

A Support Vector Machine for the Discrimination of MicroRNA Precursors from Other Genomic Hairpin Structures

ID - 217

Piccolo Viviana¹, Pesole Graziano^{2,3}, Horner David¹

¹Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Via Celoria 26, 20133 Milano

²Istituto Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, via Amendola 122/D, 70125 Bari, Italy

³Department of Biochemistry and Molecular Biology, University of Bari, Via Orabona 4, 70126 Bari, Italy

Motivation

MicroRNAs (miRNAs) are endogenous, small (~ 20 nt), single-stranded, non-coding RNAs (ncRNAs) that result from the nuclear and cytoplasmic processing of transcribed precursor hairpin structures. They are increasingly recognized as playing crucial roles as post-transcriptional antisense regulators of gene expression through regulation of mRNA stability or translational efficiency. miRNAs, first reported in *Caenorhabditis elegans*, have been identified in the genomes of most higher organisms, including worms, flies, plants, mammals and recently in viruses. Functional studies have shown that miRNAs play important roles in processes such as, cell proliferation, fat metabolism, apoptosis, neuronal cell fate, insulin secretion, haematopoietic differentiation and developmental regulation.

The detection of homologs of known miRNAs through comparative genomic approaches has proved relatively tractable. However, the ab-initio prediction of miRNA precursors through computational methods poses several additional difficulties, not least the fact that not all thermodynamically plausible transcribed hairpins are processed to yield mature miRNAs. It has not until now been possible to identify conserved sequence or structural elements that define consensus recognition elements for the enzymes that process miRNA precursors. In the light of these observations we wished to develop and improve methods for the discrimination of true miRNA precursor hairpins from spurious hairpins that do not undergo processing in-vivo.

Methods

We have developed a SVM (Support Vector Machine) that considers up to 74 features associated with the primary and secondary structures and thermodynamic characteristics of candidate hairpin structures. We use a standard heuristic approach to optimize combinations of features used and train the SVM with sets of characterized hairpin miRNA precursors and known non-miRNA hairpins.

Results

Our SVM shows highly promising results in the discrimination of true miRNA precursors from spurious hairpins (typically around 95% sensitivity) in various species. In particular, our levels of false positive predictions appear to be low relative to comparable methods.

Email: viviana.piccolo@unimi.it