

# BIGGESTS: INTEGRATED ENVIRONMENT FOR BICLUSTERING ANALYSIS OF TIME SERIES GENE EXPRESSION DATA

Joana P. Gonçalves<sup>†1,2,3</sup>, Sara C. Madeira<sup>1,2,3</sup>, Arlindo L. Oliveira<sup>1,2</sup>

<sup>1</sup>*Knowledge Discovery and Bioinformatics (KDBIO) group, INESC-ID, Lisbon, Portugal*

<sup>2</sup>*Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal*

<sup>3</sup>*University of Beira Interior, Covilhã, Portugal*

<sup>†</sup> E-mail: [jpg@kdbio.inesc-id.pt](mailto:jpg@kdbio.inesc-id.pt)

Advances in high-throughput technologies have led to a considerable increase in the amount of available data resulting from biological experiments. Microarrays, in particular, enable the simultaneous measure of the expression level of a large number of genes. Although extracting biologically relevant information from gene expression data has proved to be a very challenging task, analyzing such large quantities of results is resource and time consuming, increasing the demand for efficient computational algorithms and tools. Clustering techniques have been extensively applied to both dimensions of the gene expression matrices separately. However, it is known that many activation patterns are common to a subset of genes only under a specific subset of experimental conditions. The identification of these local patterns can only be addressed using biclustering algorithms. Although the biclustering problem is NP-hard, it becomes tractable when the gene expression level is measured over time and interesting biclusters are restricted to those with contiguous time points.

We present BiGGEsTS, a free open source graphical software tool for analyzing time series expression data and revealing local co-regulations of genes in specific intervals of time using state-of-the-art biclustering algorithms. Annotations from the Gene Ontology can be integrated and used to assess which of the identified patterns are biologically relevant. The analysis of time series gene expression data is additionally supported by a number of preprocessing and post-processing methods, together with a visualization module capable of displaying informative representations of the data including heatmaps, dendrograms, expression charts and graphs of significant biological terms.