

The integration of microRNA target data by biclustering techniques opens new roads for signaling networks analysis

Gianvito Pio¹✉, Michelangelo Ceci¹, Corrado Loglisci¹, Donato Malerba¹, Domenica D'Elia²

¹Department of Computer Science, University of Bari "Aldo Moro", Bari, Italy

²CNR, Institute for Biomedical Technologies, Bari

Motivation and Objectives

MicroRNAs (miRNAs) are key modulators of gene expression. In addition to their recognised role in embryonic and adult cell proliferation and differentiation (Ren et al., 2009), many recent studies on diverse types of human cancer have demonstrated that miRNAs are functionally integrated into those oncogenic pathways that are central to tumorigenesis (Olive et al., 2010). Although microarray profiling and next generation sequencing technologies have allowed researchers to discover much of their structural and functional features as well as many new miRNAs, the current challenge is to understand their specific biological functions and mechanisms through which they are able to ensure cell homeostasis and to control developmental timing and cancer progression. This is not a trivial task because the post-transcriptional regulation of gene expression mediated by miRNAs is rarely resolved by a simple one-to-one interaction between a miRNA and a target gene. It is much more complex, often involving multiple binding of the same miRNA and/or of different miRNAs in a co-operative manner. The combinatorial effects of different miRNAs on the same gene, or on different genes of the same pathway, is an essential part of the mechanism through which they are able to fine-tune signaling pathways (Inui et al., 2010). Indeed, the effect of a miRNA may change depending on which other miRNAs are co-expressed or silenced, which in turn depends on the specific context in which the cell, the tissue or the organism is considered. This makes the interpretation of miRNAs expression profile really difficult and a mere analysis of the list of differentially expressed genes cannot provide enough information to elucidate the multiplicity of potential miRNA:mRNA interactions. In this context, the exploitation of data mining techniques, and in particular of biclustering algorithms, is considered as a useful approach to search the correlations among miRNAs and mRNAs. However, as each miRNA may target hundreds of genes, the

selection of the most significant results for further experimental validations still remains a challenging task for many biologists.

The proposed method, which is implemented in the system HOCCLUS2, has been designed to analyse data of miRNA:mRNA interactions (derived from expression arrays or from large sets of predictions) in order to detect significant co-regulatory partnerships. In particular, the aim is to provide the biologists with a tool which can support them in two challenging tasks, that is, the detection of actual miRNAs target genes and the identification of the context-specific co-associations of different miRNAs. A further contribution to the considered research consists in the ranking of the extracted biclusters on the basis of the semantic similarity between the target genes, which allows the biologists to easily select the most significant results, from a biological view point.

Availability: <http://www.di.uniba.it/~ceci/micFiles/systems/HOCCLUS2/index.html>

Methods

HOCCLUS2 exploits and integrates multiple resources. In particular: i) a novel biclustering algorithm specifically designed for the task in hand; ii) existing SVM-based classification algorithms; iii) large sets of validated or predicted miRNA:mRNA interactions; iv) gene classification ontologies (i.e. Gene Ontology) (Ashburner et al., 2000).

The analysis of miRNA:mRNA interactions consists of three steps:

1. the extraction of a set of non-hierarchically organised biclusters in form of bicliques;
2. an iterative process in which, at each iteration, two operations are performed: i) overlap identification, in which miRNAs or mRNAs belonging to a bicluster can be added to another bicluster, by exploiting an SVM-based classification algorithm; ii) merging, in which biclusters are merged when some (distance- and density-based) heuristic criteria are satisfied. Merging implicitly defines a hierarchy of clusters;

3. a ranking of the extracted biclusters. Ranking is performed on the basis of the p-values obtained by the Student's T-Test through which we compare the intra- and inter- functional similarity of miRNA targets. The similarities between miRNA targets (belonging to the same and to different biclusters, respectively) are pairwise computed according to a semantic similarity measure which takes into account the gene classification provided in GO.

Results and Discussion

In order to identify miRNA:mRNA meaningful interactions, HOCCLUS2 has been specifically designed to identify biclusters which are:

- possibly overlapping, since mRNAs and miRNAs can be involved in multiple regulatory networks. Ignoring this aspect would lead to the identification of incomplete interaction networks;
- hierarchically organised. A hierarchical arrangement facilitates the biological interpretation of results, even when a high number of biclusters is extracted from large datasets of miRNA:mRNA interactions. More importantly, this allows us to exploit the intrinsic hierarchical organisation of miRNAs, where it is possible to distinguish among miRNAs involved in many signaling pathways (universe miRNAs) and pathway-specific miRNAs (intra-pathway miRNAs) (Shirdel et al., 2011);
- highly cohesive. This means that miRNAs and mRNAs in the same bicluster should be highly related and show (only) reliable interactions.

The results reported in this paper are referred to the application of HOCCLUS2 on miRTarBase (Hsu et al., 2011) and miRDIP (Shirdel et al., 2011) selected datasets.

By comparing the results of HOCCLUS2 with those of other biclustering algorithms we have verified that HOCCLUS2 performs significantly better in terms of biclusters cohesiveness, interpretability of the results (thanks to the hierarchical organisation) and biological significance of the extracted biclusters (according to the statistical test on GO).

We have found confirmation of multiple miRNAs co-associations in experimental results reported in the current literature for many of the most significant biclusters produced by HOCCLUS2. Moreover, mRNAs in these biclusters are significantly enriched in the same or

related pathways (Reactome mapping and over-representation statistical analysis) (Haw et al., 2011). Much importantly, we have also identified potential miRNAs combinatorial associations (likely context-specific) and specific miRNA targets (potential new target genes) not yet reported in the literature and that well correlate with existing functional hypothesis. These results suggest that the proposed method is appropriate to easily identify meaningful biological correlations otherwise impossible to discover because of the huge amount of data to deal with. Indeed, the amount of data produced by experimental approaches, if from one hand provides an invaluable resource, on the other hand requires complex and exhausting procedures for their analysis. Searching for the target genes of a miRNA in miRTarBase or in miRDip, or in any other similar database, returns thousands of potential targets and to correlate these results to those of other co-expressed miRNAs is a very complex task. Such a type of analysis may greatly benefit by the application of HOCCLUS2 because of its ability to extract and rank biologically significant interaction networks. Furthermore, the possibility to dissect functional components (miRNAs and target genes) of biclusters at higher level of the hierarchy in smaller co-regulative units (biclusters at lower levels of the hierarchy), provides the key for the interpretation of multiple and diverse co-associations of specific miRNAs which could be responsible for their context-dependent activity. These data are almost impossible to obtain by other biclustering algorithms (Caldas and Kaski, 2010; Yoon and De Micheli 2005; Prelic et al., 2006; Cheng and Church, 2000; Deodhar et al., 2000) and, at our knowledge, no similar approaches have been developed and applied in the miRNAs research domain.

The HOCCLUS2 software, the user manual, all the datasets and detailed results are available from the HOCCLUS2 web site. HOCCLUS2 is currently available as a stand-alone software. The results are available in textual format and can be used for searching significant miRNA co-associations in biclusters as well as specific miRNAs gene targeting. A web-based tool for the analysis of a given set of miRNAs (or mRNAs) which renders biclusters obtained by HOCCLUS2 is under development. A further improvement envisages the integration of gene related pathways information from Reactome.

Acknowledgements

This work is partial fulfillment of the research objective of "DM19410 - Laboratorio di Bioinformatica per la Biodiversità Molecolare" and "PON01_02589 - MicroMap project "Caratterizzazione su larga scala del profilo metatrascrittomico e metagenomico di campioni animali in diverse condizioni fisiopatologiche".

References

- Ashburner M et al. (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25:25-29. doi: [10.1038/75556](https://doi.org/10.1038/75556)
- Caldas J and Kaski S (2010) Hierarchical Generative Biclustering for MicroRNA Expression Analysis. In *Research in Computational Molecular Biology*, vol. 6044 of LNCS 2010:65-79.
- Cheng Y and Church GM (2000) Biclustering of Expression Data. In *Proc. of ISMB'00* 2000:93-103.
- Deodhar M et al. (2009) A scalable framework for discovering coherent co-clusters in noisy data. In *Proc. of ICML'09*:31.
- Haw R et al. (2011) Reactome pathway analysis to enrich biological discovery in proteomics data sets. *Proteomics*, 11 (18):3598-3613. doi: [10.1002/pmic.201100066](https://doi.org/10.1002/pmic.201100066)
- Hsu SD et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research*, 39 (Database issue):D163-9. doi: [10.1093/nar/gkq1107](https://doi.org/10.1093/nar/gkq1107)
- Inui M et al. (2010) MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell Biol.*, 11:252-263. doi: [10.1038/nrm2868](https://doi.org/10.1038/nrm2868)
- Olive V et al. (2010) mir-17-92, a cluster of miRNAs in the midst of the cancer network. *Int J Biochem Cell Biol*, 42 (8):1348-1354. doi: [10.1016/j.biocel.2010.03.004](https://doi.org/10.1016/j.biocel.2010.03.004)
- Prelic A et al. (2006) A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22 (9):1122-1129. doi: [10.1093/bioinformatics/btl060](https://doi.org/10.1093/bioinformatics/btl060)
- Ren J et al. (2009) MicroRNA and gene expression patterns in the differentiation of human embryonic stem cells. *J Transl Med.*, 7 (20). doi: [10.1186/1479-5876-7-20](https://doi.org/10.1186/1479-5876-7-20)
- Shirdel EA et al. (2011) NAViGaTing the Micronome - Using Multiple MicroRNA Prediction Databases to Identify Signalling Pathway-Associated MicroRNAs. *PLoS ONE*, 6 (2):e17429. doi: [10.1371/journal.pone.0017429](https://doi.org/10.1371/journal.pone.0017429)
- Yoon S and De Micheli G (2005) Prediction of regulatory modules comprising microRNAs and target genes. *Bioinformatics*, 21 (2):93-100. doi: [10.1093/bioinformatics/bti1116](https://doi.org/10.1093/bioinformatics/bti1116)