RANKING PREDICTED MIRNA TARGETS THROUGH THEIR BINDING PATTERNS

Leyan Tang
Computer Science, San Jose State University, One Washington Square
San Jose, CA 95112
Leyan.tang@students.sjsu.edu

miRNAs are a class of small RNA molecules of ~22nt that regulate gene expression at the post-transcriptional level. Since experimental identification of miRNA target is dreadfully difficult, computational target prediction programs remain the only way of fast identification of a potential miRNA target. Current prediction tools have produced a huge number of potential target sites, but determining which algorithm produces the most reliable prediction results remains an open problem.

This study is based on the hypothesis that binding patterns determine the mRNA:miRNA stability. There are four types of base-pairings between the miRNA sequence and its target on the mRNA: Watson-Crick pair, G-U pair, none pair and bulge. Thus, each miRNA interaction could be converted into a unique binding pattern sequence, similar to protein secondary structure. We analyze the common characteristic of binding patterns through position weight matrices (PWM) and 1st-order Markov Model. PWMs were built for known human miRNA genes grouped by length. All test data are from experimental verified human miRNA interactions. It is a well-known fact that positions 2 to 7 from the miRNA bind perfectly with the target mRNA and form the "seed". Beyond the seed, the paired type constitutes around 50% when compared to the other binding types. While the PWM yields the types of the base-pairings, it does not give the variations of the bindings as we move from one nucleotide to the next. So we modeled, studied and analyzed this variation by using a 1st-order Markov Model.