BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 15 2010, pages 1920–1921
doi:10.1093/bioinformatics/btq298

Systems biology

Advance Access publication June 6, 2010

MiRror: a combinatorial analysis web tool for ensembles of microRNAs and their targets

Yitzhak Friedman[1,†], Guy Naamati[2,†] and Michal Linial[1,*]

[1]Department of Biological Chemistry, Sudarsky Center for Computational Biology and [2]School of Computer Science and Engineering, Hebrew University of Jerusalem, 91904, Israel

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: The miRror application provides insights on microRNA (miRNA) regulation. It is based on the notion of a combinatorial regulation by an ensemble of miRNAs or genes. miRror integrates predictions from a dozen of miRNA resources that are based on complementary algorithms into a unified statistical framework. For miRNAs set as input, the online tool provides a ranked list of targets, based on set of resources selected by the user, according to their significance of being coordinately regulated. Symmetrically, a set of genes can be used as input to suggest a set of miRNAs. The user can restrict the analysis for the preferred tissue or cell line. miRror is suitable for analyzing results from miRNAs profiling, proteomics and gene expression arrays.

Availability: http://www.proto.cs.huji.ac.il/mirror

Contact: michall@cc.huji.ac.il

Received on February 11, 2010; revised on June 2, 2010; accepted on June 3, 2010

1 INTRODUCTION

Recent miRNA discovery techniques confirmed the presence of hundreds of miRNAs in healthy and diseased tissues. Currently, approximately 900 miRNAs are reported in human and approximately 600 in mouse. It was shown that miRNA binding sites are not confined solely to 3′-untranslated regions (UTRs), but also are found in coding regions and introns (Kim, 2005). Furthermore, multiple features were shown to be imperative in regulation by miRNAs (Sethupathy *et al*., 2006; Nielsen *et al*., 2007). This led to a growing number of resources and algorithms that predict the pairing of miRNAs and their targets (Mendes *et al*., 2009). Currently, mirBase (Griffiths-Jones *et al*., 2008) is the most exhaustive collection of miRNAs with over approximately 15 000 mature miRNA sequences. miRBase collection (Enright *et al*., 2003; Griffiths-Jones *et al*., 2008; John *et al*., 2004) is based on MicroCosm algorithm. Additional exhaustive miRNA targets predictors include TargetScan (Lewis *et al*., 2003) and PicTar (Krek *et al*., 2005). Other large-scale resources for microRNA–target pairing includes PITA (Kertesz *et al*., 2007), microRNA.org that is based on miRanda algorithm (Betel *et al*., 2008; Enright *et al*., 2003; John *et al*., 2004), DIANA-microT 3.0 (Maragkakis *et al*., 2009), miRDB (Wang, 2008) and ElMMO-MirZ (Hausser *et al*., 2009). For a subset of miRNAs–target pairs, *in vivo* experimental validation is available in TarBase database (Papadopoulos *et al*., 2009).

We have developed miRror, an interactive tool for analyzing experimental results under the notion of coordination in miRNAs regulation. A coordinated regulation by several miRNAs was proposed [see discussion (Ivanovska and Cleary, 2008; Krek *et al*., 2005)]. For a set of miRNAs, miRror outputs a ranked list of gene targets according to their likelihood to be targeted by the miRNA ensemble and vice versa for a set of regulated genes. miRror converts all predictor resources results into a unified platform by incorporating a statistical measure according to the hypergeometrical distribution. For example, in miRNAs overexpressing experiments in human cells (Lim *et al*., 2005), 174 genes (miR-124) and 65 genes (miR-1) were downregulated and were used as input for miRror. The original miRNAs were identified as the top candidates with a P-value <3e-11 and with 91–93% of the genes marked as targets for the relevant miRNA. Similar results were obtained by using only 50% of the gene lists, supporting the robustness of the miRror platform.

2 SYSTEM OVERVIEW

miRror is a platform by which an experimentalist can gain biological insights on sets of molecular entities. Such sets are the product of large-scale experiments such as miRNA profiling, mass spectrometry proteomics and gene expression data. miRror operates at a dual mode where ensemble of (i) miRNAs and (ii) gene targets/proteins are used as the input.

2.1 Prediction tools resources integration

Database files from multiple miRNA–target prediction resources were separated to human, mouse and fly. The resources used include TargetScan, MicroCosm implemented in miRBase, PicTar, DIANA-MicroT, PITA, ElMMO-MirZ, miRanda-based microRNA.org, TargetRank, miRDB and TarBase. miRror combines the resources following a conversion of the miRNAs and gene targets identifiers. RefSeq and UniProtKB identifiers are indicated as the primary entries. TarBase used as a validation resource rather then a prediction database.

2.2 miRtegrate analysis

miRtegrate protocol is the core of miRror (Fig. 1). It calculates the probability of matches between miRNAs in the user list and the best set of targets derived from miRNA–target predictions as reported by each of the selected resources. For each gene, we

*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.

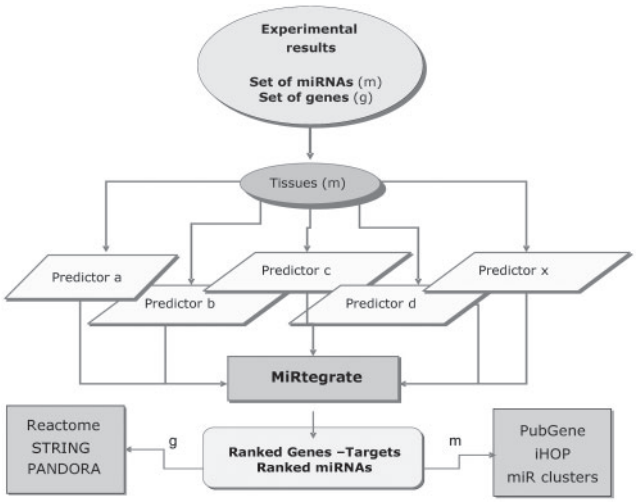

Fig. 1. A schematic view of the miRror platform. miRror supports experimental results of miRNAs (m) or genes set (g) as input. The predictor resources (a to x) are unified by miRtegrate into a statistical framework. Biological knowledge is further extracted by applying external tools.

take into account the list of miRNAs that were associated with it. We calculated the probability of the gene's interaction with the collection of miRNAs in the user set as opposed to the rest of the miRNAs in that database. The miRNA–targets coverage associated with each of the prediction resources is used for the calculation of a statistical significant threshold. We applied a default value of 0.05 as a threshold and calculated the *P*-value by considering all probabilities that are more significant that the threshold. *P*-value for miRNAs as input was performed according to the hypergeometric distribution with a correction for multiple testing. The same principle is applied with genes as input. For more details see the 'Help' section of the website.

2.3 Advanced options and filtration

The miRror results are a ranked list of best predicted molecules (miRNAs or genes, Fig. 1). A refinement of the results is based on several advanced options. Several background distributions such as different tissues and cell lines can be applied for the miRtegrate calculations. Tissues are annotated by the gene expression data from FlyAtlas (Chintapalli *et al*., 2007) and GNF atlas (Su *et al*., 2004). Using these optional backgrounds, new calculation for the *P*-value is activated to account for the apparent difference in coverage. The user can also select any subset of prediction resources to be used. A default parameter requires that a target will be selected by at least two databases. However, an intuitive option for changing the number of the miRror results is provided to the user by modifying all parameters (*P*-value threshold, minimal number of miRNAs that bind to each of the targets and minimal number of miRNA–target databases). The web tool provides multiple sorting options, optional filtration thresholds and downloading options. Furthermore, results that are experimentally validated in TarBase are highlighted. The user can forward the results to a set of analysis tools to represent enriched pathways, networks and functional annotations (Fig. 1).

3 CONCLUSIONS

We present miRror as a novel integration scheme for a dozen of miRNA–target prediction resources. miRror provides a double-sided view for ensembles of miRNAs or genes. Based on the notion of combinatorial regulation adopted by miRror, the reported list of targets for a miRNAs ensemble is a small subset of the inclusive list that is produced by the summation of predictors. As some miRNAs are predicted to pair with hundreds of targets, a reduction to a small but significant set of targets is valuable for the experimentalist. miRror application towards a set of genes as input is powerful in proposing potential miRNAs ensemble. Thus, the online tool can be used to generate hypotheses on the role of a specific genes or specific minimal set of miRNAs in any cellular settings.

ACKNOWLEDGEMENTS

We thank the ProtoNet team and Solange Karsenty for design and operation of the miRror website.

Funding: EU Framework VII Prospects consortium; grant from ISF 592/07; grant from BSF 2007219; fellowship from the SCCB, the Sudarsky Center for Computational Biology (to G.N.).

Conflict of Interest: none declared.

REFERENCES

Betel,D. *et al*. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res*., **36**, D149–D153.
Chintapalli,V.R. *et al*. (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat. Genet.*, **39**, 715–720.
Enright,A.J. *et al*. (2003) MicroRNA targets in Drosophila. *Genome Biol*., **5**, R1.
Griffiths-Jones,S. *et al*. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res*., **36**, D154–D158.
Hausser,J. *et al*. (2009) MirZ: an integrated microRNA expression atlas and target prediction resource. *Nucleic Acids Res*., **37**, W266–W272.
Ivanovska,I. and Cleary,M.A. (2008) Combinatorial microRNAs working together to make a difference. *Cell Cycle*, **7**, 3137–3142.
John,B. *et al*. (2004) Human microRNA targets. *PLoS Biol.*, **2**, e363.
Kertesz,M. *et al*. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
Kim,V.N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.*, **6**, 376–385.
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.
Lim,L.P. *et al*. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
Maragkakis,M. *et al*. (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.
Mendes,N.D. *et al*. (2009) Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Res.*, **37**, 2419–2433.
Nielsen,C.B. *et al*. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
Papadopoulos,G.L. *et al*. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res*., **37**, D155–D158.
Sethupathy,P. *et al*. (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods*, **3**, 881–886.
Su,A.I. *et al*. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
Wang,X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.