miRecords: an integrated resource for microRNA–target interactions

Feifei Xiao1, Zhixiang Zuo1, Guoshuai Cai1, Shuli Kang1, Xiaolian Gao2 and Tongbin Li1,*

1Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, 2Department of Biology and Biochemistry, University of Houston, Houston, TX 77004, USA

Received August 15, 2008; Revised September 21, 2008; Accepted October 16, 2008

ABSTRACT

MicroRNAs (miRNAs) are an important class of small noncoding RNAs capable of regulating other genes’ expression. Much progress has been made in computational target prediction of miRNAs in recent years. More than 10 miRNA target prediction programs have been established, yet, the prediction of animal miRNA targets remains a challenging task. We have developed miRecords, an integrated resource for animal miRNA–target interactions. The Validated Targets component of this resource hosts a large, high-quality manually curated database of experimentally validated miRNA–target interactions with systematic documentation of experimental support for each interaction. The current release of this database includes 1135 records of validated miRNA–target interactions between 301 miRNAs and 902 target genes in seven animal species. The Predicted Targets component of miRecords stores predicted miRNA targets produced by 11 established miRNA target prediction programs. miRecords is expected to serve as a useful resource not only for experimental miRNA researchers, but also for informatics scientists developing the next-generation miRNA target prediction programs. The miRecords is available at http://miRecords.umn.edu/miRecords.

INTRODUCTION

MicroRNAs (miRNAs) are a class of small (19–27 nt) noncoding RNAs capable of base-pairing to the transcripts of protein-coding genes (which are termed the targets of the miRNAs), leading to downregulation or repression of the targeted genes(1–3). The miRNA gene family is one of the largest in higher eukaryotes: more than 700 mature miRNAs have been identified in the human genome, according to the current release of miRBase (4), and these miRNAs account for >2.5% of all human genes. The exact mechanisms by which miRNAs regulate their target genes’ expression remain obscure, although several models have been proposed [see recent reviews, e.g. (5,6)], suggesting that miRNAs could induce translation repression at both the initiation phase and the elongation phase of translation. Alternatively, the translation may not be directly affected, but miRNAs could lead to rapid proteolysis of nascent polypeptides; or they may lead to accumulation of target miRNAs in the P-bodies, isolating them from the translation machinery (5,6). The miRNAs could also lead to degradation of the target transcripts. Recent evidence suggests that miRNA-induced target transcript degradation through a complex process that includes the deadenylation and decapping of miRNAs, distinct from the siRNA-induced RNA silencing mechanism (6,7).

Computational prediction of miRNA targets is much more challenging in animals than in plants, because animal miRNAs often form imperfect base-pairing with their target sites (3,8–11), unlike plant miRNAs which almost always bind their targets with near perfect complementarity (12). In the past several years, a large number of target prediction programs have been developed for animal miRNAs [see recent reviews, e.g. (8,10,11,13)]. Several of the earliest target prediction programs [including TargetScan (14)/TargetScanS (15), PicTar (16), miRanda (17,18), DIANA-microT (19) and MicroInspector (20)] adopted hand-derived rules based on a number of principles summarized from known miRNA–target interactions. These principles emphasize: (i) near-perfect complementarity in the 6–8 nt region close to the 5’ end of the miRNA (the so-called ‘seed’ region) with the 3’UTR region of the target sequence; (ii) evolutionary conservation of the target sequences between species; (iii) strong thermodynamic stability of miRNA–mRNA duplex; (iv) cooperativity between multiple sites in close

*To whom correspondence should be addressed. Tel: +1 612 626 3481; Fax: +1 612 626 5009; Email: toli@biocompute.umn.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2008 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
proximity; and (v) existence of a central nonmatched region (forming a loop or bulge). RNAhybrid (21), one of the earliest developed miRNA target prediction program, implemented an efficient algorithm that finds energetically most favorable hybridizations between the miRNA and mRNA molecules, avoiding intramolecular hybridizations. A newer miRNA target prediction program, Rna22 (22), applied a procedure that first finds significant sequence motifs (or ‘patterns’) among all known miRNA sequences, then defines ‘target islands’, or regions in mRNAs where reverse complements of the miRNA patterns aggregate and focuses on these target islands when searching for miRNA target sites. Most recently developed miRNA target prediction programs [miTarget (23), MirTarget2 (24) and NBmiRTar (25)] employed machine learning techniques to construct predictors directly from validated miRNA target datasets. Furthermore, several recent studies suggested that target site accessibility is an important factor for effective targeting of miRNAs (26–28). One program, PITA, was developed to make predictions based on target site accessibility features.

Comparative studies conducted with the earlier miRNA target prediction programs (including TargetScanS, PicTar, miRanda, DIANA-micoT, RNAhybrid, MicroInspector and Rna22) suggested that no one program was consistently superior to all others (13,29). Indeed, it has become a common practice for experimental researchers to look at predictions made by several target prediction programs and focus on their intersections (30,31). Machine learning methods have good potential to produce more accurate target predictors. Yet, recently developed machine-learning-based target prediction tools (miTarget, MirTarget2 and NBmiRTar) have not been through rigorous independent assessments. The performance of machine learning-based methods is known to be heavily dependent on the quantity and quality of the dataset used in the training. The lack of large, high-quality datasets of experimentally validated miRNAs will likely be the bottleneck for developing more accurate miRNA target prediction programs.

Here, we present miRecords, a new resource for animal miRNA–target interactions. miRecords consists of two components. The Validated Targets component is a large, high-quality database of experimentally validated miRNA targets resulting from meticulous manual literature curation. As the largest known collection of experimental validated miRNA targets, it emphasizes systematic and structured documentation of experimental support of miRNA–target interactions. This database not only serves the experimental researchers by providing the lists of confirmed targets of the miRNAs of their interest, but also provides a large and high-quality dataset that will facilitate the development of the next-generation miRNA target prediction programs. The Predicted Targets component of miRecords is an integration of predicted miRNA targets produced by 11 established miRNA target prediction programs. As the most complete integration of predicted miRNA targets, it is expected to provide considerable help to researchers investigating new miRNA targets.

### DATABASE CONTENT

#### Key issues in documenting validated targets

The Validated Targets component of miRecords is designed to serve as a centralized archive of confirmed miRNA–target interactions with systematic documentation of experimental support. To ensure the usefulness of this database to the miRNA research community at large, it is critical to identify key issues involving experimental validations of miRNA–target interactions, and carefully incorporate them into the documenting system.

**Endogenous miRNAs versus exogenous miRNAs.** A majority of studies of miRNA–target interactions were performed by experimental manipulations of the level of a miRNA (either by overexpression/misexpression or by underexpression) in a cell line or a tissue, followed by examination of changes in expression of the putative targets. These experiments could be generally classified as ‘exogenous miRNA experiments’.

Concerns have been raised, however, over how many of these miRNA–target interactions actually take place in endogenous, physiological conditions (32,33). Similarly to the gene regulation by transcription factors, endogenous miRNA may require favorable cellular context to bind and regulate their targets, which cannot be replicated in exogenous miRNA experiments.

An increasing number of studies have provided evidence about endogenous miRNA–target interactions. For example, in Ref. (34), when investigating whether *let-60* was a target of the miRNA *let-7* in the hypodermal seam cells in *Caenorhabditis elegans*, the authors fused *let-60 3′UTR* behind the *Escherichia coli lacZ* gene. They discovered that the reporter activity was downregulated at the L4 stage of the development, when *let-7* was known to be expressed in the seam cells. In contrast, the same reporter gene fused to an irrelevant control 3′UTR was expressed at all stages. As another example, in Ref. (33), the targeting of *cog-1* by the miRNA *ksy-6* was studied in the ASEL and ASER, two closely related bilaterally symmetric neurons in *C. elegans*. The ASEL, but not the ASER, neuron expresses endogenous *ksy-6*. The authors fused the *cog-1 3′UTR* to a green fluorescent protein (GFP) sensor construct, and found that this fusion was effectively downregulated in ASEL but not in ASER. In several other studies, the endogenous levels of miRNAs and the protein expression levels of their potential targets were measured simultaneously across several cell lines or specimens, and inverse correlations were observed between them (35–37).

In miRecords, we make a clear distinction between endogenous miRNA experiments and exogenous miRNA experiments. For each study involving endogenous miRNA experiments, we provide a brief summary about the rationale of the experiments as well as an explanation of the results.

**Target genes, target regions and target sites.** We classify any experimental evidence about miRNA–target interaction as belonging to one of the three levels: the target gene level, the target region level and the target site level. When the evidence indicates that the level of the full-length gene...
product (mRNA or protein) of a putative target has reduced following over- or misexpression of a miRNA, or that the full-length gene product has accumulated following underexpression of the miRNA, it is considered as target gene level evidence. The target gene level evidence also includes endogenous miRNA experiments leading to the finding of inverse correlations between the endogenous miRNA levels and the full-length protein products of the putative target genes. The target gene level experiments are often regarded as indirect support of the miRNA–target interactions, because the level of the gene product may change due to other reasons, e.g. a change in expression of another protein (which is a true target of the miRNA) that it interacts with.

When the experimental evidence indicates that a region of the mRNA of the putative target (shorter than the full-length transcript) is responsible for the miRNA–target interaction, it is documented as target region level evidence. Most target region level experiments were conducted with fusion of the 3′UTR of the putative target gene (or a section of the 3′UTR) to a reporter construct (e.g. luciferase or GFP), followed by observations that the reporter expression is downregulated (or upregulated) in response to overexpression/misexpression (or underexpression) of the miRNA.

When an experiment points to a very short section of the mRNA (whose length is comparable with that of the miRNA) as being responsible for the miRNA–target interaction, it is classified as target site level evidence. The target site level experiments include reporter assays with fusion constructs made with short target sites, and target site mutation experiments (discussed below).

Over- or misexpression and underexpression of miRNAs. The exogenous miRNA experiments can be broadly classified into two categories based on the methods by which the miRNA levels are manipulated: miRNA overexpression or misexpression experiments, and miRNA underexpression experiments. The methods commonly applied to over- or misexpress miRNAs include mature miRNA transfection (38), miRNA precursor transfection (39) and indirectly induced miRNA overexpression [using the DNA demethylating agent 5-Aza-Deoxycytidine (40), or the histone deacetylase inhibitor phenylbutyrate (PBA) (41)]. The techniques applied to underexpress miRNAs include miRNA knockdown by siRNAs (42), miRNA knockdown by antisense modified oligonucleotides, e.g. morpholinos (43), locked nucleic acids (44), or 2′-O-Me oligonucleotides (45) and knockout of the miRNA gene (46).

Reporter assays, mRNA- and protein-level measurements. The means by which putative target expression levels are examined can be classified into four categories: reporter assays, mRNA-level measurements, protein-level measurements and ‘others’. In a reporter assay, the putative target region or target site is fused with a reporter vector, and the expression level of the putative target region or target site is quantified by measuring the reporter’s activity. Commonly used reporters include luciferase (16), GFP (47), YFP (43) and the \( \text{lacZ} / \beta\)-galactosidase reporter (48). Several methods of measuring mRNA levels of putative targets can be applied. They include RT–PCR (28), northern blot (49), 5′RACE (50), DNA microarrays (51), ribonuclease protection assay (52) and branched DNA probe assay (53). Commonly applied protein-level measuring methods include western blot (54), ELISA (55) and immunocytochemistry (56). The ‘others’ category includes rare target expression analyses that do not belong to other categories, e.g. phenotype analysis (43).

Target site mutations. In a target site mutation experiment, point mutations are introduced to the putative target site. If the introduced point mutations lead to abolishment of the miRNA-mediated downregulation, the site is convincingly verified as a true target site. Besides validating miRNA target sites, target site mutation experiments are frequently conducted in studies investigating general features that influence the miRNA targeting, e.g. it was applied to study the importance of the 5′ seed region for miRNA–target interactions (9), and of target accessibility features in assisting miRNA target discovery (26).

Web interface

The miRecords resource can be accessed through the URL http://miRecords.umn.edu/miRecords. In the Validated Targets section, the user can first select a species, then choose from a list of miRNAs for which experimentally validated targets have been documented. Optionally, the user can provide the RefSeq accession, the Entrez Gene ID or the gene name of the target gene and initiate a search. The result of the search is presented as a list of miRNA–target interactions. For each miRNA–target interaction, the information about the miRNA and about the target gene is displayed together with the prediction results of 11 established miRNA target prediction programs. Positive predictions are indicated by lit up symbols. The user can click a lit up symbol to obtain more information about the prediction.

When the user clicks the ‘Click for detail’ link displayed in the ‘Target Interaction’ column, detailed information about a validated miRNA–target interaction is presented. The detailed information about the miRNA includes the Stemloop ID and accession number, mature miRNA ID and accession number and the sequence of the miRNA. The detailed information about the target gene includes the RefSeq accession number, name, synonyms and the description of the gene.

Underneath the general information about the miRNA and the target gene, all records, or literature accounts of the miRNA–target interaction are presented. On the top of each record is the citation information, underneath which is the experimental support of the miRNA–target interaction, listed in the order of the target gene level evidence, followed by the target region level evidence, then by the target site level evidence. In each of the three levels of evidence, endogenous miRNA experiments (when available) are described first, followed by summaries of exogenous miRNA experiments. The summary of a target gene level, exogenous miRNA experiment includes
the method of manipulating the miRNA level (e.g. ‘over-expression by mature miRNA transfection’, or ‘under-expression by siRNA knock-down’), reporter assay performed (e.g. luciferase assay), mRNA-level measurement performed (e.g. RT–PCR or 5’RACE), mRNA-level measurement result (e.g. ‘down-regulated’), protein-level measurement performed (e.g. western blot) and protein-level measurement result. The user can click the ‘Detail’ button for each individual experiment, and the description of the experiment from the original publication is displayed.

The summary of target gene level or target site level evidence also includes the section of the mRNA that was used in the fusion reporter assay, and about whether point mutation experiments were performed. When available, a drawing of the alignment between the miRNA and the putative target site is also presented.

The user can download the complete dataset of validated targets at the Download Validated Targets section accessible from the top of the web site.

The Predicted Targets section is organized similarly to the Validated Targets section. When the user submits a query, all putative targets (including unvalidated ones) predicted by one or more of the established miRNA target prediction programs are presented. The user can choose to display the result according to the predictions made by a particular target prediction program, and/or apply a filter to display putative targets predicted by at least a user-specified number of prediction programs. Moreover, an option is provided for the user to download the predicted targets of interest to his or her local computer for further examination.

Data access

The miRecords web site is publically accessible through the URL http://miRecords.umn.edu/miRecords. Additional requests can be made by emailing to miRecords@biocompute.umn.edu.

### Table 1. Comparison in data content between TarBase and miRecords

<table>
<thead>
<tr>
<th></th>
<th>Total number of miRNAs</th>
<th>Total number of target genes</th>
<th>Total number of records</th>
<th>Number of low-throughput records</th>
<th>Number of human miRNAs</th>
<th>Number of human target genes</th>
<th>Number of human records</th>
</tr>
</thead>
<tbody>
<tr>
<td>TarBase (57)</td>
<td>128</td>
<td>570</td>
<td>626</td>
<td>279</td>
<td>62</td>
<td>415</td>
<td>458</td>
</tr>
<tr>
<td>miRecords</td>
<td>301</td>
<td>902</td>
<td>1135</td>
<td>639</td>
<td>125</td>
<td>651</td>
<td>778</td>
</tr>
</tbody>
</table>

*aLow-throughput records refer to records from non-DNA-microarray studies.

### Table 2. Comparison in miRNA target prediction programs implemented among integrated miRNA target resources

<table>
<thead>
<tr>
<th>Tool</th>
<th>URL</th>
<th>DIANA-microT</th>
<th>Micro-Inspector</th>
<th>mi-Randa</th>
<th>mi-Target</th>
<th>Mir-Target2</th>
<th>NBehTar</th>
<th>PicTar</th>
<th>PITA</th>
<th>RNA22</th>
<th>RNA Hybrid</th>
<th>TargetScan/TargetScanS</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRGator (58)</td>
<td><a href="http://genome.ewha.ac.kr/miRGator/miRGator.html">http://genome.ewha.ac.kr/miRGator/miRGator.html</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>miRGen (31)</td>
<td><a href="http://www.diana.pcbi.upenn.edu/miRGen.html">http://www.diana.pcbi.upenn.edu/miRGen.html</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MiRNAMap (59)</td>
<td><a href="http://mirnamap.mbc.ntcu.edu.tw/">http://mirnamap.mbc.ntcu.edu.tw/</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>miRecords</td>
<td><a href="http://miRecords.umn.edu/miRecords">http://miRecords.umn.edu/miRecords</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
REFERENCES


