



VNIVERSIDAD
D SALAMANCA



VisBicluster

User's Guide Version 1.5

Haithem Aouabed^{1,3} Rodrigo Santamaría² and Mourad Elloumi¹

¹ Laboratory of Technologies of Information and Communication and Electrical Engineering (LaTICE), National Superior School of Engineers of Tunis (ENSIT) University of Tunis, Tunis, Tunisia.

²Departamento de Informática y Automática, Universidad de Salamanca, Salamanca, Spain.

³Faculty of Economic Sciences and Management of Sfax (FSEGS), University of Sfax, Sfax, Tunisia.

June 20, 2019

VisBicluster is an interactive, web based visualization technique designed to analyze biclusters generated from gene expression data. VisBicluster visualizes both, bicluster intersections and their properties, and the elements (genes and conditions as a heatmap) of each intersection or each bicluster.

1) Biclustering algorithms execution with R

A simple R script stored in the *R_code* folder, is used to simplify the loading of a biclustering result. It is based on **biclust R package** (Kaiser et al., 2015) to run different biclustering algorithms and **writeBiclusterResults method** (Santamaría, 2015) to write biclustering result. The example in the script is about running Xmotifs biclustering algorithm.

```
#install.packages("biclust")

require(biclust)

#usage example with Xmotifs biclustering algorithm

data = read.table("gene_expression_file_name.txt")

exprs <- as.matrix(data)


x <- discretize(exprs) #for Xmotifs biclustering execution

res <- biclust(x, method=BCXmotifs(), ns=20, nd=20, sd=5, alpha=0.01,
              number=10) # Xmotifs as an example

writeBiclusterResults("results_B.bic", res,"Xmotifs with alpha=0.5",
                      dimnames(x)[1][[1]],dimnames(x)[2][[1]])
```

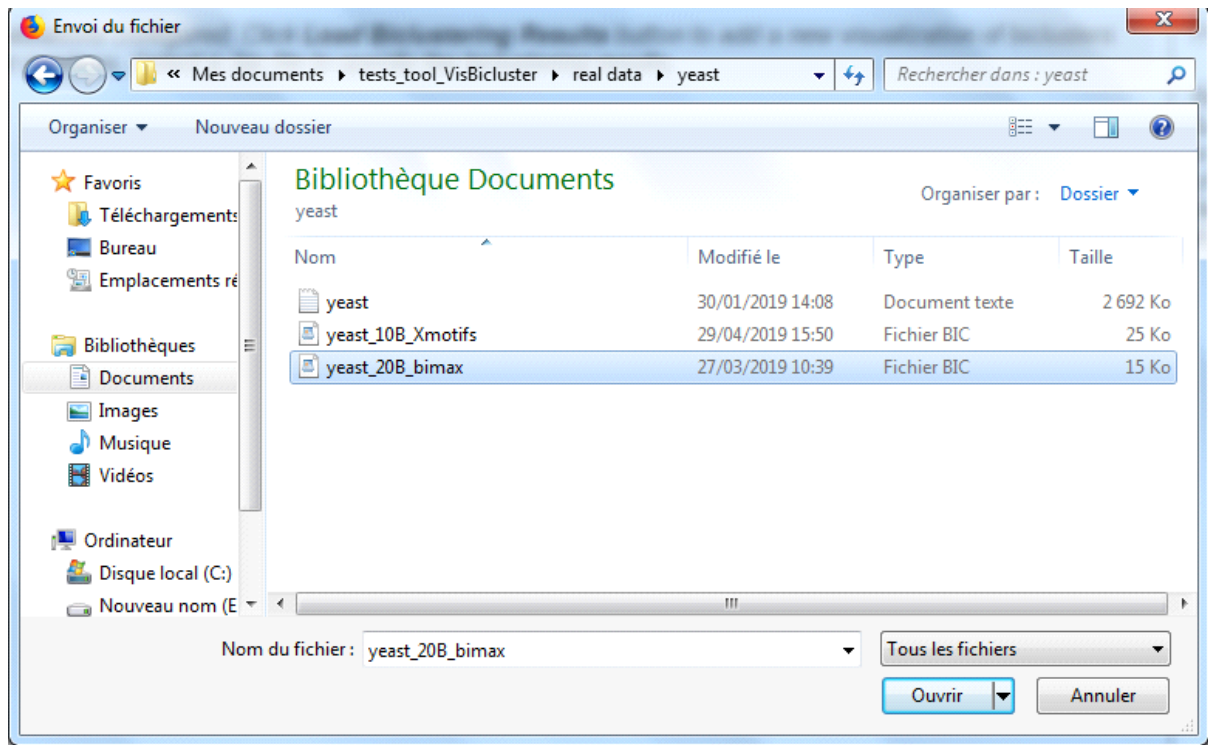
2) Data loading

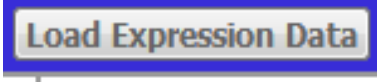
The loading process start by clicking on the **Load Biclustering Results** button



that gives the possibility to select the file

which contains the list of biclusters. The selected file has .bic as an extension. In the example below, the .bic file is *yeast_20B_bimax*.



The **Load Expression Data** button  is used to select a text file (.txt) that represents the original expression data used in the biclustering process. These data are important to draw biclusters and their overlaps as heatmaps.

We are going to analyze the data from a microarray data experiment generated by from *yeast Saccharomyces cerevisiae* data (2993 rows and 173 columns). Then, we executed *Bimax* biclustering algorithm on it to generate 20 biclusters.

3) Biclusters and their overlaps visualization

Intersections between biclusters are visualized as a matrix in the **Matrix of Overlaps Visualization** part. The rows represent the possible overlaps where

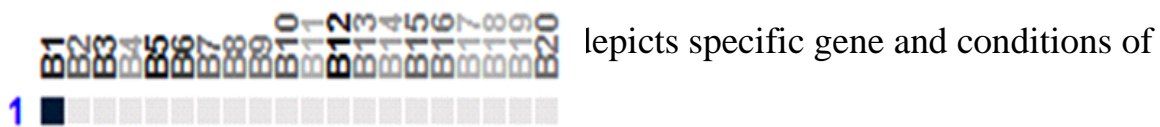
the columns depict the biclusters. Each cell in the matrix represents the list of biclusters included in a specific overlap.

Matrix of Overlaps Visualization



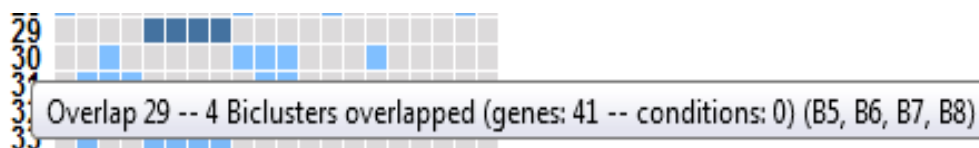
The generated matrix includes 20 biclusters with 53 overlaps.

Rows that are colored in blue represent specific data (genes and conditions in gene expression case) of a defined biclusters. for example, line 1

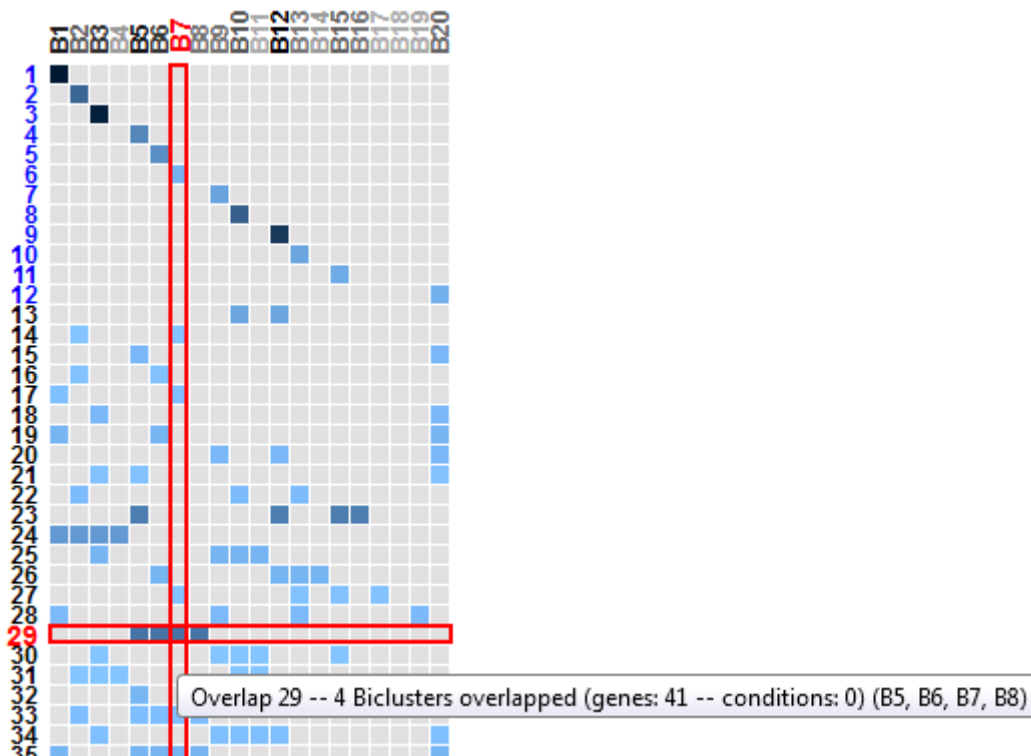


bicluster1.

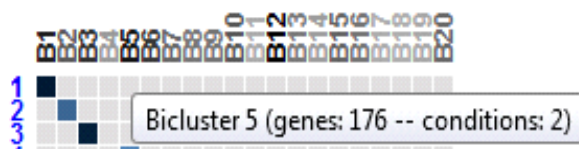
The other rows show overlaps between 2 or more biclusters. Cells of each overlap are colored according to a scale. Cells in a light blue define an overlap with small size where cells colored in dark blue show overlaps with big size. For example, overlap 20 between biclusters 9, 12 and 20 is a small one where overlap 29 between biclusters 5, 6, 7 and 8 has a big size. In order to know the interesting information of each overlap (i.e., number of overlap, total number of overlapped biclusters, size and the list of overlapped biclusters), the user can hover over either the row or one of cells of the corresponding overlap



The hover over of cell placed the corresponding row and column in a red rectangle.



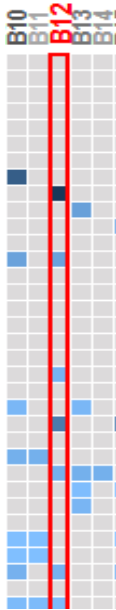
Names of columns are also colored with a scale. So, light black represents biclusters with small sizes where dark black represents biclusters with big sizes. Hover over the title of a columns gives the full name of a bicluster and its size (number of genes and conditions in microarray data case).



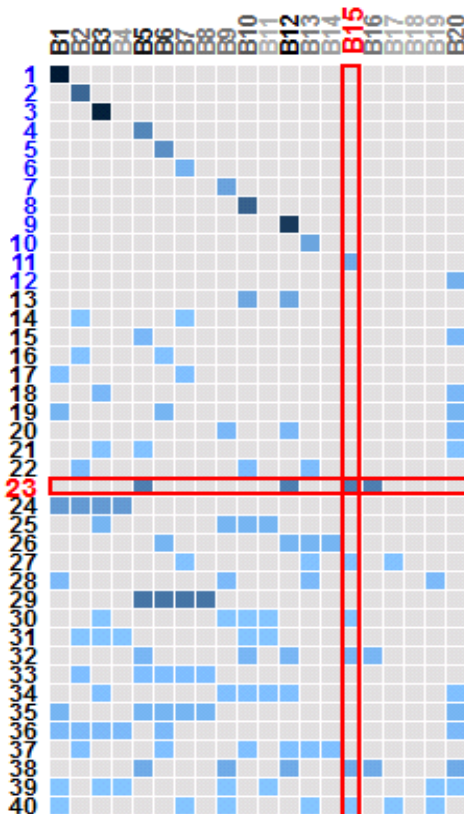
Bicluster 5 has 176 genes and 2 conditions.

In order to show the corresponding list of genes and conditions of either a bicluster, an overlap or a set of overlaps as a heatmap representation, the user must:

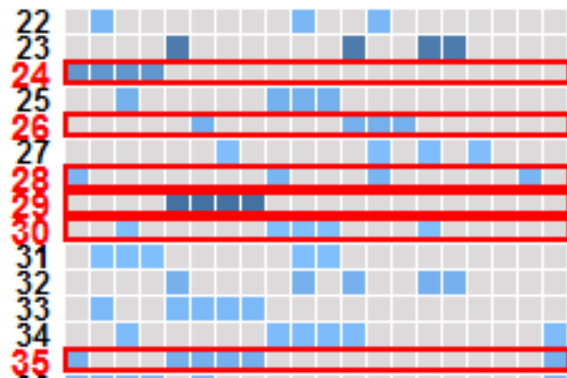
- Click on the name of a column to show genes and conditions of a bicluster. The corresponding column will be placed in a red rectangle.



- Click on a cell or double-click on the number of a row to show genes and conditions as a heatmap of an overlap. The corresponding row and column are placed in red rectangles if clicking on a cell where only row is placed in red rectangle when clicking on the number of row.

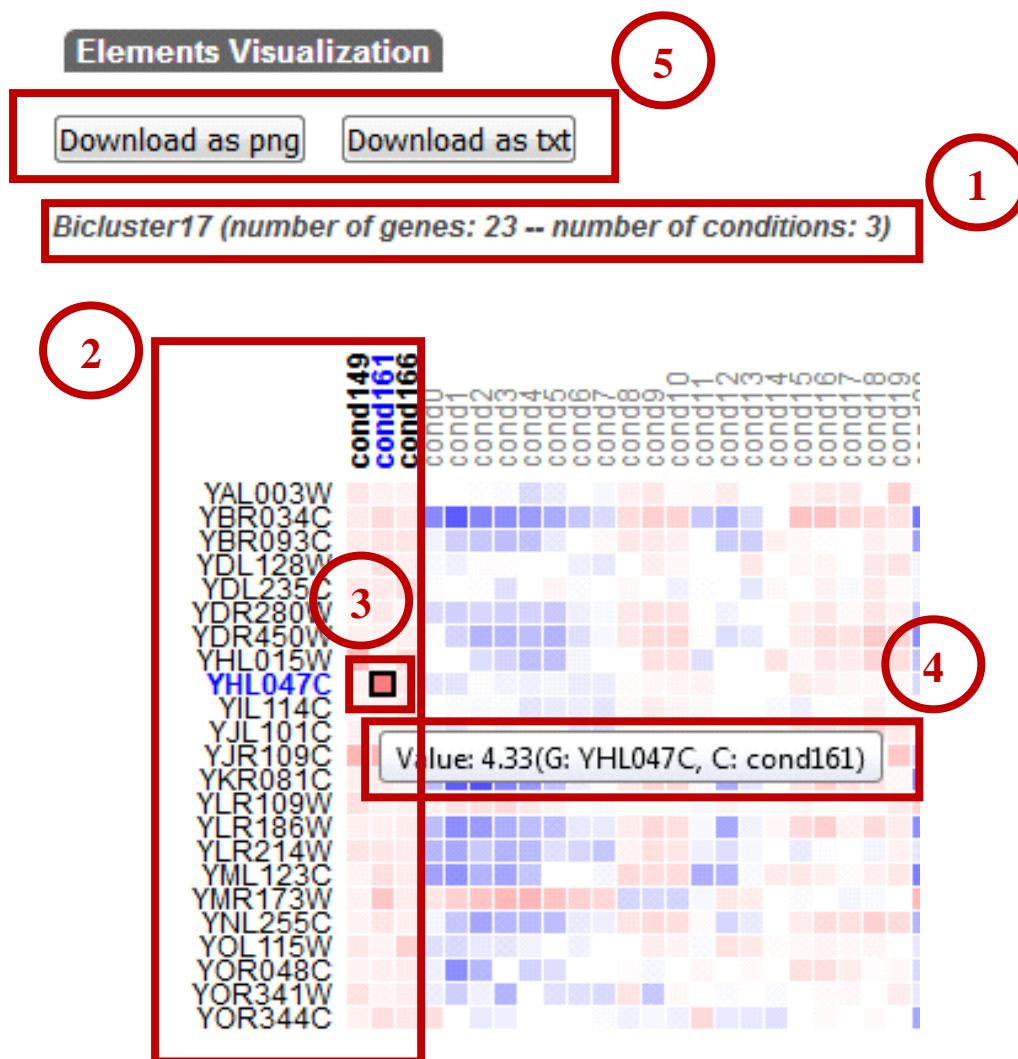


- Click one by one on a list of number of rows to allow the selection of more than one overlap in the same time and double-click on the number of one of the selected rows to show corresponding genes and conditions as a global heatmap (i.e., more than one overlap). All selected rows are placed in red rectangles.



4) Elements visualization

Genes and conditions either of an overlap, a list of overlaps or a bicluster with their corresponding gene expression values are shown by a heatmap in the **Elements Visualization** part. The following example shows the heatmap of bicluster number 17.



- 1 Description of the expression data to show (name of corresponding element to show based on the matrix of overlaps; bicluster, one overlap or set of overlaps and the size; number of genes and conditions).
- 2 Gene expression matrix of bicluster 17. The corresponding conditions are in bold.
- 3 A cell hovered. Its corresponding gene and condition are bold and colored in blue.

4

Corresponding gene expression value as well as name of condition and gene of the hovered cell.

5

The visualized heatmap can be downloaded as an image file (.png) as well as a text file (.txt) by the user on clicking on either **Download as png** or **Download as txt** buttons.

5) Data Analysis

The matrix of biclustering overlaps may suffer **different analysis** due to the **Data Analysis** part of the VisBicluster tool.

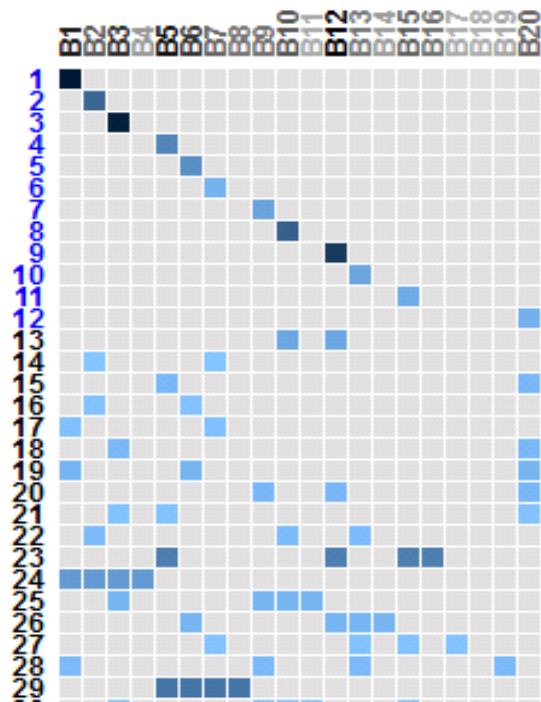
The matrix can be sorted either by the *number of overlapped biclusters* or by the using of a clustering algorithm (i.e., *Levenshtein distance*) which represents a string metric for measuring the difference between two sequences. This algorithm is used to show the closest overlaps next to each other and to define clusters from the overlapped rows.

Sort by

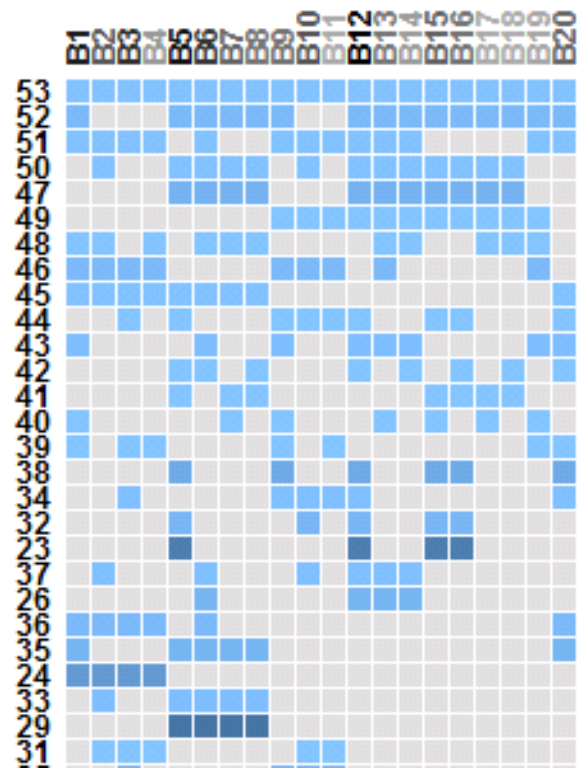
☒ Number of overlapped biclusters

☐ Clustering based on
Levenshtein Distance

Part of matrix of overlaps sorted by number of overlapped biclusters



Same matrix but sorted by Levenshtein distance

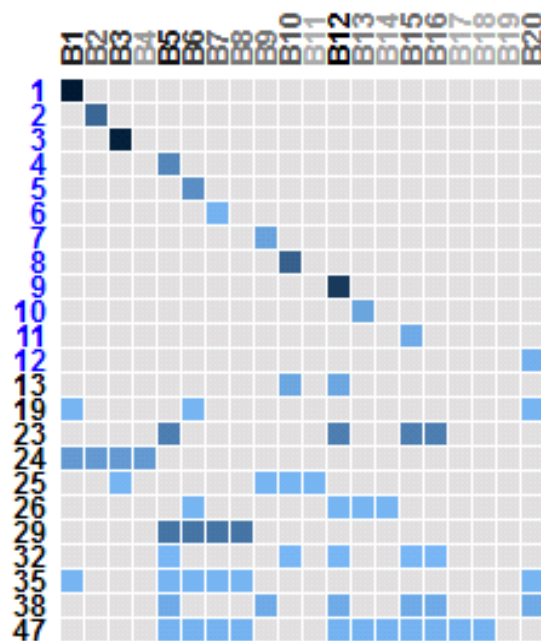


The post filtering of the matrix of overlaps can be done on rows or columns (i.e., biclusters). The user can:

- Drop rows (i.e., overlaps) with defined parameters (less or equal to a number of genes/conditions). For example, drop rows with less than 7 genes).

☒ Drop overlaps with number of
genes ▼
≤ 7 ▲▼

Matrix of Overlaps Visualization



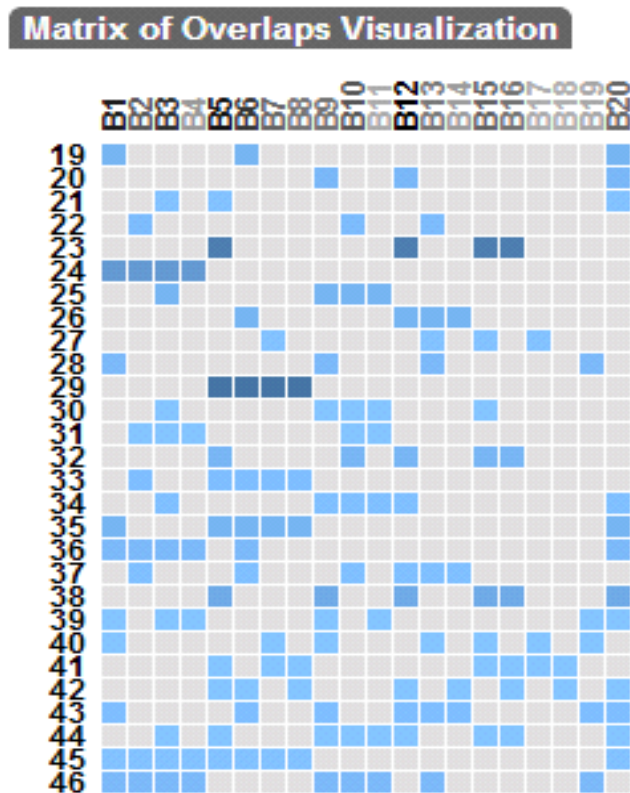
- Show only overlaps with number of overlapped biclusters less than or equal to a given number.

☒ Show only overlaps with
number of overlapped biclusters
≤ 9

- Show only overlaps with number of overlapped biclusters greater than or equal to a given number.

☒ Show only overlaps with
number of overlapped biclusters
≥ 3

Next figure shows overlaps between 3 or 9 overlapped biclusters.



- Drop biclusters with more/less/equal to a given number of genes or conditions. The next example shows the new matrix of overlaps after removal of biclusters with number of genes less than 100.

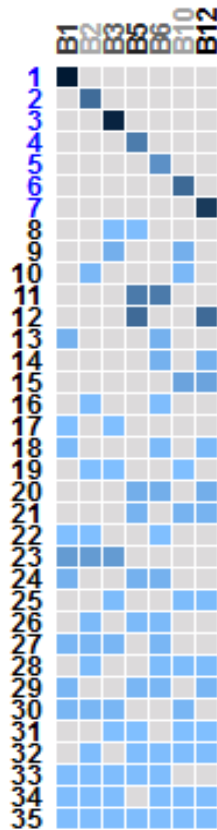
☒ Drop biclusters with

< ▾

100

genes ▾

Matrix of Overlaps Visualization



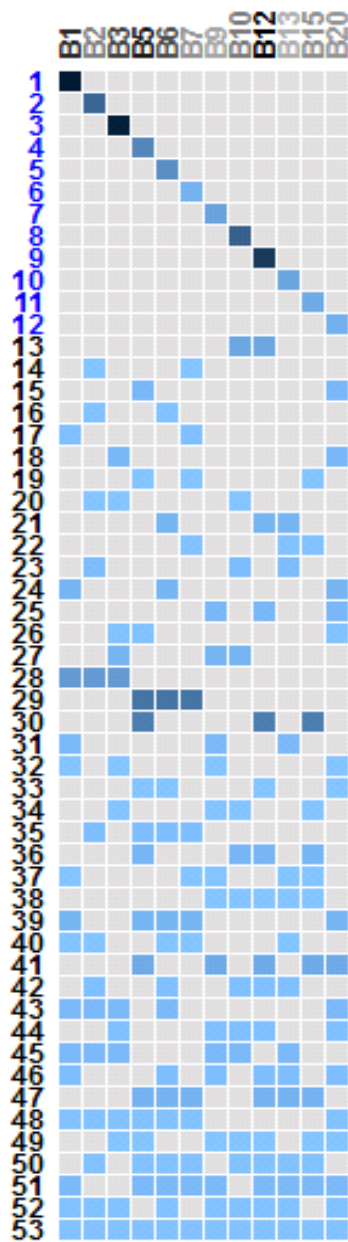
- Drop biclusters with more/less/equal to a given rate of overlap with other biclusters. The next example shows the new matrix of overlaps after removal of biclusters with rate of overlap less than 25%.

☒ Drop biclusters with

< ▾


25 %

overlap (with other biclusters)



The user can click on  button to zoom in.

 button to zoom out.

The click on  button shows the original matrix of overlaps and remove all applied data analysis.

The names of file storing biclusters (.bic) as well as the name of gene expression file (.txt) and its total number of rows and columns are shown down exactly in the **Data Information** part.

Data Information:
.bic File: yeast_20B_bimax.bic
.txt File: yeast.txt
Number of rows: 2993
Number of columns: 173

6) Setting up VisBicluster

Demo:

We are hosting a demo instance of *VisBicluster* at <http://.....>

Local Deployment

1. Clone the repository using git clone or download and extract the ZIP file at <https://.....>
2. Launch the [Python SimpleHTTPServer] (<https://docs.python.org/2/library/simplehttpserver.html>) in the project directory.

```
$ python -m SimpleHTTPServer 8000
```
3. View VisBicluster in your browser at (<http://localhost:8000>) .

Alternatively you can also ****run VisBicluster without a web server****. Simply open the **index.html** file in **Firefox** or **Google Chrome**.

Optionally, you can download some preprocessed examples at <http://.....>