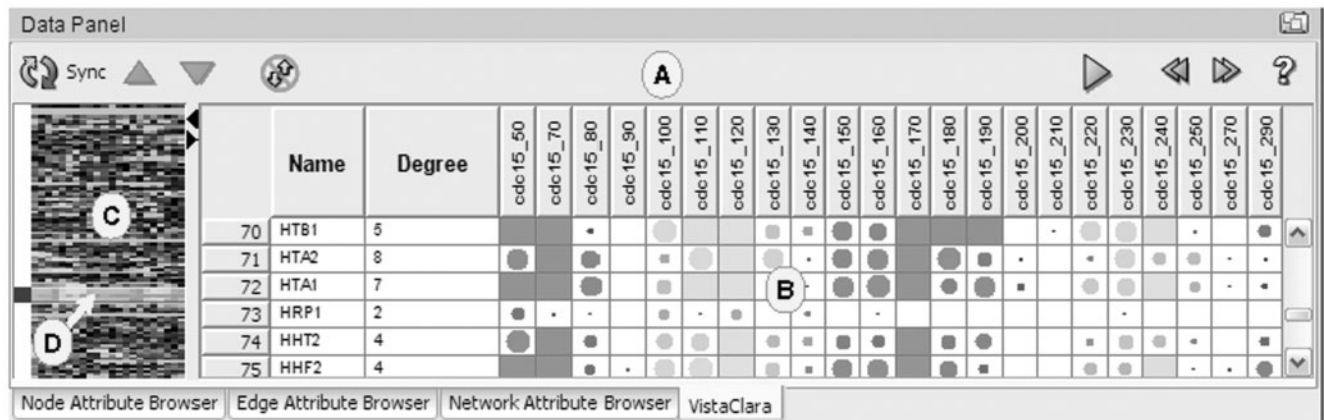*To whom correspondence should be addressed.

Data Panel

	Name	Degree	cdc15_50	cdc15_70	cdc15_80	cdc15_90	cdc15_100	cdc15_110	cdc15_120	cdc15_130	cdc15_140	cdc15_150	cdc15_160	cdc15_170	cdc15_180	cdc15_190	cdc15_200	cdc15_210	cdc15_220	cdc15_230	cdc15_240	cdc15_250	cdc15_270	cdc15_290
70	HTB1	5																						
71	HTA2	8																						
72	HTA1	7																						
73	HRP1	2																						
74	HHT2	4																						
75	HHF2	4																						

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser | VistaClara

Fig. 2. The main components of the VistaClara panel. (**A**) The main tool bar. (**B**) The graphical table view (in ink-blob mode). (**C**) Condensed heat map view of the entire data set. (**D**) A cursor showing the scroll position of the table view. Here, ink blobs become rectangular at 4-fold changes in gene expression.

VistaClara also provides an overview display of the entire data set in the form of a dynamic heat map (Fig. 2C). As rows and columns are rearranged, the overview is updated to reflect the change and any emerging correlations that might be visible beyond the tabular view. The overview can also be used to navigate to regions of interest by simply clicking on the appropriate location in the overview. A cursor (Fig. 2D) shows the user the portion of the full data set currently visible in the table view (Fig. 2B).

2.2 Coordintated network interactions

Integrating VistaClara directly into Cytoscape enables several useful coordinated interactions:

Mapping of experimental conditions is easily managed by simply clicking on the appropriate table column header to select that column for network node coloring.

For studying the dynamics of a system, the user can click the play button to automatically select each condition in succession for network node coloring. For time series experiments this permits the creation of an animated view of the network as it changes over time. Forward and reverse buttons allow user control for replaying specific conditions of interest.

A mode is provided where selections are linked between the table and network views. This permits coordinated navigation. Selecting a single node in the network scrolls the table to reveal the corresponding row in the table. Selecting a cell in the table zooms into the neighborhood of the corresponding node in the network. Multiple selections are also supported in either view with full coordination between views.

2.3 Heat strips

VistaClara also provides a feature to view all expression data simultaneously as described previously in a study of coronary atherosclerosis (King *et al.*, 2005). Each node is shown with an accompanying glyph in the form of a bar graph which depicts all expression values for that node. Redundant heat map coloring is used to reinforce both the heat map scheme as well as the interpretation of the bar graph. We call these glyphs heat strips, referring to their
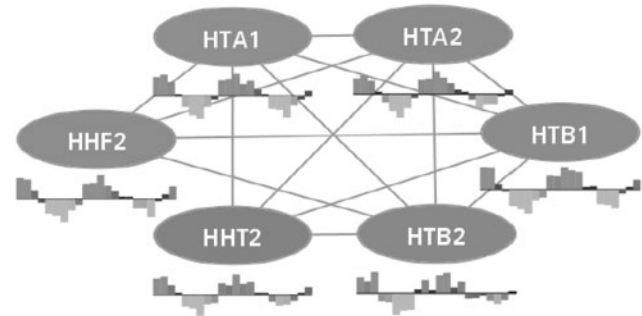
Fig. 3. Heat strips display bar-graphs of all expression data for each node. Temporal patterns in time series data are readily apparent.

encoding relationship to heat maps. Figure 3 shows a display of cell cycle data (de Lichtenberg *et al.*, 2005; Spellman *et al.*, 1998). For these particular nodes, the sinusoidal behavior of the gene expression through two phases of cell cycle is quite obvious.

3 INSTALLATION

The current version of the VistaClara plug-in (Version 1.0) requires Cytoscape 2.6 and is freely available from the Cytoscape website or through the built-in Cytoscape plug-in manager. After installation, plug-in specific online help is accessible at run time and includes a short tutorial about data syncing as well as more detail about general usage. The help button is the yellow question mark seen in the toolbar (Fig. 2A).

Conflict of Interest: none declared.

REFERENCES

Bertin,J. (1981) *Graphics and Graphic Information-Processing*. de Gruyter, Berlin, New York.

Cline,M.S. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.

de Lichtenberg,U. *et al*. (2005) Dynamic complex formation during the yeast cell cycle. *Science*, **307**, 724–727.

Kincaid,R. (2004) VistaClara: an interactive visualization for exploratory analysis of DNA microarrays. In *Proceedings of the 2004 ACM Symposium on Applied Computing,* pp. 167–174. ACM, Nicosia, Cyprus.

King,J.Y. *et al*. (2005) Pathway analysis of coronary atherosclerosis. *Physiol. Genomics*, **23**, 103–118.

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

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**, 3273–3297.

Vailaya,A. *et al*. (2005) An architecture for biological information extraction and representation. *Bioinformatics*, **21**, 430–438.