Improving microRNA Prostate Cancer Target Genes detection

Masulli F[1,2], Parini A[1], Rovetta S[1], Russo G[2,1]

[1] Department of Computer and Information Sciences, University of Genova,
Via Dodecaneso 35, 16146 Genova, Italy
[2] The Sbarro Institute for Cancer Research and Molecular Medicine,
Center for Biotechnology, Temple University, BioLife Science Bldg.,
1900 N 12th Street Philadelphia, PA 19122 USA

Motivation

MiRNA (microRNAs) are a class of short, non coding RNA capable of base-pairing with imperfect complementarity to the transcripts of animal protein-coding genes (targets), generally within the 3' untranslated region[1]. They are able to negatively regulate gene expression; however, the exact mechanism remains still unclear. Since each miRNA targets several hundreds of transcripts and each transcript can be the target of multiple miRNAs, the complexity of this class of RNA is remarkable. Computational prediction of miRNA targets is not an easy task. More than 10 target prediction programs have been developed. They typically analyze sequences by a sliding window and, in order to reduce false positive predictions, they test each site of the gene with some algorithms in cascade. Usually only the 3'-UTR of the gene is considered and sometimes we find more than one target site prediction for the same gene. The Mirecords web site, http://mirecords.umn.edu/miRecords/, shows the list of biologically validated gene targets of all known miRNAs in humans and other animals, together with the target prediction obtained by various programs. Much work has still to be done in order to obtain reliable target prediction methods. We present an approach to improve computational estimation of target sites, maximizing the adherence of results to biological evidence. The increased selectivity of the method may also be used to guide experimental validation in a more focused direction. Prostate cancer is very widespread, with variable prognosis and evolution, which suggest diversity at the molecular level. A deep understanding of the mechanics underlying the disease is necessary, since therapies are often very aggressive and potentially dangerous. Being able to discriminate between cancer subtypes is therefore a first step in more targeted therapies. MiRNAs have been proved to be involved in the development of prostate cancer, so a very important goal is to ensure reliability of MiRNA target prediction software, the topic of the present work.

Methods

MiRanda [2], proposed in 2003, is the most popular miRNA target prediction method. It splits the target site prediction task into three distinct steps carried out in sequence: Homology evaluation, based on the Smith-Waterman algorithm, used for complementarity rather than matching, with ad-hoc penalty values; free energy computation, carried out using the RNA folding routine RNAfold included in the Vienna RNA secondary structure library; evolutionary conservation computation, as a third filter applied to binding sites that passed the previous two filtering stages, and performed by PhastCons, a third party tool. The parameters used in MiRanda have been optimized according to biological knowledge and software tools available in 2003. The method we present provides some improvements on the original MiRanda program. We employ the updated knowledge of biologically validated MiRna gene targets nowadays available at Mirecords web site for tuning the parameters, and adopt the new RNAcofold routine introduced in Vienna RNAlib in 2006. Another goal is to improve the match between the set of estimated target genes and the set of biologically validated target genes by optimizing the algorithm parameters. To this aim we used a Genetic Algorithm [3] (GA) as implemented in the package GALOPPS. Genetic algorithms are global search heuristic techniques inspired by mechanisms of evolution such as selection, crossover and mutation.

Results

We have selected a dataset of 50 candidate target genes from those involved in pathways related to prostate cancer. The MiRNAs considered are miR-15a and miR-1, both also proven to be involved in prostate cancer. Eight genes in the dataset are also biologically validated on the Mirecords web site, i.e., they actually feature one or more target sites for these MiRNAs. The available daset of candidate target genes was split into two data sets labeled as A and B. Training was performed wit miR-15a on A, whereas testing was done with miR-1 on A and both miR-1

and miR-15a on B. The GA finds an optimal solution after 18 generations (iterations). On the attached table (see Supplementary Informations) we compare the results obtained with MiRanda and with our proposed method. Our method is more selective than MiRanda. The validated targets with the proposed method have consistently better free energy than the others in the list; this is not true for MiRanda.

Supplementary informations

1) Experimental results
PDF format: http://mlsc.disi.unige.it/projects/miRNA/BITS-abstract/table.pdf
As a PNG image: http://mlsc.disi.unige.it/projects/miRNA/BITS-abstract/table.png
2) Web links
Mirecords: http://mirecords.umn.edu/miRecords/
MiRanda: http://www.microrna.org/
RNAlib: http://www.tbi.univie.ac.at /~ivo/RNA/RNAlib/
PhastCons: http://compgen.bscb.cornell.edu/phast/
3) Bibliography for the abstract
[1] Y.Wang, H.M. Stricker, D.Gou, and L.Liu, "MicroRNA: past and present", Frontiers in Bioscience, vol.12, pp. 2316-2329, January 2007.
[2] John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. "Human MicroRNA targets." PLoS Biol. 2005 Jul;3(7):e264.
[3] D.E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning." Boston, MA, USA:Addison-Wesley Longman Publishing Co., Inc., 1989.

Contact : ste@disi.unige.it