Interactive exploration of RNA22 microRNA target predictions

Phillipe Loher and Isidore Rigoutsos*

Computational Medicine Center, Thomas Jefferson University, Philadelphia, PA 19107, USA

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: MicroRNA (miRNA) target prediction is an important problem. Given an miRNA sequence the task is to determine the identity of the messenger RNAs targeted by it, the locations within them where the interactions happen and the specifics of the formed heteroduplexes. Here, we describe a web-based application, RNA22-GUI, which we have designed and implemented for the interactive exploration and *in-context* visualization of predictions by RNA22, one of the popular miRNA target prediction algorithms. Central to our design has been the requirement to provide informative and comprehensive visualization that is integrated with interactive search capabilities and permits one to selectively isolate and focus on relevant information that is distilled on-the-fly from a large repository of pre-compiled predictions. RNA22-GUI is currently available for *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster* and *Caenorhabditis elegans*.
Availability: http://cm.jefferson.edu/rna22v1.0/.
Contact: Isidore.Rigoutsos@jefferson.edu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

MicroRNA (miRNAs) are non-coding RNAs, ~22 nucleotides in length, that regulate messenger RNA (mRNA) expression in a sequence-dependent manner (Bartel, 2009; Pasquinelli, 2012). Efforts during the past decade have shown the importance of miRNA-based regulation in diverse contexts and in health and disease (Esteller, 2011). Thus, the ability to predict miRNA targets has attracted a lot of attention and several algorithms have been proposed (for a review, see Rigoutsos and Tsirigos, 2010). Despite progress, the problem of miRNA target prediction has proven to be a vexing one and improved solutions continue to be sought (Thomas *et al.*, 2010).

RNA22 (Miranda *et al.*, 2006) differs from other algorithms in several ways: it is trained with known miRNAs, not with experimentally validated heteroduplexes; it permits combinations of G:U wobbles and bulges in the seed region of a heteroduplex; and, it neither relies on nor requires a target's conservation across genomes. RNA22 has successfully predicted non-canonical targets as well as targets beyond the 3′-untranslated region (3′-UTR) (Duursma *et al.*, 2008; Lal *et al.*, 2008, 2009; Tay *et al.*, 2008). Also, independent studies (Hammell *et al.*, 2008; Ritchie *et al.*, 2009) have shown that RNA22 exhibits a high signal-to-noise ratio. To facilitate work with RNA22 targets, we implemented

RNA22-GUI, an application that permits user interaction with the server, exploration of various attributes of the predicted targets, visualization of the predictions in context, link-out to external expression repositories and databases, and so forth.

2 IMPLEMENTATION

Our design had the following key requirements: (i) informative and comprehensive visualization that is integrated with context-dependent search and filtering capabilities—a flexible and user-friendly interface permits one to zoom-in on relevant information that is distilled from a large (hundreds of GBytes) repository of pre-computed predictions; (ii) speedy access to a succinct collection of mRNA:miRNA predictions in a multitude of formatting choices that facilitate downstream analyses; (iii) collaborative sharing enablement through a dynamic-in-nature system that permits information transfer through URL exchange; (iv) ability to store the generated data on the user's computer to enable offline studies without the need to connect to our web server; and (v) remote harvesting or downloading of our pre-compiled data (see Supplementary Material).

3 FUNCTIONALITY

The main page of RNA22-GUI allows the user to choose an organism. Currently, predictions are available for *Homo sapiens, Mus musculus, Drosophila melanogaster* and *Caenorhabditis elegans*. The user is then taken to a new page, referred to in what follows as the 'main page'. For the remainder of the presentation, we will work with *H. sapiens*.

3.1 Interactive 'search by …'

Input boxes on the main page permit searches using a miRNA identifier, an ENSEMBL gene or ENSEMBL transcript identifier, or, a common gene name (Supplementary Figure S1). Help buttons next to each input box link to help information specific to the input type, examples of use and an interactive demo with annotated pictures. Input validation is provided throughout the application: ambiguous queries (e.g. hsa-miR-29) lead the system to recommend more specific instances and to prompt the user to select the intended one (e.g. hsa-miR-29**a**, hsa-miR-29**b**, etc). This being a transition period for miRNA nomenclature, we have opted for a solution that allows for maximal flexibility (see Supplementary Material). Malformed or non-existent identifiers also result in user notification. For queries with an ENSEMBL gene identifier, the user is presented with a list of known transcripts for the gene and links to the relevant ENSEMBL records and the gene's entries in the Human Protein Atlas, the Gene Expression Atlas and the Allen Institute for Brain Science repository.

---

*To whom correspondence should be addressed.

A pre-populated link to FACTA+ (Tsuruoka *et al.*, 2011) allows the generation of term associations for the gene's function. Other links allow one to view the 500 most accessed transcripts and to submit questions/comments through email.

3.2 Boolean operations on miRNAs

For queries that use a miRNA identifier, we permit implicit use of the AND operator by allowing the user to enter multiple space-separated miRNA identifiers. The output will include those transcripts that are predicted to be targeted simultaneously by the listed miRNAs. We are in the process of implementing a parser that will allow the use of more complicated Boolean expressions e.g. (miR-17 AND miR-18a) AND (miR-19a OR miR-19b-1) to minimize the amount of needed offline processing by the user.

3.3 Interactive tabular output of predicted targets

After selecting the filtering criteria, a simple or a composite miRNA query will generate a multi-column list (Supplementary Figure S2) of all predicted targets satisfying the criteria. This output includes common gene name, ENSEMBL gene and transcript identifiers, the total number of predicted targets in the transcript for the specified miRNA(s), chromosome and strand information, links to tabular and html-based maps of the predicted targets for the transcript for all currently known miRNAs, links to the matching ENSEMBL records and a description of each transcript's function. The user can interact with the output in multiple ways sorting the rows in ascending or descending order of the entries in the various columns, or can cut-and-paste it into a spreadsheet program. Importantly, the user can save the entire web page on his/her computer for subsequent offline analysis. The user can also generate single column lists with gene or transcript identifiers or common gene names for use with other tools, e.g. DAVID (Huang *et al.*, 2009).

3.4 *In-context* interactive visualization of predictions

Queries with a transcript identifier allow the retrieval and visualization of predicted targets in the form of a navigable map. The engine will render the corresponding cDNA on consecutive lines each containing 60 nucleotides. In the map, the 5′ and 3′-UTRs are shown in a separate color than the coding sequence and the amino-acid translation is juxtaposed to the cDNA. Predicted targets sites (Fig. 1 and Supplementary Fig. S3) are optionally delineated with thin red rectangles superimposed on the cDNA. Regions of cDNA regions predicted to be hotspots of miRNA targeting are rendered in gradually darker shades of gray and are automatically updated if the user changes the filtering criteria. A pull-down menu at the top of the page lists the targeting miRNAs and their target locations in miRNA numerical order. Selecting a miRNA/target-location allows the user to jump to the corresponding location of the page. Pull-down menus next to each cDNA line permit one to selectively show one or more predicted targets and the corresponding heteroduplexes. The heteroduplexes are aligned to the cDNA and shown in a semi-transparent pop-up window. Filtering criteria (e.g. folding energy of a heteroduplex) can be changed on-the-fly; detailed help is available within the filtering dialog. Notably, toggling target visualization is separate and distinct from the changing of the filtering options. As with the tabular output, the cDNA map can be saved on the user's
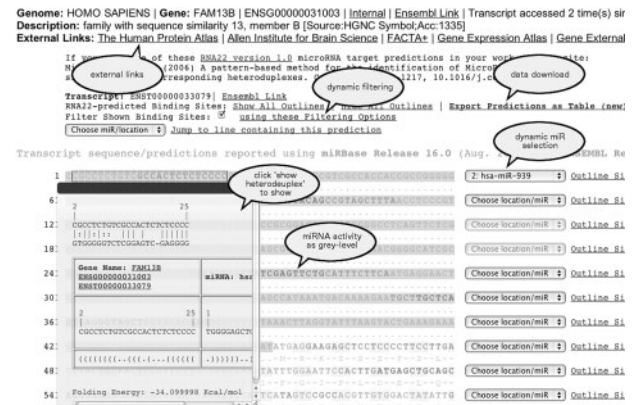

Fig. 1. An interactive working session with the graphical representation of a cDNA map. See text and Supplementary Material for a discussion

local computer and studied 'offline'. Portions or all of it can be shared with colleagues as an attachment, included in presentations, and so forth.

REFERENCES

Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

Duursma,A.M. *et al.* (2008) miR-148 targets human DNMT3b protein coding region. *RNA*, **14**, 872–877.

Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.

Hammell,M. *et al.* (2008) mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods*, **5**, 813–819.

Huang,D.W. *et al.* (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.

Lal,A. *et al.* (2008) p16(INK4a) translation suppressed by miR-24. *PLoS One*, **3**, e1864.

Lal,A. *et al.* (2009) miR-24 inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to 'seedless' 3′UTR microRNA recognition elements. *Mol. Cell*, **35**, 610–625.

Miranda,K.C. *et al.* (2006) A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.

Pasquinelli,A.E. (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.*, **13**, 271–282.

Rigoutsos,I. and Tsirigos,A. (2010) MicroRNA target prediction. In Slack,F. (ed.) *MicroRNAs in Development and Cancer*. Imperial College Press/World Scientific Press, UK, pp. 237–259.

Ritchie,W. *et al.* (2009) Predicting microRNA targets and functions: traps for the unwary. *Nat. Methods*, **6**, 397–398.

Tay,Y. *et al.* (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**, 1124–1128.

Thomas,M. *et al.* (2010) Desperately seeking microRNA targets. *Nat. Pub. Group*, **17**, 1169–1174.

Tsuruoka,Y. *et al.* (2011) Discovering and visualizing indirect associations between biomedical concepts. *Bioinformatics*, **27**, i111–i119.