**STarMirDB: a database of microRNA binding sites**

William Rennie[1], Shaveta Kanoria[1], Chaochun Liu[1], Bibekanand Mallick [1,2], Dang Long [1,3], Adam Wolenc[1], C. Steven Carmack[1], Jun Lu[4] and Ye Ding[1,*]

*Correspondence:

Email: ye.ding@health.ny.gov; Phone: (518) 486 1719; Fax: (518) 474 3181

[1] Wadsworth Center, New York State Department of Health, Center for Medical Science, 150 New Scotland Avenue, Albany, NY 12208, USA

[2] Present address: RNA Biology and Functional Genomics Laboratory, Department of Life Science, National Institute of Technology, Rourkela-769008, Odisha, India

[3] Present address: Biotechnology Department, Faculty of Chemistry, Danang University of Science and Technology, 54 Nguyen Luong Bang St., Danang, Vietnam

[4] Department of Genetics and Yale Stem Cell Center, Yale University, New Haven, CT 06520, USA

**Abstract**

microRNAs (miRNAs) are an abundant class of small endogenous non-coding RNAs (ncRNAs) of ~22 nucleotides (nts) in length. These small regulatory molecules are involved in diverse developmental, physiological and pathological processes. miRNAs target messenger RNAs (mRNAs) for translational repression and/or mRNA degradation. Predictions of miRNA binding sites facilitate experimental validation of miRNA targets. Models developed with data from CLIP studies have been used for predictions of miRNA binding sites in the whole transcriptomes of human, mouse and worm. The prediction results have been assembled into STarMirDB, a new database of miRNA binding sites available at http://sfold.wadsworth.org/starmirDB.php. STarMirDB can be searched by miRNAs or mRNAs separately or in combination. The search results are categorized into seed and seedless sites in 3′ UTR, CDS and 5′ UTR. For each predicted site, STarMirDB provides a comprehensive list of sequence, thermodynamic and target structural features that are known to influence miRNA: target interaction. A high resolution PDF diagram of the conformation of the miRNA:target hybrid is also available for visualization and publication. The results of a database search are available through both an interactive viewer and downloadable text files.

**Keywords**

microRNA, CLIP, binding site prediction, seed site, seedless site

## Introduction

miRNAs are a class of single-stranded, non-coding RNAs of ~22 nucleotides in length. They have been discovered in plants, animals as well as in some viruses [1-3]. miRNAs play essential roles in cell proliferation, differentiation, development, and are associated with human diseases [2, 4]. A mature miRNA can guide miRNA-induced silencing complex (miRISC) for target recognition by sequence complementarity between the miRNA and sequences typically in the 3′ untranslated regions (3′ UTRs) of the cognitive messenger RNAs (mRNAs). Successful target binding usually results in translational repression and/or mRNA degradation [5]. Each human miRNA is predicted to be able to regulate several hundred different mRNAs [6].

Computational prediction algorithms have proven to be valuable in the discovery of new miRNA targets. Most of the existing algorithms are based on the seed rule, i.e., the target site within 3′ UTR forms Watson-Crick (WC) pairs with bases at positions 2 through 7/8 of the 5′ end of the miRNA [7]. However, numerous exceptions to the seed rule have been well-documented [8-13]. Other sequence features have been proposed based on their enhancement of targeting specificity. These include sequence conservation, strong base-pairing to the 3′ end of the miRNA, local AU content and location of miRNA binding sites (near either end of the 3′ UTR is favorable)[14]. The importance of target structural accessibility for miRNA target recognition has been supported by numerous studies [15-21].

In recent years, experimental methods based on cross-linking immunoprecipitation (CLIP) have been developed. For human and mouse studies, these include HITS-CLIP [22], PAR-CLIP [23] and variations of such techniques [24]. The CLIP approach has also been successful in worm [25]. The CLIP studies have provided high throughput quality datasets for regions of mRNAs containing miRNA binding sites. These data allowed us to develop models for improved predictions of miRNA binding sites [18, 26]. The models are based on a comprehensive list of sequence, thermodynamic and target structure features that were enriched for miRNA binding sites identified from CLIP data, and were validated by intra-dataset, inter-dataset as well as cross-species validations. For human, mouse and worm, we have used these models to carry out transcriptome-scale predictions of both seed and seedless sites in the 3′ untranslated region (3′ UTR), coding sequence (CDS) region, and 5′ untranslated region (5′ UTR) of mRNAs. The results have been assembled into STarMirDB, a new database application module of the Sfold

RNA package [27, 28]. In this article, we describe this new resource. The unique tools of STarMirDB shall complement the existing miRNA target resources for computational predictions and experimental target data. Examples of these include, but are not limited to, TargetScan [29], Diana-microT [30], TarBase [31] , StarBase [32], miRecords [33] and miRTarBase [34].

**Generation of transcriptome-scale data for STarMirDB**

The database currently contains records for three species, *H. sapiens* (human), *M. musclus* (mouse) and *C. elegans* (worm). For human and mouse, we used complete mRNA sequences from NCBI RefSeq build 36.3 and 37.2, respectively. For worm, 3' UTR sequences were obtained from the Wormbase version WS-190. The current release of STarMirDB includes 38,745 transcripts for human, 34,631 for mouse and 22,926 3' UTRs for worm. miRNA sequences were obtained from miRBase release 18 [35]. We collected 1,921 miRNA sequences for human, 1,157 for mouse and 368 for worm.

Our CLIP based models were used to make transcriptome scale predictions of both seed and seedless binding sites [18]. For each site, a comprehensive list of sequence, thermodynamic and target structure features are computed (**Table 1**). A logistic probability is provided as a measure of confidence in the predicted site. The number of binding sites is astronomical, so that in the database we only included those with a probability of 0.5 or higher. This filter also helps assure a reasonable response time for database search queries. In the case of interest in those low confidence sites with probabilities under 0.5, the user can use the STarMir web server that presents all predicted sites for single or multiple miRNAs and a target mRNA [36]. The database can be searched by one or more miRNAs or targets, separately or in a combination. For worm, we provide a user interface that allows developmental stage specific search of miRNA binding sites within the 3′ UTR of transcripts. This interface is activated when *C. elegans* is selected as the species for database search. Additionally, for *C. elegans*, all the prediction data for miRNA binding sites within the 3′ UTR of transcripts are also provided as downloadable files.

**Input of database query**

STarMirDB presents a collection of predicted miRNA binding sites on mRNAs through a web interface that enables both search and retrieval of data and visualization of the conformation of predicted miRNA:target hybrid for each predicted site. The web interface has four input fields:

species, miRNAs, mRNAs, and logistic probability threshold. The requirements for each input field are described below in detail.

To start the database search, the user should first select the species from a dropdown menu. Currently three species are included in the database: human (*Homo sapiens*), mouse (*Mus musculus*) and worm (*C. elegans*). Next, one or a set of miRNAs can be selected from the miRNA scroll down list, which displays all available miRNAs assembled for the selected species. Additionally, one or more miRNA names can be entered in the text box. The database follows the naming convention used by miRBase [35], i.e., all the miRNAs can be identified by their miRNA name/identifier (e.g., *hsa-let7-5p* for human, *mmu-let7a* for mouse, and *cel-mir-1018* for worm).

Target mRNA information has to be entered into the provided text box. For human and mouse, either Genbank accession number or Gene symbol, as assigned by the HUGO Gene Nomenclature Committee (HGNC), can be provided. For worm, Wormbase ID is required. For search result display through an interactive site viewer, a user can choose to display only the most relevant site features for each binding site, or the complete list of features. The most relevant features are considered by us to be the most informative. They were selected from those used in the development of the prediction models[18, 26]. A user may choose to input merely miRNAs while leaving the target input box blank. In this case, the database server will retrieve predicted sites for the entire transcriptome assembled for the species. The user can also choose to input merely mRNA IDs, which will prompt the database server to identify all miRNAs assembled for the species that have binding sites on those mRNAs. This can be useful, e.g., when the question is whether an mRNA is targeted by any miRNA. A database search is typically instantaneous. However, if the database is queried with only miRNA(s) without target information, the search will take minutes. Finally, the user can use a drop down menu to filter out miRNA binding sites with logistic probabilities below the specified threshold.

**Output of database query**

Relevant data in the database are retrieved in response to a specific database query and are available through both an interactive site viewer and downloadable files. For the interactive site viewer, the data is classified into three mRNA regions (5′ UTR, CDS and 3′ UTR) and seed and seedless sites. To facilitate online viewing, the number of sites displayed in the interactive area is limited to top-ranked sites according to the decreasing order of their logistic probabilities. By

default, 100 binding sites are displayed. Alternatively, the user can choose to display the top 250, 500 or 1,000 sites. The results of a search are also available for download as text files, wherein all of the retrieved binding sites are listed. The interactive viewer presents the results with either the most relevant site features or all of the site features as specified by the user in the input page. The downloadable text files provide all site features. In the text files, features marked with an asterisk are those used in the model computations of the logistic probabilities. In addition to comprehensive sequence, thermodynamic and target structural features (**Table 1**), a high resolution PDF diagram of the conformation of the miRNA:target hybrid is also provided. The diagram was developed to be high quality so that the user can choose to use them for publication purposes. When both the miRNA and the mRNA were included in the CLIP study for the prediction model development[18, 26], an indicator field named "CLIP" will be given a value of 1 if the predicted site is support by the CLIP data, and 0 otherwise. CLIP studies are limited to abundant miRNAs and expressed transcripts. When either the miRNA or the mRNA was absent in the CLIP study, a value of "NA" is assigned to the CLIP indicator. In the database, less than 1% of sites have a CLIP indicator value of 0 or 1. Thus, our prediction data complement the CLIP data. A file providing definitions of site features is available via the link for 'Feature definitions' under the table listing predicted sites.

## Illustration of database search

For an illustration of the database search, **Figure 1** shows the input screen for a query starting with 'Human (V-CLIP; NCBI RefSeq Build 36.3)' selected in the species dropdown menu. From the dropdown list of miRNAs, *hsa-7a-5p*, *hsa-7b-5p*, and *hsa-7d-5p* were selected. In addition, two miRNAs, *hsa-let-7c* and *hsa-let-7e* were manually entered. For mRNA targets, accession numbers NM_0000024, NM_0000021, NM_0000017, and Gene symbol *Lin54* were entered. Next, the option of "Show predictions with the most relevant features in interactive viewer" was selected. Finally, a logistic probability threshold of 0.6 was selected. The "Search" button was then clicked for submitting the query input information for processing by the database server.

Upon completion of data retrieval by the database server, the user is presented with an interactive site viewer (**Figure 2**). By default, the list of the top 100 sites is displayed in decreasing order of logistic probabilities. An alternative number of sites can be selected from a dropdown menu. The tab for "3′ UTR-seedless" was selected for presenting seedless sites in

the 3′ UTR of the target. For example, the first entry in the site table has a logistic probability of 0.8669, which indicates a high confidence in this predicted site. A rather low value of −21.26 kcal/mol for $\Delta G_{total}$ indicates a high structural accessibility at the target site[17]. In the "Hybrid Conformation" column, a link is provided for a high resolution PDF diagram of the conformation of the miRNA:target hybrid at the predicted site. Clicking this link will open the diagram in a new tab or window, depending on the configuration of the user's web browser. Multiple windows/tabs facilitate comparison of hybrid conformations for multiple binding sites. **Figure 3** shows hybrid diagrams for a seed site and a seedless site.

Under the interactive site viewer, links are provided for downloading files of the query results for the six combinations of regions and site types (**Figure 2**). For site feature information, the downloadable files provide all site features whereas the interactive viewer displays either all or the most relevant features as selected by the user in the query input page. The user can initiate a new search by clicking on the link at the bottom of the page.

**Conclusions**

STarMirDB is a new bioinformatics resource for facilitating miRNA target studies. The current release of database includes 96,302 mRNAs and 3,446 miRNAs for human, mouse and worm. It will be periodically updated and likely extended to other species. It presents predictions for all three mRNA regions and for both seed and seedless sites. Importantly, it presents a probability for each site as an indicator of confidence in the prediction. In addition to use for visualization and publication, high quality diagrams of miRNA:target hybrids can facilitate design of nucleotide mutations for experimental validation of binding sites. The option for search by developmental stage shall be useful for studies of miRNAs in worm. The unique tools from STarMirDB will complement the existing miRNA target resources for computational predictions and experimental target data. The database can retrieve miRNA binding sites for single or multiple miRNAs and/or one or more targets. For example, this capability will be useful for elucidating miRNA regulation of genes of interest. It will also be useful in miRNA overexpression and knockout studies, wherein differentially expressed genes can be further examined by prediction and validation of miRNA binding sites.

We have also developed STarMir, a web server for prediction of miRNA binding sites [36]. STarMir and STarMirDB are complementary tools. While the database allows fast search of pre-

computed results, STarMir makes predictions for any miRNA:mRNA pair from any species of interest. For example, the user can use STarMir in making predictions for a new isoform absent in the current database release.

The provision of extensive predictions of seedless sites (i.e., non-canonical sites) is a major feature for both the database and the web server. The functionality of seedless sites has been demonstrated by numerous studies based on diverse methods, which include reporter assays, nucleotide mutation analysis, analysis of microarray data, analysis of proteomics data, and phenotypic analysis [11, 37-43]. However, a study primarily based on microarray data failed to find support for functional seedless sites [44]. Further experimental investigations will be helpful for addressing this lack of consensus. Our tools will facilitate experimental testing of predicted seedless sites, especially those with high logistic probabilities.

**Acknowledgements**

The Bioinformatics Core at the Wadsworth Center is acknowledged for supporting computing resources for this work. This work is supported in part by the National Science Foundation (DBI-0650991 to Y.D.), National Institutes of Health (GM099811 to Y.D. and J. L.).

**Author Contributions**

Y.D. conceived and supervised the study. W.R. performed development, implementation and deployment of the database. C.L. and B.M. contributed to generation of data for the database; S.K. and J.L. performed testing of database interface. D. L. and A.W. wrote the initial software for the computation of several target site features used by the database, and C.C. provided hardware and system support. W.R., S.K. and Y.D. wrote the paper with contributions from all authors. All authors read and approved the final manuscript.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Figure legends**

**Figure 1.** An illustrative example of the STarMirDB input page. Binding sites are searched for miRNAs selected from a pre-stored list as well as manually entered by a user, multiple targets,

and a specific logistic probability threshold selected by the user. The option of the most relevant site features is selected for output display.

**Figure 2.** STarMirDB output page for the default display of top 100 sites, with the tab selected for displaying seedless sites in the 3' UTR.

**Figure 3.** Conformation diagrams of miRNA:target hybrids for a seed site (A) and a seedless site (B).

| **Table 1. Description of site information and features for STarMirDB output** | |
|---|---|
| Site ID | Predicted sites are sequentially numbered along the target sequence |
| Target | Accession number of the target mRNA |
| Gene | Gene symbol of the target mRNA |
| miRNA | Name of the microRNA (miRNA) |
| Target_Len | Length of the target |
| Site_Position | Start and end position of the target region (site) predicted to be bound by miRNA |
| Seed_Position | Start and end position of the target sub-region complementary to the miRNA seed (i.e. positions 2-7/8 of the miRNA) |
| Seed_Type | 6mer, offset 6mer, 7mer-A1, 7mer-m8, and 8mer seed sites [14, 45] |
| Site_Access | A measure of structural accessibility as computed by the average probability of a nucleotide being single-stranded (i.e., unpaired) for the nucleotides in the predicted binding site [18] |
| Seed_Access | A measure of structural accessibility as computed by the average of single-stranded probabilities of the nucleotides in the target sub-region complementary to the miRNA seed [18] |
| Upstream_Access (# nt) | A measure of structural accessibility as computed by the average of single-stranded probabilities for the block of nucleotides upstream of the predicted binding site (# nt: block size of 5, 10, 20, 25 or 30) [18] |
| Dwstream_Access (# nt) | A measure of structural accessibility as computed by the average of single-stranded probabilities for the block of nucleotides downstream of the predicted binding site [18] |
| Upstream_AU (# nt) | Percentage of AU for the block of nucleotides upstream of the binding site |
| Dwstream_AU (# nt) | Percentage of AU for the block of nucleotides downstream of the binding site |
| Site_Location | Relative starting location of the predicted binding site along the length of the sequence (e.g., for 3′ UTR, 0 indicates the 5′ end of the UTR, and 1 corresponds to the 3′ end)[14] |
| 3′_bp | Presence of contiguous Watson Crick base pairing for miRNA nucleotide positions 12-17 (sites with 3′_bp are also called 3′ compensatory/supplementary sites)[14] |
| Site_Consv | Conservation score by the PhastCons program [46] for the binding site |
| Seed_Consv | Conservation score by the PhastCons program for the target sub-region complementary to the miRNA seed |
| Offseed_Consv | Conservation score by the PhastCons program for nucleotides within the target site, but outside the seed complementary region |
| dG_hybrid | $\Delta G_{hybrid}$ (in kcal/mol): a measure of stability for miRNA:target hybrid as computed by RNAhybrid [45] |
| dG_nucl | $\Delta G_{nucl}$ (in kcal/mol): a measure of the potential of nucleation for miRNA:target hybridization[17] |
| dG_total | $\Delta G_{total}$ (in kcal/mol): A measure of the total energy change of the hybridization[17] |
| LogitProb | Logistic probability of the site being an miRNA binding site as predicted by our logistic model) [18] |
| Target_Mismatch | Nucleotides in the target binding site that are not base paired with the miRNA |
| Target_Match | Nucleotides in the target binding site that are base paired with the miRNA |
| Mir_Match | Nucleotides in the miRNA that are base paired with the target mRNA |
| Mir_Mismatch | Nucleotides in the miRNA that are not base paired with the target mRNA |
| Hybrid Conformation | The last four fields above present information for the miRNA:target hybrid conformation predicted by RNAhybrid. In each of the fields, spaces are included so the fields can be easily aligned to produce a simple diagram of the hybrid conformation as illustrated below:<br><br>`Target_Mismatch:  U    UUUCC     U           A`<br>`Target_Match:      GACU     AUGUA     CUACCUC`<br>`Mir_Match:         UUGA     UACGU     GAUGGAG`<br>`Mir_Mismatch:                   UGGAU       A` |

**References**


1. Li C, Zhang B. MicroRNAs in Control of Plant Development. J Cell Physiol 2015.
2. Ambros V. The functions of animal microRNAs. Nature 2004; 431:350-5.
3. Liu DG. MicroRNAs in human virus genomes: helping hands for viral infection. Microrna 2014; 3:75-85.
4. Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer. Annu Rev Med 2009; 60:167-79.
5. Fabian MR, Sonenberg N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol 2012; 19:586-93.
6. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009; 19:92-105.
7. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005; 120:15-20.
8. Didiano D, Hobert O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nat Struct Mol Biol 2006; 13:849-51.
9. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature 2008; 455:1124-8.
10. Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. Genes & development 2004; 18:132-7.
11. Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, et al. 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 2009; 35:610-25.
12. Ha I, Wightman B, Ruvkun G. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. Genes & development 1996; 10:3041-50.
13. Stern-Ginossar N, Elefant N, Zimmermann A, Wolf DG, Saleh N, Biton M, et al. Host immune system gene targeting by a viral miRNA. Science 2007; 317:376-81.
14. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009; 136:215-33.
15. Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 2005; 436:214-20.
16. Long D, Chan CY, Ding Y. Analysis of microRNA-target interactions by a target structure based hybridization model. Pac Symp Biocomput 2008:64-74.
17. Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y. Potent effect of target structure on microRNA function. Nat Struct Mol Biol 2007; 14:287-94.
18. Liu C, Mallick B, Long D, Rennie WA, Wolenc A, Carmack CS, et al. CLIP-based prediction of mammalian microRNA binding sites. Nucleic Acids Res 2013.
19. Robins H, Li Y, Padgett RW. Incorporating structure to predict microRNA targets. Proc Natl Acad Sci U S A 2005; 102:4006-9.
20. Hammell M, Long D, Zhang L, Lee A, Carmack CS, Han M, et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. Nat Methods 2008; 5:813-9.
21. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nature genetics 2007; 39:1278-84.
22. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 2009; 460:479-86.
23. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 2010; 141:129-41.

24. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nature methods 2011; 8:559-64.
25. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, et al. Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat Struct Mol Biol, 2010:173-9.
26. Liu C, Rennie WA, Mallick B, Kanoria S, Long D, Wolenc A, et al. MicroRNA binding sites in C. elegans 3' UTRs. RNA Biol 2014; 11:693-701.
27. Ding Y, Lawrence CE. A statistical sampling algorithm for RNA secondary structure prediction. Nucleic Acids Res 2003; 31:7280-301.
28. Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. Nucleic Acids Res 2004; 32:W135-41.
29. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007; 27:91-105.
30. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res 2013; 41:W169-73.
31. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Res 2009; 37:D155-8.
32. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 2014; 42:D92-7.
33. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 2009; 37:D105-10.
34. Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 2014; 42:D78-85.
35. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res, 2014:D68-73.
36. Rennie W, Liu C, Carmack CS, Wolenc A, Kanoria S, Lu J, et al. STarMir: a web server for prediction of microRNA binding sites. Nucleic Acids Res 2014; 42:W114-8.
37. Chi SW, Hannon GJ, Darnell RB. An alternative mode of microRNA target recognition. Nat Struct Mol Biol 2012; 19:321-7.
38. Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, et al. Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. Mol Cell 2012; 48:760-70.
39. Khorshid M, Hausser J, Zavolan M, van Nimwegen E. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. Nat Methods 2013; 10:253-5.
40. Grosswendt S, Filipchyk A, Manzano M, Klironomos F, Schilling M, Herzog M, et al. Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol Cell 2014; 54:1042-54.
41. Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W, et al. Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. Cell Rep 2014; 8:1225-39.
42. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 2013; 153:654-65.
43. Zhang H, Artiles KL, Fire AZ. Functional relevance of "seed" and "non-seed" sequences in microRNA-mediated promotion of C. elegans developmental progression. Rna 2015.
44. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. eLife 2015; 4.

45. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA 2004; 10:1507-17.
46. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res, 2005:1034-50.

Sfold

*Software for Statistical Folding of Nucleic Acids and Studies of Regulatory RNAs*

HOME | LICENSE INFO | MANUAL | FAQ | CONTACT | Monday March 14, 2016

# STarMirDB

**10386 searches since April 21, 2013**

| | |
|---|---|
| Select species | Human (V-CLIP; NCBI RefSeq Build 36.3) |
| Select one or more microRNAs from the dropdown list, or enter one or more microRNA names into the text box. | hsa-let-7a-5p<br>hsa-let-7b-5p<br>hsa-let-7c<br>hsa-let-7d-5p<br>hsa-let-7e-5p<br>hsa-let-7f-5p<br>hsa-let-7g-5p<br>hsa-let-7i-5p<br>hsa-miR-98<br>hsa-miR-103a-3p<br><br>hsa-let-7c<br>let-7e<br><br>Enter one or more mirBase names separated by spaces or commas.<br>(e.g. hsa-let-7-5p, mmu-miR-299) |
| Enter the Genbank accession number(s) or gene symbols(s) of the target(s) | NM_000024<br>NM_000021<br>NM_000017<br>Lin54<br><br>Enter one or more accession numbers or gene IDs separated by spaces or commas.<br>(e.g. NM_000024, NM_000021, NM_000017, LIN54) |
| | (•) Show the most relevant features in the interactive viewer<br>( ) Show all computed features in the interactive viewer |
| Select a probability. No site with a logistic probability less than the selected value will be displayed. | 0.60 |
| | Search Reset |

# STarMirDB

**Prediction model training data (species): V-CLIP (Human)**

**Interactive Site Viewer** Display the top 100 sites

3' UTR-seed | 3' UTR-seedless | CDS-seed | CDS-seedless | 5' UTR-seed | 5' UTR-seedless

| Target | Gene | miRNA | Site_Start | Site_End | Hybrid Conformation | LogitProb | Site_Consv | $\Delta G_{hybrid}$ | $\Delta G_{nucl}$ | $\Delta G_{total}$ | CLIP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NM_194282 | LIN54 | hsa-let-7b-5p | 5119 | 5127 | view | 0.9126 | 0.7227 | -18.7 | -5.935 | -17.3 | 0 |
| NM_194282 | LIN54 | hsa-let-7b-5p | 5119 | 5140 | view | 0.879 | 0.2959 | -21.6 | -6.089 | -20.19 | 0 |
| NM_194282 | LIN54 | hsa-let-7c | 4973 | 5003 | view | 0.8733 | 1 | -17.8 | -5.183 | -13.43 | 1 |
| NM_194282 | LIN54 | hsa-let-7c | 4973 | 5005 | view | 0.8681 | 1 | -17.9 | -5.223 | -13.29 | 1 |
| NM_000021 | PSEN1 | hsa-let-7b-5p | 1793 | 1812 | view | 0.8669 | 0.5912 | -21.1 | -7.179 | -21.26 | 0 |
| NM_194282 | LIN54 | hsa-let-7c | 4973 | 5000 | view | 0.8621 | 1 | -15.9 | -4.826 | -11.53 | 1 |
| NM_000024 | ADRB2 | hsa-let-7a-5p | 1524 | 1544 | view | 0.8565 | 0.8208 | -16.8 | -4.358 | -16.44 | 0 |
| NM_000024 | ADRB2 | hsa-let-7b-5p | 1516 | 1539 | view | 0.8543 | 0.659 | -17.3 | -5 | -17.45 | 0 |
| NM_000024 | ADRB2 | hsa-let-7a-5p | 1524 | 1539 | view | 0.8528 | 0.7788 | -15.2 | -4.876 | -15.89 | 0 |
| NM_000024 | ADRB2 | hsa-let-7c | 1524 | 1544 | view | 0.8525 | 0.8208 | -16.5 | -4.173 | -16.14 | 0 |
| NM_194282 | LIN54 | hsa-let-7d-5p | 4779 | 4798 | view | 0.8524 | 0.9964 | -15.5 | -0.462 | -8.907 | 0 |
| NM_194282 | LIN54 | hsa-let-7b-5p | 4973 | 5003 | view | 0.8505 | 1 | -17.8 | -5.023 | -13.43 | 1 |

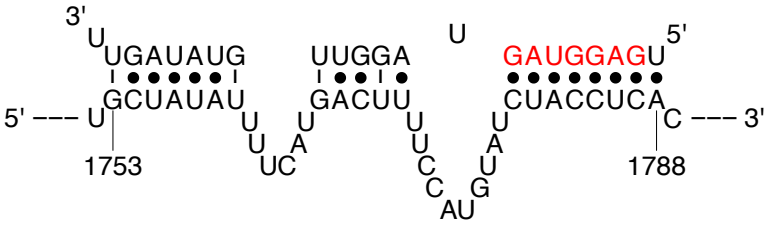
Feature definitions

**Download results from the search above**

| | |
|---|---|
| Features and predictions for 3' UTR-seed sites | Download |
| Features and predictions for 3' UTR-seedless sites | Download |
| Features and predictions for CDS-seed sites | Download |
| Features and predictions for CDS-seedless sites | Download |
| Features and predictions for 5' UTR-seed sites | No sites for this case |
| Features and predictions for 5' UTR-seedless sites | Download |

△G= −23.2 kcal/mol

hsa−let−7a−5p

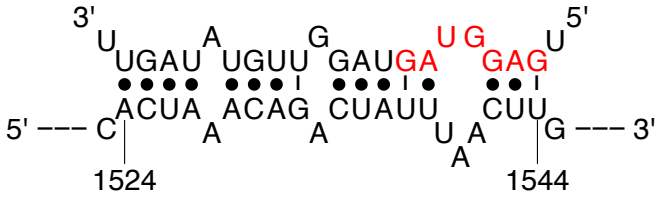



NM_000024

**Fig. 3A**

△G= −16.8 kcal/mol

hsa−let−7a−5p




NM_000024

**Fig. 3B**