A preliminary study on distribution and conservation of MicroRNAs

ID - 226

Maselli Vincenza[1,2], Diego Di Bernardo[1,2], Sandro Banfi[1,2]
[1]TIGeM Telethon Institute of Genetics and Medicine, Napoli
[2]SEMM European School of Molecular Medicine, Napoli

Motivation

MicroRNAs (miRNAs) are a large family of endogenous non-coding RNAs of approximately 22nt small RNAs, which are abundant, highly conserved, and predicted to regulate a large number of transcripts. Most miRNAs are located inside a gene either coding or non-coding, often into an intron, but they can also be localized in intergenic regions.Our aim is to perform a semi automatic analysis of distribution, conservation and evolution of miRNAs. We made an analysis to look at the distribution and the conservation of genomic localization and structure of microRNAs, which are annotated in miRBase, among the species. We plan to develop a web-based database to collect and to show information about distribution and localization on the genome of both known and predicted miRNAs. In addition we plan to verify whether there are microRNAs that do not show a conservation of their genomic location in order to infer more insights into the biology of the miRNA, and in particular transcription regulation and identification of regulatory regions.

Methods

First we identify the host gene looking at the pre-miRNA sequence overlapping with a gene and for the coding host gene we look at the conservation in the other species.
Until now our analysis is based on ensEMBL hortology prediction but we are performing a pipeline to improve this prediction. To study the conservation of the location we use as reference the exons that flank the intron on which the miRNA falls or the exon itself if the miRNA falls on an exon. To identify putative specie-specific miRNAs we run BLAST for each pre-miRNA sequence of a species A which is absent in the species B against the genome of B, then we use further blast analysis to identify new predicted miRNAs and putative specific miRNAs.

Results

Our preliminary results show that 50% of miRNA has a known host gene and for these there is a conservation of location (both the surrounding exons are conserved). We also found that almost 26% of human and 10% of mouse miRNA are absent in other species. Since the data in miRBase, from which we start, are too stringent because they are based only on experimental evidence, these results will be validated by further analysis such as running BLAST and using conservation score and secondary structure prediction and in some cases experimental test.
On the other hand, we believe that a manual and experimental validation of our results should increase the number of predicted miRNAs and it should help to have a deeper insight into the miRNAs world.
All data will be stored into a database under construction and will be freely available at the address http://maselli.tigem.it/Projects/microrna/htdocs/

Availability: http://maselli.tigem.it/Projects/microrna/htdocs/

Email: maselli@tigem.it