BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 12 2014, pages 1785–1786
doi:10.1093/bioinformatics/btu120

Gene expression

Advance Access publication March 3, 2014

BicOverlapper 2.0: visual analysis for gene expression

Rodrigo Santamaría[1,*], Roberto Therón[1] and Luis Quintales[1,2]

[1]Department of Computer Science, University of Salamanca, 37008 Salamanca, Spain and [2]Instituto de Biología Funcional y Genómica, CSIC/USAL, 37007 Salamanca, Spain

Associate Editor: Janet Kelso

ABSTRACT

Motivation: Systems biology demands the use of several point of views to get a more comprehensive understanding of biological problems. This usually leads to take into account different data regarding the problem at hand, but it also has to do with using different perspectives of the same data. This multifaceted aspect of systems biology often requires the use of several tools, and it is often hard to get a seamless integration of all of them, which would help the analyst to have an interactive discourse with the data.

Results: Focusing on expression profiling, BicOverlapper 2.0 visualizes the most relevant aspects of the analysis, including expression data, profiling analysis results and functional annotation. It also integrates several state-of-the-art numerical methods, such as differential expression analysis, gene set enrichment or biclustering.

Availability and implementation: BicOverlapper 2.0 is available at: http://vis.usal.es/bicoverlapper2

Contact: rodri@usal.es

Received on November 19, 2013; revised on January 27, 2014; accepted on February 19, 2014

1 INTRODUCTION

BicOverlapper 1.0 (Santamaria *et al.*, 2008) focused on the visualization of complex gene expression analysis results coming from biclustering algorithms. Based on Venn-like diagrams and overlapping visualization layers, it successfully conveyed biclusters. With the use of BicOverlapper by the authors and third-party users, several new requirements arose, and it has evolved to support other analysis techniques and additional steps of the analysis process. Similar evolutions have occurred on other tools on the field. For example, Expander has extended microarray data analysis with relational and functional information (Ulitsky *et al.*, 2010). Hierarchical Clustering Explorer, although originally designed for general use, added new methods for bioinformatics analysis (Seo *et al.*, 2006). Treeview (Saldanha, 2004) is developing toward a new version that will address high-throughtput biology needs (see https://www.princeton.edu/~abarysh/treeview/).

2 APPROACH

During the design of BicOverlapper 2.0, we focused on a high level of interaction and a visual analytics (Thomas and Cook, 2005) approach. Another important design principle was the simplification of installation and interfaces. Finally, following the original 'overlapping' philosophy, we designed linked visualizations and an agglomerative use of standard numerical analyses. For example, differential expression analysis compares two experimental conditions, but BicOverlapper 2.0 allows to compare several combinations of experimental conditions at once and then to visualize the relationships between the differentially expressed groups.

3 METHODS

The tool is implemented as two interconnected layers: visualization and analysis. The analysis layer is R/Bioconductor-dependent, using several packages and *ad hoc* scripts. Data retrieval from Gene Expression Omnibus (GEO) and ArrayExpress is supported by its corresponding packages (Davis and Meltzer, 2007; Kauffmann *et al.*, 2009), although it requires high bandwith and not all of the experiments are supported. Data analysis includes the following:

- Differential expression with *limma* (Smyth, 2005). In addition to one-to-one comparisons, BicOverlapper allows to perform multiple comparisons at once, visualized as intersecting differentially expressed groups. This way, analysis time is reduced, and the differences between the comparisons can be inspected.
- Gene set enrichment analysis is also implemented via *GSEAlm* (Oron and Gentleman, 2008). Enriched gene sets are visualized as overlapping groups.
- Biclustering, as in the previous version, is computed with *biclust* (Kaiser *et al.*, 2013) package. The Iterative Search Algorithm (ISA) algorithm is now also available by the *isa2* package.
- Correlation networks. This is a simple yet powerful method to find groups. Genes with low overall expression variation are filtered out, and the rest are linked if they have a profile distance below some standard deviations. The resulting network is visualized as a force-directed layout, where nodes can be colored by the expression under selected conditions.

The visualization layer is developed in Java and it communicates with the analysis layer via rJava (Urbanek, 2007). This layer contains several visualization techniques, with implementations based on Prefuse (Heer *et al.*, 2005) (networks, scatterplots), Processing (Reas and Fry, 2007) (overlapper, heatmap) and plain Java (parallel coordinates, word clouds).

4 RESULTS

To involve biology specialists on bioinformatics analyses, we need simpler and highly interactive tools. For example, Figure 1 was generated only by clicking two menu options and selecting one visual item and gene/condition labels, on a process that takes not more than 5 min (see Supplementary Video at http://vis.usal.es/bicoverlapper2/docs/tour.mp4). Underneath, this requires the seamless connection of different

*To whom correspondence should be addressed.

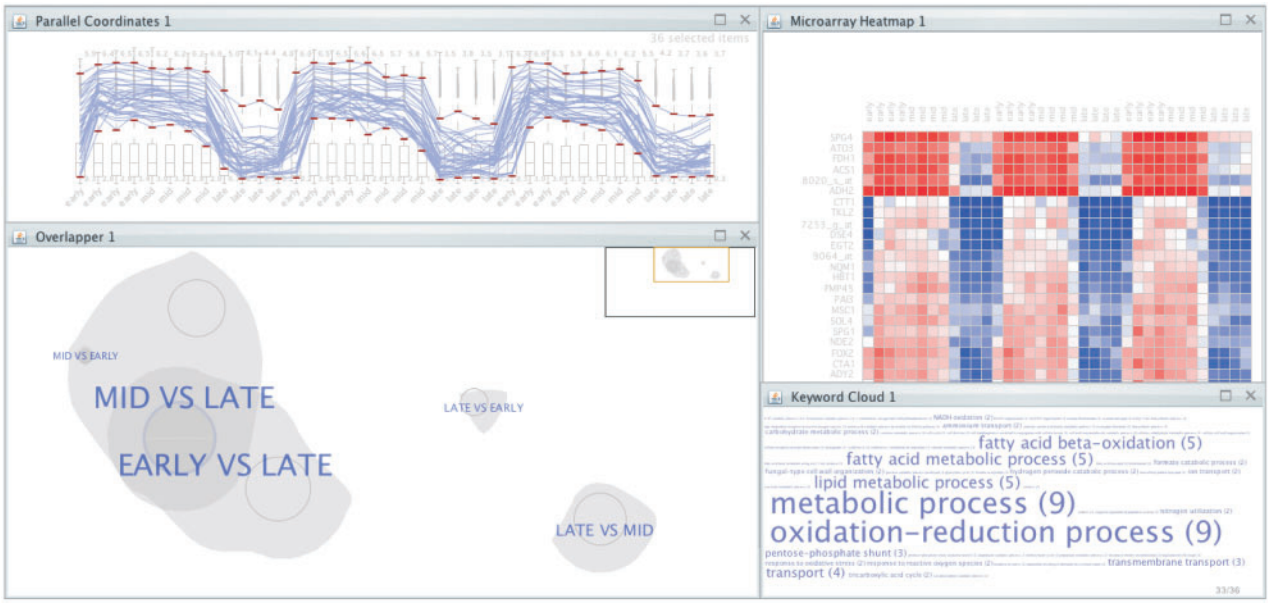

Fig. 1. Yeast gene expression profile along three cell cycles, from experiment GSE3431 (Tu *et al*., 2005). Each cell cycle is divided into three time intervals (early, mid and late). Differential expression for every combination of such intervals is computed and visualized as overlapping groups. Thirty-six genes high-regulated at early and mid intervals have been selected (intersection between 'early versus late' and 'mid versus late' groups at the bottom left); their expression profiles are shown in parallel coordinates and heatmap visualizations. Finally, the functional annotations, stacked by term, are shown as a word cloud, indicating, for example, that 9 of the 36 genes are related to metabolic and oxidation–reduction processes

steps: expression data loading, computation of distribution statistics, three differential expression analyses (for up- and downregulation), gene annotation retrieval and the visualization of four interactive representations.

Figure 1 provides a considerable amount of information about the experiment. First, parallel coordinates (Inselberg, 2009) indicate with boxplots that the data are normalized, although probably skewed towards upregulation. Second, differential expression groups, displayed as Venn diagrams, present a large overlap for genes upregulated at mid and early timepoints with respect to late timepoints. These intersecting genes have a clear pattern under heatmap and parallel coordinates and include nine genes related to the Gene Ontology (GO) terms 'oxidation–reduction process' and five related to 'fatty acid beta-oxidation'.

5 CONCLUSION

BicOverlapper is a simple-to-use, highly visual and interactive tool for gene expression analysis. Easily and without programming knowledge, the user can have an overall view of several expression aspects, from raw data to analysis results and functional annotations. This may significantly reduce the analysis time and improve the analytical discourse with the data. For the future, we are working on the support of high-throughput data, especially RNA-Seq and a comprehensive report and image generation.

Funding: This work was supported by the Spanish Government, under the Ministerio de Economía y Competitividad-MINECO (projects BFU2011-28804 and Consolider-Ingenio CSD007-00015) and by the Ministerio de Ciencia e Innovación - MICINN (project FI2010-16234)

Conflict of Interest: none declared.

REFERENCES

Davis,S. and Meltzer,P.S. (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846–1847.

Heer,J. *et al.* (2005) Prefuse: a tool for interactive information visualization. In: *CHI '05 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, pp. 421–430.

Inselberg,A. (2009) *Parallel Coordinates Visual Multidimensional Geometry and its Applications*. Springer Science+Business Media, New York.

Kaiser,S. *et al.* (2013) biclust: BiCluster Algorithms. R package version 1.0.2.

Kauffmann,A. *et al.* (2009) Importing ArrayExpress datasets into R/Bioconductor. *Bioinformatics*, **25**, 2092–2094.

Oron,A. and Gentleman,R. (2008) GSEAlm: linear model toolset for gene set enrichment analysis. Bioconductor package version 1.20.0.

Reas,C. and Fry,B. (2007) *Processing: A Programming Handbook for Visual Designers and Artists*. Mit Press, Cambridge, Massachusetts.

Saldanha,A.J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.

Santamaria,R. *et al.* (2008) BicOverlapper: a tool for bicluster visualization. *Bioinformatics*, **24**, 1212–1213.

Seo,J. (2006) An interactive power analysis tool for microarray hypothesis testing and generation. *Bioinformatics*, **22**, 808–814.

Smyth,G.K. (2005) limma: Linear Models for Microarray Data. Bioconductor package version 3.16.8.

Thomas,J.J. and Cook,K.A. (2005) *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Press, Richland, Washington.

Tu,B.P. *et al.* (2005) Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*, **310**, 1152–1158.

Ulitsky,I. *et al.* (2010) Expander: from expression microarrays to networks and functions. *Nat. Protoc.*, **5**, 303–322.

Urbanek,S. (2007) rJava: Low-level R to Java interface. R package version 0.9.4.