MiRonTop: mining microRNAs targets across large scale gene expression studies

Kevin Le Brigand[1,2,†], Karine Robbe-Sermesant[1,2,†], Bernard Mari[1,2] and Pascal Barbry[1,2,*]

[1]Centre National de la Recherche Scientifique, Institut de Pharmacologie Moléculaire et Cellulaire, UMR6097, 06560 Sophia-Antipolis and [2]University of Nice Sophia Antipolis, Nice, France

Associate Editor: David Rocke

ABSTRACT

Summary: Current challenges in microRNA (miRNA) research are to improve the identification of *in vivo* mRNA targets and clarify the complex interplay existing between a specific miRNA and multiple biological networks. MiRonTop is an online java web tool that integrates DNA microarrays or high-throughput sequencing data to identify the potential implication of miRNAs on a specific biological system. It allows a rapid characterization of the most pertinent mRNA targets according to several existing miRNA target prediction approaches. It also provides useful representations of the enrichment scores according to the position of the target site along the 3′-UTR, where the contribution of the sites located in the vicinity of the stop codon and of the polyA tail can be clearly highlighted. It provides different graphs of miRNA enrichment associated with up- or down-regulated transcripts and different summary tables about selections of mRNA targets and their functional annotations by Gene Ontology.

Availability: http://www.microarray.fr:8080/miRonTop/index

Contact: barbry@ipmc.cnrs.fr

1 INTRODUCTION

MiRNAs appear capable to block translation and also stimulate mRNA degradation, with translational repression often preceding mRNA degradation. Several mechanisms of miRNA action, including block of cap binding or elongation, increased ribosome drop-off rate or increased de-adenylation, have already been proposed (Guo *et al.*, 2010). It remains, however, that target recognition follows a complex set of rules dominated by the existence of a perfect match between 6 and 8 nt located in the 5′ region of the miRNA (called the 'seed') and the target mRNA (Grimson *et al.*, 2007). The ability of miRNAs to interact with many targets and the possibility for some of them to share a same target increase tremendously the complexity of many biological networks. In that context, an experimental validation of *in silico* predictions appears quite necessary. Many available target prediction softwares combine a search for a perfect seed match between miRNA and 3′-untranslated region (3′-UTR) and a conservation of the interaction across species. In fact, comparison of results

obtained with several of these computational tools show significant divergences. MiRonTop aims at detecting influence of miRNAs on gene expression profiles by analyzing potential overrepresentation of specific miRNAs targets across an experimental gene set. Other tools have been recently described to explore miRNA influence on datasets (Alexiou *et al.*, 2010; Antonov *et al.*, 2009; van Dongen *et al.*, 2008). MiRonTop provides several novel and complementary information including (i) the possibility to screen and compare miRNA signatures using different target prediction tools; (ii) options to calculate enrichment scores according to the spatial distribution of predicted targets sites along the transcript, since true sites may be preferentially located in close vicinity of stop codons and polyadenylation sites (Chi *et al.*, 2009; Majoros and Ohler, 2007); and (iii) access to Gene Ontology functional annotations of the different gene lists. Overall, these different modes of representation increase the sensitivity of the research in order to define a set of either mRNA or miRNAs associated with the selected experiment.

2 IMPLEMENTATION

2.1 Description of the interface

MiRonTop takes as input the results of a global mRNA expression data. An acceptable format corresponds to a tab-delimited text file containing a unique identifier (HGNC gene symbol, NCBI gene ID, Affymetrix or Agilent probe identifiers), a signal intensity, a $\log_2$ ratio between two experimental conditions and a statistical value (*P*-value or odds ratio). MiRonTop maintains a relational database linking DNA microarray probes identifiers with NCBI gene IDs, which are used as references by the application. The database also keeps targeting relationships between NCBI gene IDs and miRNA predicted by Targetscan, MicroCosm, Miranda, Pictar and an exact seed search (2–7, 1–8 or 2–8 first nucleotides of the miRNA) across genes 3′-UTRs and coding sequence (CDS). Predictions can be tuned with several adjustable parameters defining expressed genes (signal intensity $\geq$ cut-off$_1$) and statistically significant genes (expressed genes having $|\log_2$ ratio$| \geq$ cut-off$_2$ and *P*-value $\leq$ cut-off$_3$ or odds ratio $\geq$ cut-off$_{3'}$). The three cut-offs, selected by the end-user, are used to define two sets: the down-regulated gene set (where enrichment for mRNA targets of up-regulated miRNAs should be detected) and the up-regulated gene set (where mRNA targets for down-regulated miRNAs are expected). For each miRNA considered, an enrichment of its predicted targets according to one selected prediction tool is defined for the 'Down' and the 'Up' gene sets, significance

---

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
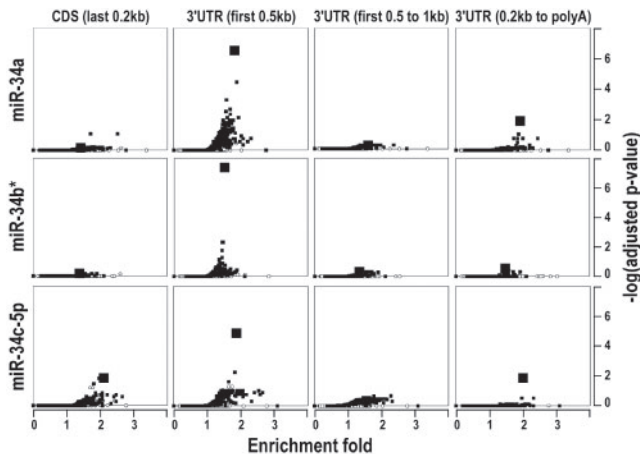
Fig. 1. Graphs derived from MiRonTop showing the significance of the enrichment [represented as −log10(adjusted *P*-value)] according to the fold enrichment, in three experiments of overexpression of miR-34a, miR-34b* and miR-34c-5p in four distinct regions of the transcripts. On each panel, a large square highlights the result for the overexpressed miRNA.

being evaluated using the hypergeometric law. The interface then displays for each miRNA the relationship between a Benjamini and Hochberg-adjusted hypergeometric *P*-value and the fold enrichment, relative to the Down and Up gene sets. In this representation, miRNAs having the largest influence on 'Down' or 'Up' gene set (green or red circles, respectively) are located in the top right quadrant of the plot. A possibility is given to highlight the effect of a specific miRNA (large square). The interface also provides a summary table. Functional annotations of the different gene lists can be accessed through several hyperlinks that show the results of Gene Ontology enrichment analyses. Additional information provided for each miRNA corresponds to the expression levels of the transcripts encoded by the same host gene. Each table can be easily downloaded for further analysis.

2.2 Database resources

Genes set references are based on Refseq v14/06/2010. Gene Ontology annotations are linked to Refseq genes through NCBI Gene IDs. The set of miRNAs corresponds to release 15 of Mirbase (Griffiths-Jones, 2006). The interface accommodates several miRNA prediction softwares: Targetscan v5.1 (Lewis *et al.*, 2003), MicroCosm Targets v5 (Griffiths-Jones *et al.*, 2008), Miranda-sept.2008 (Enright *et al.*, 2003) and PicTar (Krek *et al.*, 2005) and exact search of miRNA seeds 2–7, 1–8 and 2–8 into the 3′-UTR and into the CDS of the reference set of genes.

2.3 Example of dataset exploration

Figure 1 summarizes the results of a typical experiment where miR-34a, miR-34b* and miR-34c-5p were individually overexpressed into HeLa cells. Expression data have been publicly deposited in Gene Expression Omnibus (reference GSE15957). Corresponding top tables are also accessible through MiRonTop web portal. Cut-offs were equal to 8 for the $\log_2$ signal, 0.5 for the $\log_2$ ratio (miR-34a,b*,c-5p versus controls) and 0 for the odds ratio. Large variations in the −log(adjusted *P*-value), which depicts

the significance of the enrichment, were noticed for miR-34abc (large black square) when different zones of the transcripts were analyzed. From this representation and in agreement with Majoros and Ohler, the most specific profiles were observed for the regions of the 3′-UTR adjacent to the stop codon, and, to a lesser extent, to the polyA tail. Low enrichment was noticed for the CDS and in the middle of the 3′-UTR. This was consistent with Hafner *et al.* showing that CDS sites were not as functionally active as 3′-UTR sites (Hafner *et al.*, 2010). Similar relationships were made for about 10 distinct miRNAs, regardless of the target prediction software that was used (data not shown).

3 CONCLUSION

MiRonTop is an online application allowing the detection of miRNAs that significantly affect gene expression at a large scale. The flexibility of the interface allows a fine-tuning of the 'Down' and 'Up' gene sets according to three parameters that correspond to the level of expression, the extent of differential expression and a significance index. Such a selection allows experimenters to easily explore a large-scale experiment and get a summary report of the different predictions made by several miRNA target identification softwares. Recent investigations on other datasets have indicated that miRonTop is not only useful in experiments of miRNA overexpression/knock-down but can also detect more subtle miRNA activities in a physiological context. In the future, MiRonTop will be easily updated in order to integrate any additional predictions made publicly available, such as experimentally validated pairs of miRNA and mRNA that could now be provided by recent experimental approaches (Chi *et al.*, 2009; Hafner *et al.*, 2010).

Conflict of Interest: none declared.

REFERENCES

Alexiou,P. *et al.* (2010) The DIANA-mirExTra web server: from gene expression data to microRNA function. *PLoS One*, **5**, e9171.

Antonov,A.V. *et al.* (2009) GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res.*, **37**, W323–W328.

Chi,S.W. *et al.* (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.

Enright,A.J. *et al.* (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.

Griffiths-Jones, S. (2006) miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**, 129–138.

Griffiths-Jones,S. *et al.* (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.

Grimson,A. *et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.

Guo,H. *et al.* (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.

Hafner,M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.

Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.

Lewis,B.P. *et al.* (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.

Majoros,W.H. and Ohler,U. (2007) Spatial preferences of microRNA targets in 3′ untranslated regions. *BMC Genomics*, **8**, 152.

van Dongen,S. *et al.* (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nat. Methods*, **5**, 1023–1025.