mirhunt: an approach to predict microRNA binding sites using different prediction tools

Jaqueline Ramalho[1,2], Michelle A. Paz[1,2], Iane O. P. Porto[1,2], Celso T. Mendes-Junior[3,] Erick C. Castelli[1,2]

(1) Programa de Pós-Graduação em Patologia – Faculdade de Medicina de Botucatu, UNESP, (2) Laboratório de Genética Molecular e Bioinformática, Unidade de Pesquisa Experimental, Bloco 5, Faculdade de Medicina de Botucatu – UNESP and (3) Departamento de Química – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, USP

One of the mechanisms for gene expression regulation occurs post transcriptionally by the binding of microRNAs to the mRNA. microRNAs are small noncoding RNA molecules of about 22 nucleotides that usually interacts with the mature mRNA 3' untranslated tail. Their binding occurs through the seed region (usually 2-7 nucleotides) in the microRNA 5' end. This binding might suppress translation or trigger mRNA degradation. The same microRNA may bind to several targets, and one gene might be regulated by several microRNAs, configuring regulatory networks. Several studies have reported the up- and down-regulation of specific microRNAs in pathological contexts. In this matter, it is important to infer which microRNA may influence the expression of clinically relevant genes for future functional studies. The prediction of microRNA targets or binding sites usually requires the use of different prediction tools, mainly because they might predict different targets for a same microRNA/mRNA pair as well as to avoid false-positive interactions. In addition, the use of long mRNA sequences to predict microRNA/mRNA interactions might bias the analyses detecting only the most stable bindings. *mirhunt* is an application designed to automate the searching for human microRNA binding sites by using three different prediction algorithms, miRanda, RNAhybrid and IntaRNA, comparing the results of these methods and applying a scoring system to classify each interaction. The scoring system is based on the strength of each microRNA/mRNA interaction found by comparing it with a database (matrix of 26,414 human mRNA sequences inferred from the annotations provided by the human genome draft version hg38 and 2,588 known microRNAs (miRBase)), which containing the interactions previously detected. Long sequences are managed by fragmenting them on a series of overlapped subsequences and processed independently, maximizing the ability of these algorithms to predict miRNA/mRNA interactions on long mRNA sequences. Reports are generated presenting all the interactions found and the scores for each interaction. The scoring system used minimizes the presence of false-positive interactions, since it considers the results from three prediction algorithms, and also allows the selection of the most stable bindings and most specific bindings by comparing the target results with the *mirhunt* database. For example, *mirhunt* was used to infer the microRNA binding sites to the *HLA-G* 3'UTR sequence and the most specific bindings detected were related to miR-148a-3p and miR-148b-3p, whose influence on the HLA-G expression profile was functionally proved. The data provided by *mirhunt* may provide background for functional studies. Available: http://www.castelli-lab.net/apps/apps_mirhunt.php.

Financial Support: FAPESP, CNPq
E-mail: castelli@fmb.unesp.br