A novel biclustering algorithm for the discovery of meaningful biological correlations between miRNAs and mRNAs

G. Pio[1], M. Ceci[1], D. D'Elia[2] ✉, C. Loglisci[1], D. Malerba

[1]Department of Computer Science - University of Bari, Bari, Italy
[2]CNR, Institute for Biomedical Technologies, Bari, Italy

Motivations

microRNAs (miRNAs) are post-transcriptional regulators which represent one of the major regulatory gene families in animals, plants and viruses and that plays a key role in almost all main cellular processes. The computational prediction of miRNA target genes is important for the functional annotation of genomes and, on the other side, functional annotation of target genes can be of great help in suggesting specific biological functions of miRNAs [1]. This work aims to contribute to the elucidation of miRNAs role in the regulation of gene expression, by proposing a method for the hierarchical and overlapping biclustering of miRNAs and target messenger RNAs (mRNAs). The method allows to discover possible miRNA:mRNA functional relationships, at different granularity levels, in large datasets produced by miRNA target site prediction algorithms, thus reducing the impact of noise on the significance of the resulting biclusters.

Methods

In order to properly work on miRNA:mRNA interactions, three important issues have to be considered. In particular, extracted biclusters should be: i) possibly overlapping, because miRNAs can be involved in different regulatory networks; ii) exhaustive, i.e. each miRNA or mRNA should belong to at least one bicluster, thus preventing the loss of possible co-regulations; iii) hierarchically organised, thus facilitating biological interpretability of results even when a high number of biclusters is extracted from large miRNA:mRNA datasets.

We propose an algorithm for the efficient discovery of overlapping, exhaustive and hierarchically organised biclusters. Our algorithm effectively deals with a kind of "relational" imbalance (i.e. miRNAs and mRNAs participate with significantly different cardinalities in the interactions). Moreover, it combines the notions of bicluster separability and bicluster distance in order to extract significant biclusters according to both density and distance-based criteria.

The performance of our method is evaluated on miRNAs target predictions in the human genome dataset extracted by miRNAMap 2.0 [2]. The strength of the miRNA:mRNA interactions is estimated on the basis of predictions provided by miRanda, RNAHybrid and TargetScan algorithms. It is computed as a linear combination of three criteria: 1) the number of algorithms which predict the miRNA target site; 2) the number of miRNA target sites found in the same UTR region; 3) the accessibility of the target site. Weights of the linear combination are selected such that criterion 1 dominates over the other two, while criterion 3, ceteris paribus on criterion 1, dominates over criterion 2.

Results

The performance of our method is evaluated in terms of execution time, bicluster compactness (the intra-bicluster cohesion) and bicluster co-regulation (the ability to group together miRNAs that target the same mRNAs). A comparative analysis shows that our method is able to extract a smaller number of (hierarchically organised) biclusters, with higher compactness and co-regulation values than ROCC [3]. Execution times of our method and ROCC are comparable. The significance of the extracted hierarchies is evaluated in terms of the F-Measure, on a set of synthetic datasets generated at different levels of noise. The analysis reveals that the hierarchy structure is almost correctly discovered, even for high levels of noise.

The effectiveness of the algorithm in extracting biologically related biclusters is tested on the basis of: i) classification of biclustered miRNAs in the same miRNA family or gene cluster; ii) validated functional associations of biclustered miRNAs reported in the literature and in major web specialised resources, iii) GO classification and functional clustering of biclustered mRNAs [4]. Results show that the proposed algorithm allows

to extract a relatively small number of biclusters that preserve both compactness and co-regulation. Biclusters extracted by our method represent meaningful biological correlations between miRNAs and mRNAs.

Availability
http://dl.dropbox.com/u/66737165/biclustering/index.html

Competing interests statement
Domenica D'Elia is on the Editorial Board of the EMBnet.journal

References

1. Grun D, Wang YL, Langenberge D, Gunsalus KC, Rajewsky N (2005) microRNA Target Predictions across Seven Drosophila Species and Comparison to Mammalian Targets. PLoS Comput Biol. 1(1)

2. Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, Hsu PW Wong YH., Chen YH, Chen GH, and Huang HD (2008) miRNAMap2.0: genomic maps of microRNAs in metazoan genomes, Nucleic Acids Res. 36

3. Deodhar M, Gupta G, Ghosh J, Cho H, Dhillon IS (2009) A scalable framework for discovering coherent coclusters in noisy data, ICML

4. Huang DW, Sherman BT and Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources, Nature Protoc. 1(4), 44-57