

A Kernel-Based Method for MicroRNA Precursor Identification

Authors : Bin Liu

Abstract : MicroRNAs (miRNAs) are small non-coding RNA molecules, functioning in transcriptional and post-transcriptional regulation of gene expression. The discrimination of the real pre-miRNAs from the false ones (such as hairpin sequences with similar stem-loops) is necessary for the understanding of miRNAs' role in the control of cell life and death. Since both their small size and sequence specificity, it cannot be based on sequence information alone but requires structure information about the miRNA precursor to get satisfactory performance. Kmers are convenient and widely used features for modeling the properties of miRNAs and other biological sequences. However, Kmers suffer from the inherent limitation that if the parameter K is increased to incorporate long range effects, some certain Kmer will appear rarely or even not appear, as a consequence, most Kmers absent and a few present once. Thus, the statistical learning approaches using Kmers as features become susceptible to noisy data once K becomes large. In this study, we proposed a Gapped k-mer approach to overcome the disadvantages of Kmers, and applied this method to the field of miRNA prediction. Combined with the structure status composition, a classifier called imiRNA-GSSC was proposed. We show that compared to the original imiRNA-kmer and alternative approaches. Trained on human miRNA precursors, this predictor can achieve an accuracy of 82.34 for predicting 4022 pre-miRNA precursors from eleven species.

Keywords : gapped k-mer, imiRNA-GSSC, microRNA precursor, support vector machine

Conference Title : ICBCSB 2018 : International Conference on Bioinformatics and Computational Systems Biology

Conference Location : Sydney, Australia

Conference Dates : January 29-30, 2018