

BCTool: a biclustering comparison toolbox

Fadhl M. Al-Akwaa^{1,*}, Mohammed M. Alkhawlan² and Hussein H. Mazzar³

¹Biomedical Engineering Department, University of Science & Technology, Sana'a, Yemen.

²Electronics Engineering Department, University of Science & Technology, Sana'a, Yemen.

³Faculty of Computer Science, Cairo University, Giza, Egypt.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

ABSTRACT

Summary:

BCTool is an open-source software tool for comparing and evaluating the results of biclustering algorithms. BCTool helps researchers to evaluate the results of different biclustering methods, compare the results against each others, and allow viewing the comparison results via convenient graphical displays. The objects in this toolbox are open source so users may modify the program and add further algorithms or extensions.

Availability: BCTool is a Matlab toolbox and requires Matlab to run. The BCT toolbox and user manual are freely available at <http://fadhl-alakwa.weebly.com/bctool.html>.

Contact: fadlwork@gmail.com

1 INTRODUCTION

Microarray experiments have been used to measure gene expression levels under different cellular conditions or along certain time course. The analysis of microarrays data poses a large number of exploratory statistical aspects including clustering and biclustering algorithms which help to identify similar patterns in gene expression data and group genes and conditions into subsets that share biological significance. There are several biclustering methods that have been proposed to achieve this target (Cheng and Church, 2000; Liu and Wang, 2007; Madeira and Oliveira, 2004), but the question is: which algorithm is better? Do some algorithms have advantages over others?

Generally, comparing different biclustering algorithms is not straightforward as they differ in strategies, approaches, time complexity, number of parameters and prediction ability. Also they are strongly influenced by user-selected parameter values (Prelic, et al., 2006). Desirable is a software tool that offers different biclustering comparison methodology within a common framework—to our best knowledge, such a tool has not been available so far.

BCTool tries to fill this gap and provides the following functionality:

- (1) Evaluating the quality of each biclustering algorithm results after applying the Gene Ontology (GO) functional analysis and displaying the percentage of the enriched biclusters at the standard p-values (significance levels) which are: 0.00001, 0.00005, 0.0001, 0.0005, 0.001, 0.005, 0.01 and 0.05.
- (2) Comparing between the different biclustering algorithms according to the percentage of the functionally enriched biclusters at the required significance levels, the selected GO category and with certain filtration criteria for the GO terms.
- (3) Evaluating and comparing the results of external biclustering algorithms. This gives the BCTool the ability to be a generic tool that does not depend on the employed methods only. For example; it can be used to evaluate the quality of the new algorithms introduced to the field and compare against the existing ones.
- (4) Displaying the analysis and comparing results using graphical and statistical charts visualizations in multiple modes.

2 METHODS

Our tool is based on GO assessment to investigate whether the set of genes discovered by biclustering methods present significant enrichment with respect to a specific GO annotation provided by the Gene Ontology Consortium (Ashburner, et al., 2000).

BCTool provides reasonable methods (Figure 1) for comparing the results of different biclustering algorithms by:

1. Identifying the percentage of enriched or overrepresented biclusters for each algorithm with one or more GO term per multiple significance levels (p-values). (Option 1 & 2 in Figure 1).

The definition of significance depends on the user selection of the threshold p-values. A bicluster is said to be significantly overrepresented (enriched) with a functional category if the p-value of this functional category is lower than the preset threshold P-value. The results are displayed using a histogram for the entire compared algorithms at the different preset significance levels. The algorithm which gives higher proportion of enriched biclusters per all significance levels is considered to be the optimum one as it does group effectively the genes sharing similar functions in the same bicluster (Figure 2).

*To whom correspondence should be addressed.

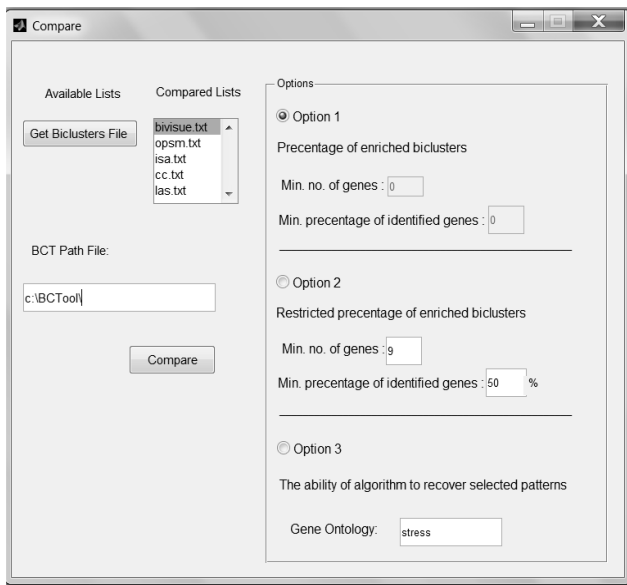


Fig. 1. Graphical user interface of the BCTool software. The window on the left hand side displays the compared list of the biclustering algorithms. The panel on the right shows the available options for the comparison.

2. Estimate algorithms predictability power to recover user interested pattern (option 3 in Figure 1). Genes whose transcription is responsive to a variety of stresses have been implicated in a general yeast response to stress. Other gene expression responses appear to be specific to particular environmental conditions (Gasch, et al., 2000). BCTool make the user able to compare the predictability power of biclusters algorithms to interested pattern defined by the user.

BCTool has been developed in Matlab. Besides biclusters comparison, BCTool also provides functions for running three different biclustering algorithms, exporting and filtration the results and visualization the results using heatmap.

3 RESULTS

For demonstration purposes, the BCTool program is applied on yeast *Saccharomyces Cerevisiae* cell cycle dataset (Gasch, et al., 2000), which consists of 2993 genes and 173 conditions. The dataset contains of diverse environmental transitions such as temperature shocks, amino acid starvation, and nitrogen source depletion. Figure 2 shows the comparison charts for 5 different biclustering algorithms using BCTool option1(Figure 1).

Investigating Figure 2, we observed that OPSM provides a high portion of functionally enriched biclusters at all significance levels. The BCTool user manual contains the details of the other BCTool comparison mythologies.

4 CONCLUSION

An open-source software tool known as BCTool for comparing and assessing biclusters from a gene expression matrix is described. BCTool allows viewing the comparison results via convenient graphical displays.

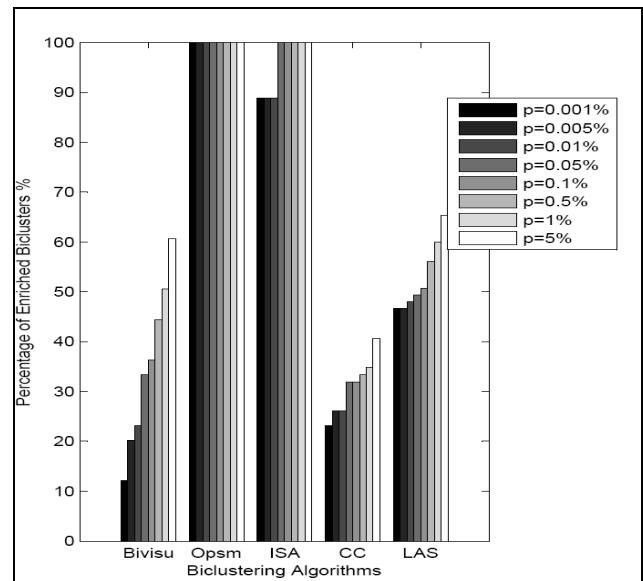


Fig. 2. Percentage of biclusters significantly enriched by GO Biological Process category (*S. Cerevisiae*) for the five selected biclustering methods at different significance levels p .

BCTool enables researchers and biologists to compare between the different biclustering methods based on set of biological merits and draw conclusion on the biological meaning of the results. The toolbox is described in the context of gene expression analysis, but is also applicable to other types of data, e.g. data from proteomics or synthetic lethal experiments.

ACKNOWLEDGEMENTS

This work is supported by the Centre for Bioinformatics, Department of Biomedical Engineering, University of Science and Technology, Yemen.

Conflict of Interest: none declared.

REFERENCES

- Ashburner, M., et al. (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nature Genetics*, **25**, 25 - 29.
- Cheng, Y. and Church, G.M. (2000) Biclustering of expression data, *Proceedings of 8th International Conference on Intelligent Systems for Molecular Biology*, 93 - 103.
- Gasch, A.P., et al. (2000) Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes, *Mol. Biol. Cell*, **11**, 4241-4257.
- Liu, X. and Wang, L. (2007) Computing the maximum similarity bi-clusters of gene expression data, *Bioinformatics*, **23**, 50-56.
- Madeira, S.C. and Oliveira, A.L. (2004) Biclustering algorithms for biological data analysis: a survey, *IEEE/ACM Trans Comput Biol Bioinform*, **1**, 24 - 45.
- Prelc, A., et al. (2006) A Systematic comparison and evaluation of biclustering methods for gene expression data, *Bioinformatics*, **22**, 1122 - 1129.