

iVici 0.91 for Windows, Mac OS X, Linux, Unix

Introduction

iVici is designed for visualization and comparative analysis of cellular gene or protein networks represented as two-dimensional matrices. The typical data flow consists of creating a first dataset that contains an association matrix describing, for example, protein-protein interactions. Gene/protein names in the dataset are row and column names. Prior to visualization, such datasets may be subjected to symmetric hierarchical clustering or other methods that can reveal groups of genes/proteins with similar association patterns. iVici can interpret two tab delimited text data formats: a) a general text file containing association matrix values together with row and column names; b) a *minimal* CDT file format, which is the output of Cluster software by Eisen [ref]. After the first dataset is loaded, the second dataset can be added and compared. A second dataset may describe other types of relationships between the same genes/proteins as in the first dataset. Thus, for example, protein-protein interaction data can be visually correlated with GO annotations or mRNA expression profiles. iVici is not limited to analysis of symmetric datasets. Matrices of any dimension may be visualized and correlated using general and superimposed modes (see Examples).

In this manual, you will find information on how to use the software and prepare datasets for analysis. For more information on principles of heat-map representation of biological networks, please, refer to the original iVici paper [1].

Data Formats

iVici recognizes two tab-delimited text file formats: CDT file format, which is used for output of Cluster software [2], and text files containing a matrix of values together with row and column names.

Txt format:

Text files contain a table where names in a first row and a first column represent genes or proteins. Values in the table describe some association value between genes/proteins. For example, if GeneA and GeneB are proteins and the table describes protein-protein interactions, then value +1 indicates that there is a direct interaction between these proteins (Fig.1).

Figure 1. Text file format

ID	GeneA	GeneB	GeneC	GeneD
GeneA	0	+1	0	+1
GeneB	+1	0	0	0
GeneC	0	0	0	0
GeneD	+1	0	0	0

This text file format is the same as a *minimal* Cluster input file. ID field (in blue) is optional in iVici and there are no specific requirements for its value, however it is required for Cluster.

CDT format:

This version of iVici works with minimal CDT Files (.cdt) (Fig1) and corresponding Tree Files (.atr and .gtr) (Fig2). The *minimal* CDT output file will be

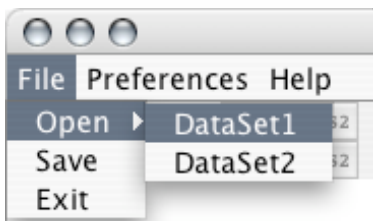
produced by Cluster when a *minimal* text file (see Txt format) is used as an input for hierarchical clustering.

The first row in a .cdt file contains the names of mandatory fields (GID, YORF and GWEIGHT), followed by unique names of entities in the columns (generally referred to as array names). The second line contains AID headers and AID identifiers for array names. The third line contains EWEIGHT values. In the current version of iVici EWEIGHT and GWEIGHT values are considered to be equal to 1 and kept only for purpose of direct compatibility with the output of Cluster. The following lines contain GID, gene/protein identifier and association values with genes/proteins; the names are given in the first row. For more information on CDT file format, please, refer to the original Cluster manual.

Figure 2. Minimal CDT File Format

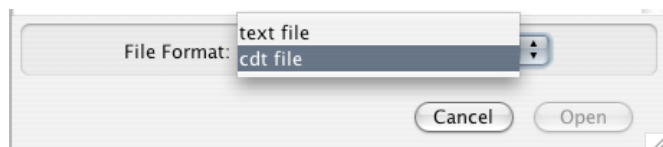
GID	YORF	GWEIGHT	GeneA	GeneB	GeneC	GeneD	GeneE	GeneF
AID			ARRY0X	ARRY1X	ARRY2X	ARRY3X	ARRY4X	ARRY5X
EWEIGHT			1	1	1	1	1	1
GENE0X	GeneA	1	-1	-1	-1	-0.5	-0.5	0
GENE1X	GeneB	1	-1	-1	-1	-0.5	-0.5	0
GENE2X	GeneC	1	-1	-1	-1	0	0	-0.5
GENE3X	GeneD	1	-0.5	-0.5	0	-1	-1	-0.5
GENE4X	GeneE	1	-0.5	-0.5	0	-1	-1	-0.5
GENE5X	GeneF	1	0	0	-0.5	-0.5	-0.5	-1

Loading Data



In order to start using iVici you should load a first data set starting from the File pulldown menu as follows: File > Open > DataSet1. A file menu will appear and you will be

provided with an option to choose the appropriate file format (TXT or CDT). If CDT file

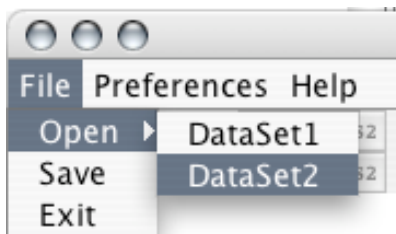


format is selected, iVici will attempt to locate corresponding .atr and .gtr files and will display data matrix and clustering trees. However, if .atr and .gtr files are missing there will be no warning message and only the data matrix will be displayed.

iVici allows for loading of two datasets and their visual comparison. DataSet2 is selected by choosing File > Open > DataSet2. This option will be available only if DataSet1 is already loaded. DataSet2 can be in one of the described formats.

Datasets binding:

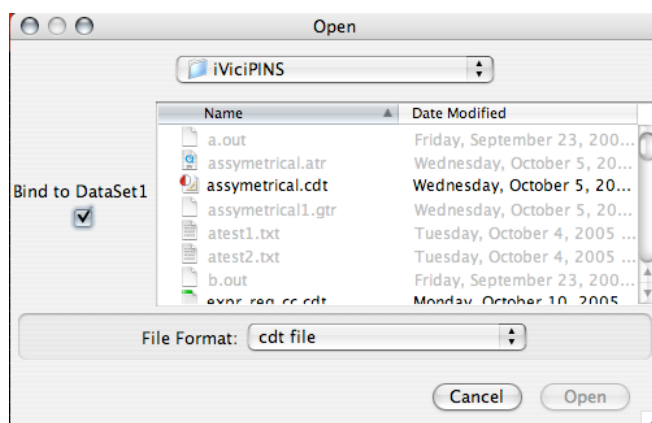
DataSet2 is not required to be of the same dimensions as DataSet1 and can have an



arbitrary order of rows and columns. iVici will automatically put datasets in accordance. When loading DataSet2 there are two options of how to

correlate it with DataSet1. DataSet2 may be bound to DataSet1. In this case columns and rows values from the DataSet2 will be displayed only if corresponding columns and rows

names are present in the DataSet1.



Thus, DataSet1 will serve as a reference set and DataSet2 will be trimmed and sorted accordingly to the reference. If DataSet1 was loaded in CDT format together with clustering trees, only these trees will

be displayed in iVici when two datasets are bound. In order to bind datasets, select the option "Bind to DataSet1" when opening DataSet2. If you wish to change the reference

data set, another dataset should be opened as DataSet1 and this will become the new reference. When DataSet1 is changed, DataSet2 is kept if it has the same dimensions, otherwise it will be deleted from iVici.

If “Bind to DataSet1” option is not selected the original dimensions and ordering of datasets is preserved. Datasets can be viewed in General and Superimposed modes (see Visualization Modes).

Basic Visualization (Fig 3)

Accessing information:

When a dataset is loaded heat-maps and clustering trees (if corresponding files are provided) will be displayed (see Loading Data). Column and row names will be used for heat-maps annotations. A single click on a cell i,j of the heat-map will display the following information in the Node Info field: i -th row name, j -th column name, data value i,j and data value i,j from the second dataset (if DataSet2 is loaded). The same information will be displayed as a popup following a left mouse click. A double click on a cell i,j will highlight it together with the cell j,i .

Links:

By clicking on heat-map cell, row and column name will appear in the Links field. You can configure (see Configuring Links) up to 4 different links that will connect you to external or internal web resources based on row/column names of a selected cell.

Navigation:

Visual layout is designed in such a way that clustering trees (if present) and annotations will always be displayed even if only a portion of the heat-map can fit into

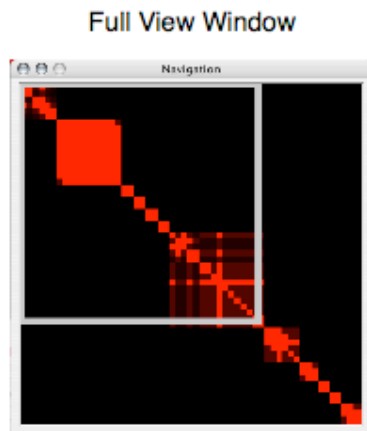
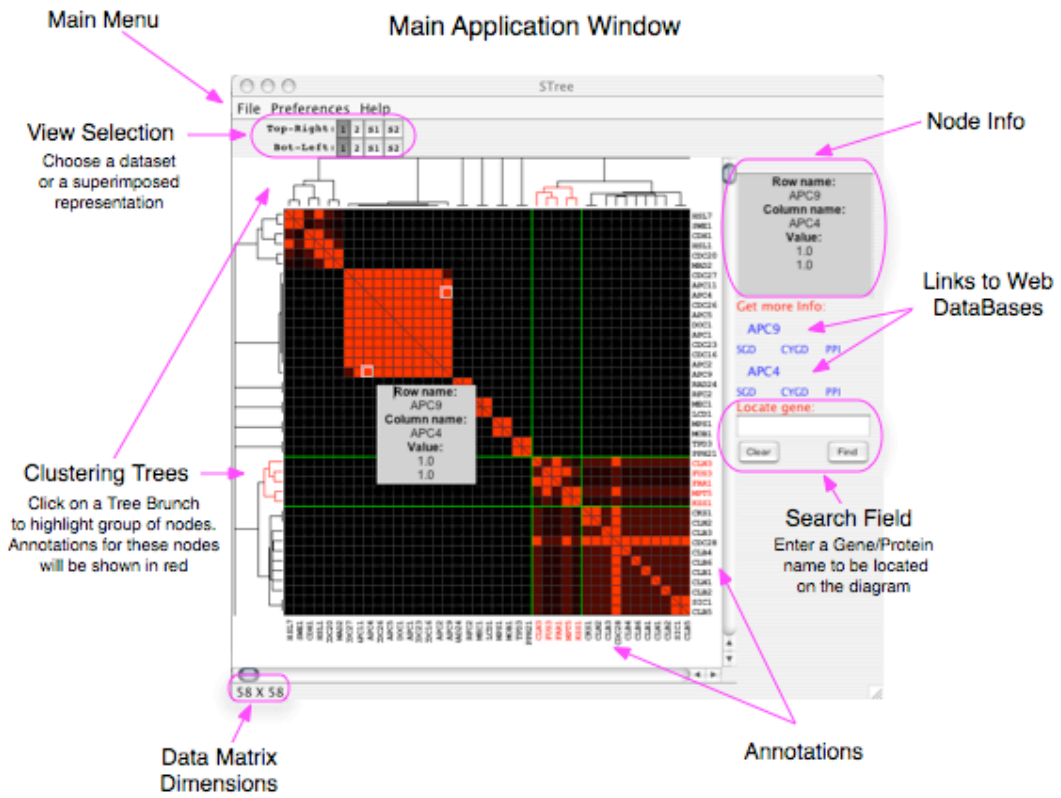


Figure 3.

the application window. Moving the scrollbars will move the heat-map and correspondingly adjust annotations and clustering trees. In order to zoom into a specific region of the complete heat-map select View > Full View. A full heat-map will be displayed and a white rectangle will highlight the region shown in the main application window. Drag the rectangle with a mouse to display a different heat-map fragment in the main application window. The navigation window may be resized by dragging its bottom-left corner. The size of a full heat-map inside the window will be changed accordingly. The original scale of a full heat-map will be preserved.

Search:

Enter your query name in the Search Field and click Find in order to highlight a row and a column with that name. To locate the name iVici will search through column and row names (see Data Formats). The search is case insensitive. Pressing the Clear button will hide search results.

View Selection:

View Selection toolbars let you choose a dataset or a superimposed view that will be displayed in the main application window (see Visualization Modes). When you load a symmetric dataset, two toolbars appear. The Top-Right toolbar is used for controlling the view in the top right triangular part of the rectangular heat-map. Bot-Left toolbar is used for controlling the view in a bottom left triangular part of the rectangular heat-map. Thus, different datasets may be simultaneously visualized on rectangular heat-map. Asymmetric dataset cannot be divided into two triangles. However, when an asymmetric

dataset is loaded, you are provided with a toolbar, which lets you specify a visualization mode for the whole heat-map.

Displaying Color Scale:

In order to display Color Scale window select View > Color Scale from iVici main manu.

Visualization Modes

General Mode:

DataSet1:

In order to display DataSet1 on a heat-map, select option 1 in the View Selection toolbars (Fig. 3). This view will appear by default after loading DataSet1.

DataSet2:

In order to display DataSet2 on a heat-map, select option 2 in the View Selection toolbars (Fig. 3). This view will appear by default after loading DataSet2.

Superimposed Mode:

S1:

In this mode, the color intensities for rendering of a heat-map are based on the values of DataSet1. The actual colors will difer from those used for DataSet1 visualization. A non-zero data value i,j , which exists only in DataSet1, will be displayed in the original DataSet1 color. Supperimposed colors will be used when a non-zero value i,j exists in both datasets (see Colors).

S2:

In this mode, the color intensities for rendering of a heat-map are based on the values of DataSet2. The actual colors will differ from those used for DataSet2 visualization. A non-zero data value i,j , which exists only in DataSet2, will be displayed in the original DataSet2 color. Superimposed colors will be used when a non-zero value i,j exists in both datasets (see Colors).

When datasets are not bound to each other, iVici will attempt to find datapoints with the same row and column names in both datasets and corresponding values will be used for superimposition.

Comparison of multiple datasets

In order to keep the visualization color scheme simple, comparative and superimposed modes are used to display two datasets at a time. However, when correlation of more than two datasets is necessary, an export feature of iVici can be used for displaying an overlap between the first two datasets in a separate application window where it can be further compared with other datasets.

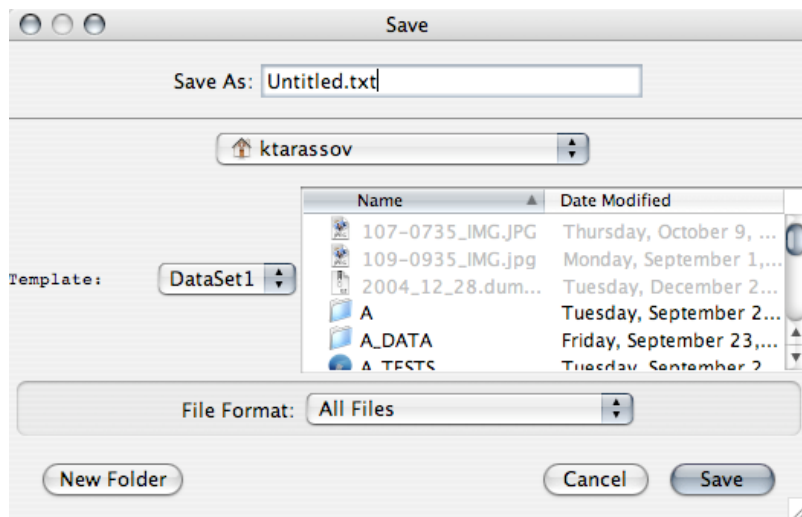
Saving the overlap:

You can save either a union or a superimposition of two datasets in a text format (see Data Formats).

When superimposition is saved, only datapoints that are present in both datasets are saved. In order to save a superimposition of two datasets, select “File > Save Data > Superimposed DataSet1” for saving only those datapoints of DataSet1 that have corresponding non-zero values in DataSet2. Select “File > Save Data > Superimposed DataSet2” for saving only those datapoints of DataSet2 that have corresponding non-zero values in DataSet1.

A union of datasets combines all non-zero values from two datasets. To save a union of two datasets select “File > Save Data > Union of DataSets.

The following dialog will allow you to choose a template dataset (Template:



DataSet1 or DataSet2).

When a template dataset is chosen it will be saved in its original rows and columns order. Data values that are not present in the template dataset but

exist in the second dataset will be added to the template dataset. If non-zero value exists in both datasets, a template dataset value will be used for saving the union of two datasets. Note that if two datasets are bound to each other (see Binding Datasets) the union will be of the same dimensions as DataSet1, because when datasets are bound iVici reorganizes DataSet2 accordingly to DataSet1.

Please refer to the original iVici paper [1] for examples on correlation of multiple datasets.

Comparative Mode for Analysis of Symmetric Datasets:

Two symmetric datasets may be visualized simultaneously by rendering these datasets on different sides of main matrix diagonal. When symmetric datasets are loaded two selection toolbars appear, which allows to choose a dataset for displaying below or above the main matrix diagonal (Fig 3). Note that for asymmetric datasets only one selection toolbar will be available, which will allow to switch between general and superimposed modes (see Visualization Modes).

Examples:

A. Symmetric datasets:

We have selected 128 proteins that are annotated as regulators of the cell cycle (SGD database). According to information in CYGD database, 58 of these proteins interact with at least one other protein within the group of cell cycle regulators. We have calculated an association matrix for the interacting proteins and applied hierarchical clustering using standard approaches [3]. The clustering output is loaded to iVici as DataSet1 (Fig. 4A). The DataSet2 consists of a table, where row and column names are identical and are in the same order as in DataSet1. A value +1 is recorder in the table in a position i,j and due to symmetry of the table in position j,i , if proteins i and j share the same annotation for physiological process [4] (Fig. 4B). In figure 4D, DataSet2 is displayed in the top right triangle and bottom left triangle is displayed in Superimposed Mode S1. In Superimposed Mode S1 you see DataSet1 colored accordingly to its correlation with DataSet2. Values i,j are rendered in red if a non-zero value i,j does not exist in DataSet2, meaning that a protein i and a protein j are not linked to exactly the same term describing a physiological process. If such correlation exists, value i,j is rendered in yellow. The color intensity for a value i,j is the same as in DataSet1 (Fig. 4A)

as well as the overall dataset structure. In figure 4E, Superimposed Mode S2 is displayed side by side with DataSet1. In this mode, non-black cells indicate that corresponding proteins are linked to the same term and yellow color is used to highlight those protein pairs, which are associated through a protein-protein interaction.

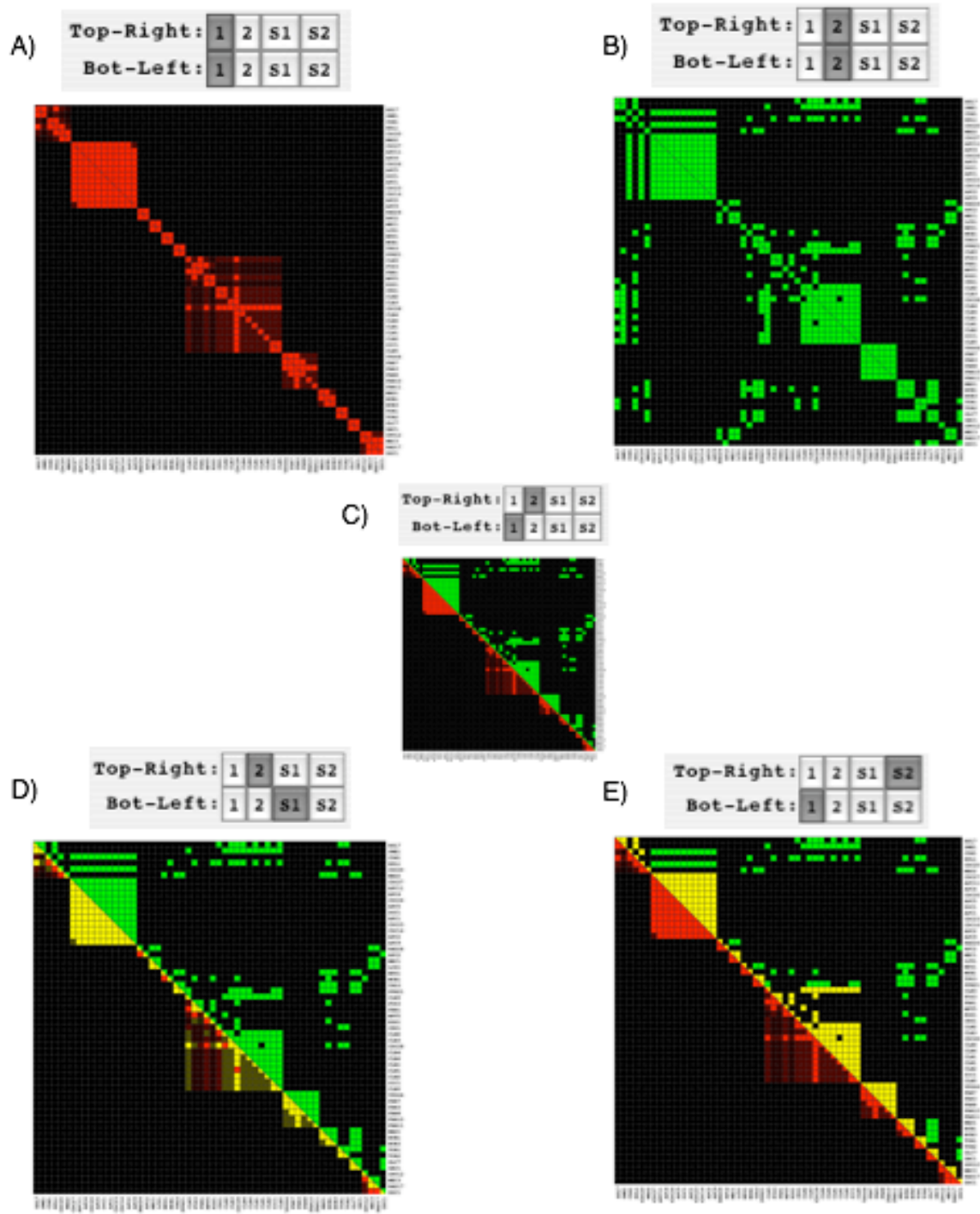


Figure 4. Visualization Modes

B. Asymmetric datasets:

Asymmetric datasets cannot be displayed on the same visualization space as shown in the above example. However, two asymmetric datasets may be visualized separately and compared in superimposed mode. Asymmetric, datasets 1 (Fig. 5A) and 2 (Fig. 5B) are computer generated asymmetric matrices. Two datasets are visualized in general mode and their superimposition is rendered in superimposed mode (Fig. 5C) using the same coloring scheme as it is described above.

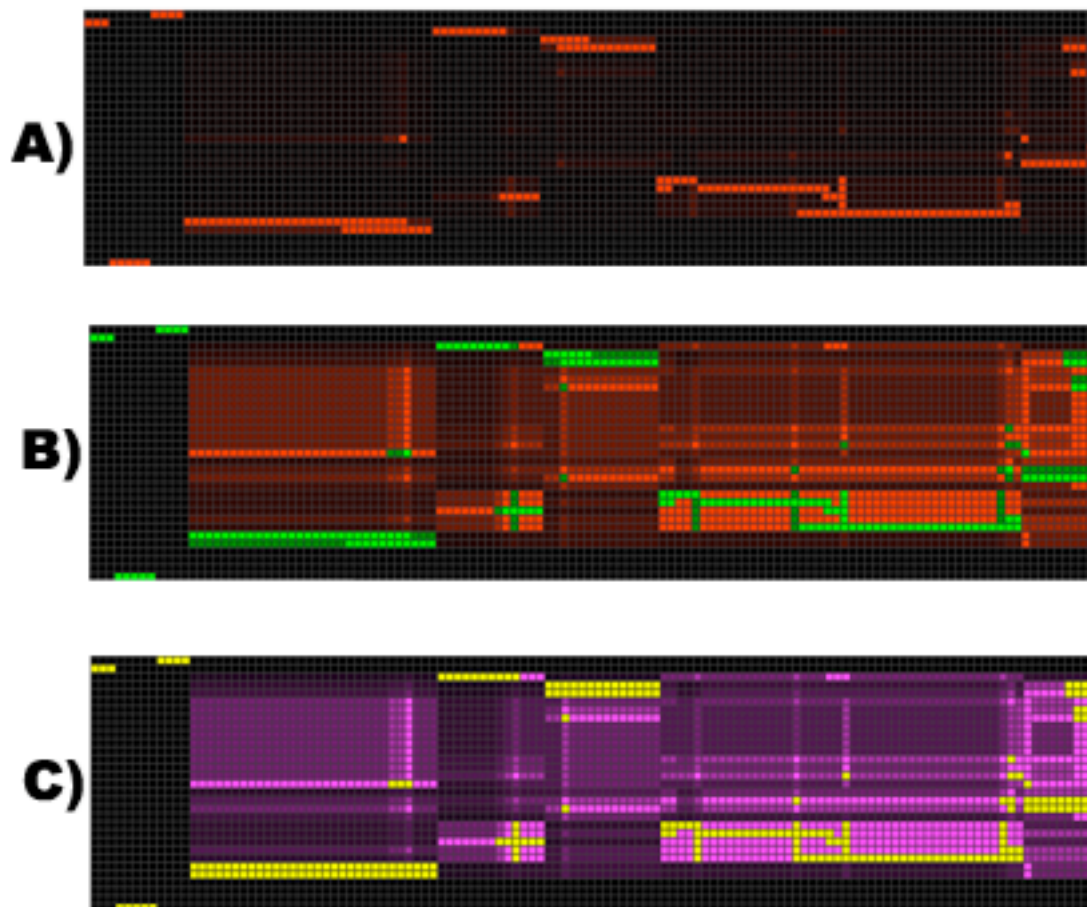


Figure 5. Asymmetric datasets

Colors

Default colors:

In this section a default selection of colors is explained. When both datasets contain values of the same sign within the dataset, DataSet1 will be rendered in red, DataSet2 will be rendered in green and yellow will be used as the superimposed color. The color will not depend on the actual signs (+/-) of the values.

If one or both datasets contain positive and negative values, both datasets will be rendered in red (for positive values) and green (for negative values). In superimposed modes, magenta is used for rendering of correlated positive values, cyan for correlation of negative values and yellow for correlation of values with different sign. Despite usage of the same colors for the datasets in comparative mode they are discriminated by plotting them on different sides of the main diagonal of the heat-map.

User defined colors:

In order to define your own colors, select Preferences > Colors and the dataset or superimposed colors you'd like to change. You will be provided with choices of colors for rendering positive, negative and superimposed values (Fig. 6).

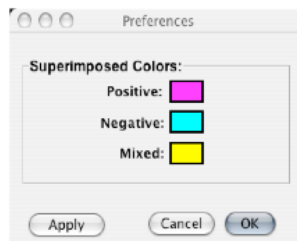
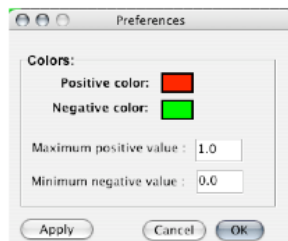
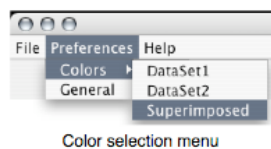


Figure 6. Color selections

Display Options

Available options for controlling visualization parameters are shown in Figure 7. It can be accessed by selecting Preferences > General.

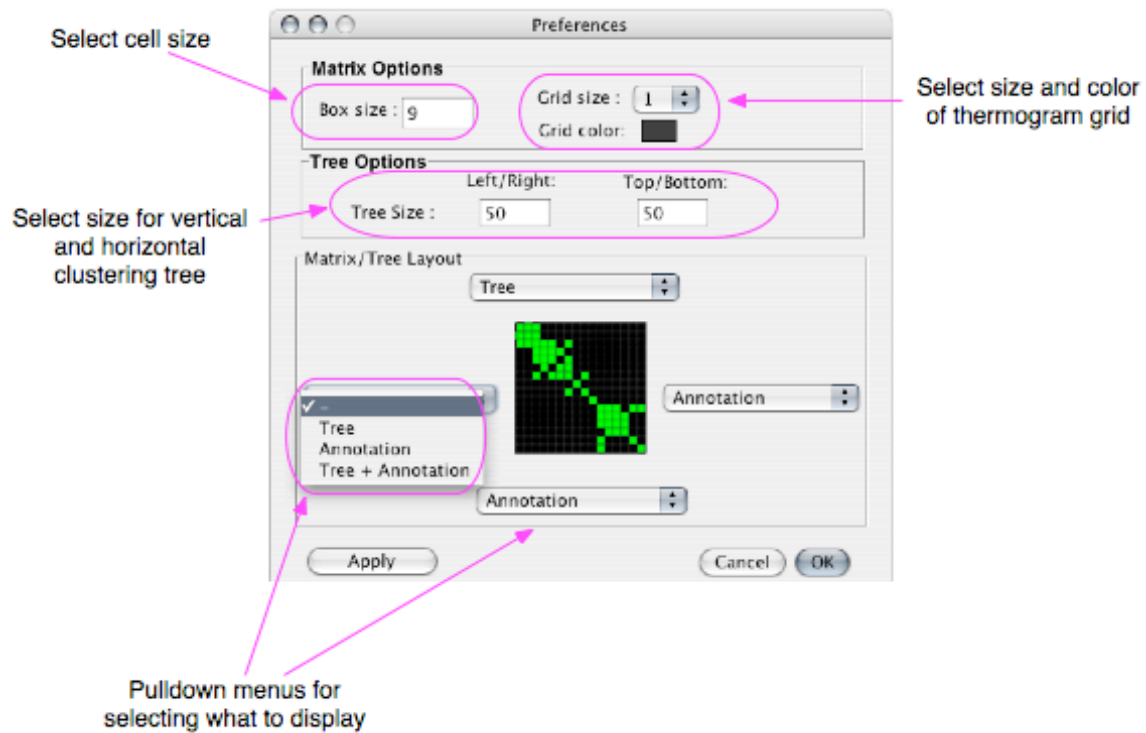


Fig. 7. General Visualization Preferences

Images can be saved in two formats: png and postscript. Please, chose “File > Save Image” for saving a png image or “File > Export to PostScript” to save an image in postscript format.

References:

1. Tarassov K. and Michnick SW., *iVici: interrelational visualization and correlation interface*. Submitted to Genome Biology.
2. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
3. Rives, A.W. and T. Galitski, *Modular organization of cellular networks*. Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1128-33.
4. GO Consortium. *Gene ontology: tool for the unification of biology*. Nat. Genet., 2000. 25: p. 25-29.