MiRnaBoost: Multi-view AdaBoost for microRNA target prediction

Jocelyn Brayet, Remy Nicolle, Mohamed Elati
institute of Systems and Synthetic Biology
Genopole Campus 1 - Genavenir 6
5 rue Henri Desbruères - F-91030 EVRY cedex
mohamed.elati@issb.genopole.fr

MicroRNAs are short (21-25 nt) non-coding RNAs that repress the expression of their direct targets (Bartel, 2009). Building an accurate binding model for a microRNA is essential to differentiate its true binding targets from spurious ones (Khorshid 2013). So far, conventional approaches to prediction of microRNA binding sites have all relied on local sequence information only, in a way or another. In this work we devise a novel machine learning system, MiRnaBoost, to build a microRNA binding classifier by combining sequence, expression and position information-based classifiers. Currently, sequence-based prediction methods are not fully capturing microRNA target preferences, nor context specific regulations. To overcome the limitation of sequence-only miRNA-gene interaction prediction, MiRnaBoost complements a sequence based classifier (miRanda) with two additional supervised models trained on different views i) the expression levels of both the miRNA and the target gene (Huang 2007), ii) the pattern of the genomic position (Elati 2013) of the targets of a miRNA. MiRnaBoost combines these weak classifiers using a modified version of the Adaboost algorithm, which manages to combine and improve together classifiers trained on the same instances but on different views (Zhijie 2010; Elati 2013).
Based on cross-validation analysis over the microRNAs with the most validated targets in TarBase, MiRnaBoost consistently outperforms conventional methods exploiting only sequence information. The main advantage of MiRnaBoost is that it lowers the false positive rate. Furthermore, MiRnaBoost predicted miRNA target sets are more consistently annotated with GO terms than similar sized random subsets of genes with conserved miRNA seed regions.

References

Bartel D: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136(2):215–233.
Khorshid M, Hausser J, Zavolan M, van Nimwegen E: A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. Nature Methods 2013
Elati M, Nicolle R, Junier I, Fernandez D, Fekih R, Font J, Kepes F: PreCisIon: PREdiction of CIS-regulatory elements improved by gene's positION. Nucleic Acids Res. 2013 February; 41(3): 1406,1415.
Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, Hughes TR, Blencowe BJ, Frey BJ, Morris QD: Using expression profiling data to identify human microRNA target. Nature Methods 2007, 4:1045–1049.
Zhijie X, Shiliang S: An Algorithm on Multi-View Adaboost. ICONIP 2010, Sydney, Australia, November 22-25, 2010, Proceedings, Part I, LNCS 6443, pp. 355-362