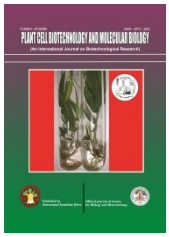

Plant Cell Biotechnology and Molecular Biology

Volume 25, Issue 9-10, Page 26-34, 2024; Article no.PCBMB.12232
ISSN: 0972-2025

Micro Target: MicroRNA Target Prediction and Validation with Experimentally Positive and Negative Examples

Shibsankar Das [a*]

[a] *Department of Mathematics, Uluberia College, Uluberia, Howrah, West Bengal, India.*

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: https://doi.org/10.56557/pcbmb/2024/v25i9-108783

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://prh.ikprress.org/review-history/12232

Original Research Article

Received: 11/05/2024
Accepted: 15/07/2024
Published: 19/07/2024

ABSTRACT

MicroRNAs (miRNAs) usually controls the gene by binding to complementary sites of 3' untranslated region of its target genes. Numerous criteria-based and machine learning approaches are available in the literature to predict miRNA–mRNA interactions, but most of them struggle with either high false positive or false negative rates and also don't show good validation with experimentally validated positive and negative examples. Here we present microTarget, a new computational approach for identifying miRNA target genes which are based on complementarity score, thermodynamic duplex stability and also independent of conservation of target sites in related genomes. In this article, we validated our algorithm using positive and negative data from the literature in various human tissues, and our method outperformed existing computational methods such as miRanda, RNA22, and PITA. Receiver operating characteristic curves (ROC) and Matthew's correlation coefficient (MCC) were calculated using experimentally validated data, and

**Corresponding author: E-mail: ssdas80@gmail.com;*

Cite as: *Das, Shibsankar. 2024. "Micro Target: MicroRNA Target Prediction and Validation With Experimentally Positive and Negative Examples". PLANT CELL BIOTECHNOLOGY AND MOLECULAR BIOLOGY 25 (9-10):26-34. https://doi.org/10.56557/pcbmb/2024/v25i9-108783.*

they reveal that microTarget greatly improves miRNA target prediction compared to the three algorithms employed individually. Additionally, an F-score analysis demonstrated that microTarget greatly enhances the relevance of the other techniques. Thus, microTarget is a useful tool for biologists looking for miRNA targets and integrating them into biological contexts.

Keywords: miRNA target interaction; target prediction algorithm; target validation; complementarity score.

1. INTRODUCTION

"MicroRNAs (miRNAs) are small noncoding RNAs that play a central role in gene regulation by base pairing to mRNAs in animal and plant genomes either by repressing mRNA translation or mediating mRNA degradation for regulating the expression of target genes in animals and plants" [1-5] (Chen, CW. et al. 2020). In recent years, more 30,000 mature miRNAs have been discovered in all over the species; however, only a limited number of functions of these miRNAs are identified. Experimental and computational evidence suggests that they are involved in various diseases like cancer [6-9]. Since the interaction between miRNA and mRNA is very complex, so the recognizing miRNA and mRNA interaction is the current interest among scientists who want to pursue research to study the functional behavior of miRNAs. However, because of the laborious, costly, and time-consuming experimental methods used to predict miRNA targets, scientists are now looking beyond direct experimentation to predict miRNA-mRNA interactions. This is because computational approaches can offer more useful and efficient target prediction methods.

Several miRNA-target prediction algorithms are available in the literature based on the different procedures and measures including base pairing, target accessibility, contrary to the position and localization of Watson-Crick pairings and mismatches and evolutionary conservation of target site like miRanda [10] PITA [11], RNA22 [12], miRgo [5] miRabel [14] etc. However, these methods have high sensitivity and specificity and gives low MCC and F score calculated in experimentally validated positive and negative examples. Brennecke et al. [13] showed experimentally in D. melanogaster and D. pseudoobscura that strong complementary in 5' end of miRNA requires conferring regulation and sites with weaker 5' complementarity require compensatory pairing to the 3'end of the miRNA in order to function. They also displayed the experimental results that miRNA having more than one G=U base-pair or bulge or mismatch in seed region might become totally". It was established by Wang Xiaowei [15] that 6-mer or 7-mermiRNAwith perfect or one G=U base pairing in between positions 2-10 counted from miRNA 5' end, showed decent enrichment ratio in CLASH data. Grimson *et al*. [16] revealed that "additional Watson-Crick Pairing in between positions 12–17 improves miRNA Targeting efficiency". Thus we have incorporated the results of Brennecke *et al.* [13] Wang Xiaowei [15] and Grimson *et al*. [16] in our proposed miRNA-target prediction algorithm (detailed described in Materials and Methods section) to improve false positives and shown that it performs good MCC and F score calculated in experimentally validated positive and negative examples.

In this study, we have proposed a new computational method to detect miRNA targets. We first validated the results with experimentally validated positive and negative examples and compare validation results with miRanda, PITA, and RNA22. We have performed various measures like ROC curve, AUC, MCC score and F-measure of our algorithm and compared results with other popular miRNA target prediction algorithms i.e. miRanda, PITA, and RNA22.

2. MATERIALS AND METHODS

2.1 miRNA Sequences

Experimentally validated miRNA sequences of *D. melanogaster* (Fruit Fly) were downloaded from MirTarBase database [17] (mirtarbase.mbc.nctu.edu.tw). Also, all miRNA sequences of human genome used in positive and negative examples have collected from Metabases database [17].

2.2 3' UTR Sequences

All 3' UTR sequences of target genes of *D. melanogaster* were collected from UTRdb database [18]. Also, all 3' UTR sequences of

target genes sequences used in positive and negative examples of the human genome have collected from UTRdb database [18] (utrdb.ba.itb.cnr.it/).

2.3 MicroTarget Algorithm

microTarget algorithm [19] is similar to miRanda algorithm [10], however instead of using empirical rules (flowchart shown in Fig. 1). “It uses similar complementarity parameters as miRanda algorithm at individual alignment positions: +5 for

G≡C, +5 for A=U, +2 for G=U and -3 for all other nucleotide pairs. The algorithm uses affine penalties for gap-opening (-8) and gap-extension (-1). Also, complementarity scores (positive and negative values) at the first eleven positions are multiplied by a scaling factor (here set at 2.0)”. [19]. The following five rules are applied to positions counted starting at the 5' end of the miRNA:

(1) There must be 6 to 8 base pairs between positions 1 to 10.
(2) Seed region with 8 base pairs and starting from position 1, may have up to two G=U base- pairs or one bulge (either of the miRNA or of the 3’UTR) or single non-G=U mismatch in between seed region (i.e. from positions 2-7).
(3) Seed region with 7 base pairs and starting from positions 1-4, may have one G=U base-pair or one bulge (either of the miRNA or of the 3’UTR) or single non-G=U mismatch in between seed region.
(4) (4) Seed region with 6 base pairs and starting from positions 2-5, may have only one G=U base- pair in between seed region.
(5) If G=U base pair or bulge or mismatch are used in seed region and starting from positions either 3-4 or 4-5, there must be at least 4 base pairs (including G=U base-pairs) from positions 12 to 3’ UTR end of miRNA.

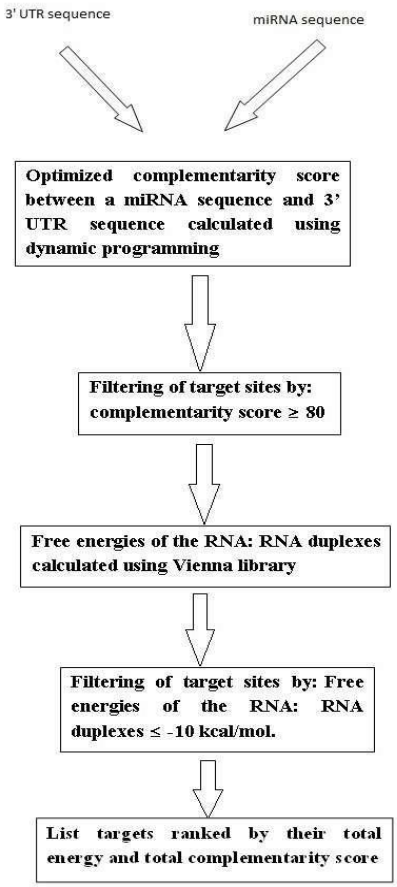

Fig. 1. Flow chart for the systematic prediction of miRNA–target duplex by microTarget.

"Using these parameters and rules, complementarity score between a miRNA sequence and 3' UTR sequence is optimized using dynamic programming and summed over all aligned positions" [19]. This miRNA and 3' UTR interaction will be predicted as a possible target if its complementarity score is greater than 80. The default cut-off value for complementarity score is taken 80 as we observed that complementarity score calculated by our method in all experimentally validated miRNA-3' UTR examples (collected from MirTarBase database [17] in D. melanogaster (Fruit Fly), Danio rerio (Zebrafish), Gallus gallus (Chicken) and one plant, Arabidopsis thaliana (Thale cress) are greater than their cut-off value and complementarity score between all randomized experimentally validated miRNA and 3' UTR pairs are less than their cut-off values. "All non-overlapping hybridization alignments in decreasing order of complementarity score are found. In order to calculate free energies of the RNA: RNA duplexes, we use folding routines from the Vienna RNA secondary structure programming library (RNAlib)" [20]. "The thresholds used for the possible target are complementarity score ≥ 80 and energy of the duplex structure ≤ -10 kcal/mol. Each possible target site between a miRNA and a UTR sequence is then scored according to the total energy and total score of all possible targets sites between those two sequences. The top ten ranked genes are selected as its candidate target genes for each miRNA. A target gene binding by multiple miRNAs is selected by the miRNAs that assign the highest scoring and lowest free energy of the miRNA target duplex to each potential site so that different miRNA target sites cannot overlap" [19].

3. RESULTS AND DISCUSSION

3.1 Experimentally Verified Positive and Negative Examples for Validation of the Proposed Method

In this paper, we have collected a set of 190 negative miRNA-target interactions (set of negative examples are provided in the Supplementary Table ST4 as used in [21] to access the prediction performance of our proposed method. The negative mina-Target interactions are authenticated experimentally with the dataset investigated in Selbach *et al*. [22] using pulse-labeling stable isotope labeling with amino acids in cell culture (pSILAC) technology and witnessed that the gene expression levels of the target set moved toward more negative set than compared with the non-target set. It is also observed that all the mRNAs of true target set showed a higher negative log2 fold change of (greater than −0.5) and mRNAs of non- target set showed relatively higher fold changes. Those miRNA-mRNA pairs both of which are overexpressed or under-expressed in the same tissue are extracted as potential negative examples as these examples do not support the biology of miRNA-mediated target repression event.

In this paper, we have collected a set of 187 miRNA transcript pairs (positive examples) extracted from them Records database [23], and set of positive examples are provided in the Supplementary Table ST1. These examples are validated by the two experiments proposed by Lim *et al*. [24] and Wang and Wang [15].

3.2 Features Selection to Compare the Performance of microTarget at the Target Level

In this paper, we have used three popular target prediction methods, namely miRanda, Pita and RNA22 (their software are publicly available) in addition to our proposed algorithm microTarget. The latest versions of miRanda (microrna.org; [10] PITA [11] and RNA22 [12] executables were downloaded and executed with its default parameters as described by the package.

We have chosen four features- frequency of A's in seed region, number of mismatches in seed region, number of GT matches in total region and free energy in seed region as these features are common in all three target prediction algorithms and assessed the marginal distribution of features in the form of histogram in both positive and negative sets. Although the marginal distribution of the features will not show any combinatory discriminative importance, they will reveal the discriminative power of each individual feature. Fig. 2 shows the histogram of four selected features. Histograms are drawn by taking *x*-axes as the feature values and the *y*-axes denote the relative frequency. It is clear from the Fig. 2 that A's in seed region, number of mismatches in seed region, number of GT matches in the total region have good discriminative power and free energy in seed region do not perform relatively good feature for target prediction [25].

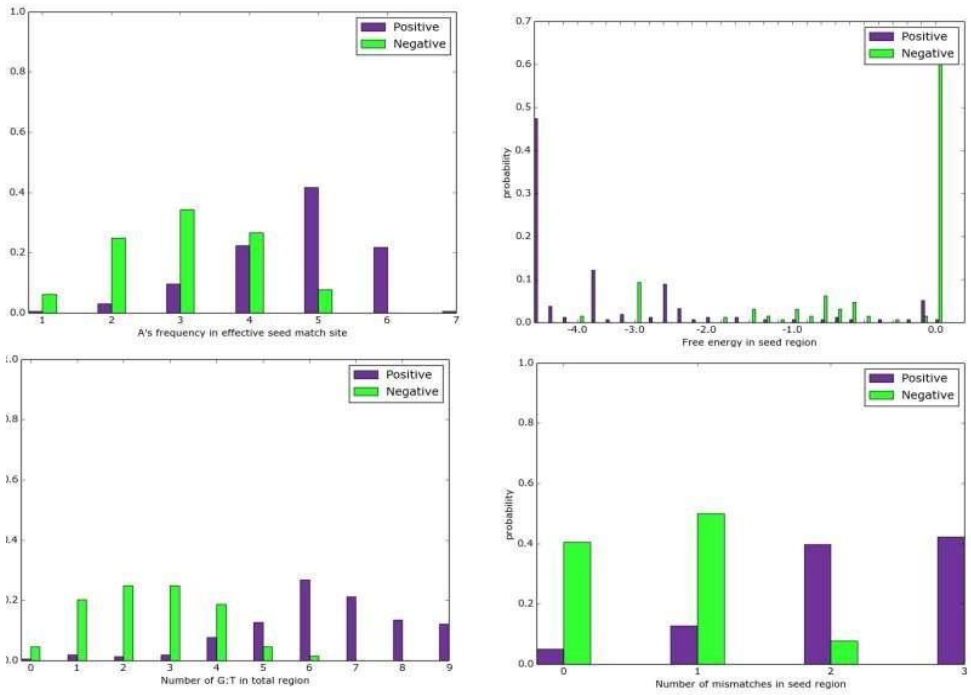

Fig. 2. Histogram of four different features

3.3 Comparison of Prediction Methods

The predictions made by all algorithms for the set of 187 positive examples are shown in Supplementary Tables ST2 and ST3. From this table, we observed that microTarget predicted 155 positive examples out of 187 positive examples, whereas RNA22, miRanda, and PITA made 57, 125 and 168 positive examples. It was additionally observed that, among of all the methods used for this article, only our method was able to predict two positive examples as targets which were not detected by other methods. Out of 190 negative examples (shown in Supplementary Tables ST5 and ST6), microTarget predicted only 64 negative examples as targets, whereas PITA predicted highest (123) negative examples as targets. RNA22 and miRanda made 79 and 41 respectively negative examples as targets. In addition, we observed that none of the other methods used in this paper detected any negative targets predicted only by our methods. These results showed that PITA predicted the highest number of positive and negative targets, indicating that it had a high false positive rate due to bias in its prediction of all input instances as positive examples. Both miRanda and RNA22 predicted fewer positive and negative targets, indicating a high likelihood of false negatives. On the other hand, our approach predicted low negative and high positive targets, which constitute essential aspects of an effective prediction method.

3.4 Comparative Performance of microTarget at the Binding Site Level

We have evaluated the performance of microTarget with other algorithms in terms of sensitivity, $Sn = \left(\frac{\mathrm{TP}}{\mathrm{TP+FN}} \right)$, specificity $Sp = \left(\frac{\mathrm{TN}}{\mathrm{TN+FP}} \right)$, Matthew's correlation coefficient (MCC)

$$MCC = \left(\frac{\mathrm{TP \times TN - FP \times FN}}{\sqrt{(\mathrm{TP+FP}) \times (\mathrm{TN+FN}) \times (\mathrm{TP+FN}) \times (\mathrm{TN+FP})}} \right) \text{ and } F - measure = \left(\frac{2\mathrm{TP}}{2\mathrm{TP+FP+FN}} \right)$$

where TP = true positive, TN= true negative, FN = false negative and FP = false positive.

Table 1 shows the sensitivity and specificity of different target prediction algorithms. It shows that PITA has high sensitivity (*Sn*=0.903226) but low specificity (*Sp*=0.352632), whereas RNA22 has high specificity (*Sp*=0.784211) but low sensitivity (*Sn*=0.30645). miRanda has specificity (*Sp*=0.584211) and sensitivity (*Sn*=0.67204), whereas our proposed algorithm shows specificity (*Sp*=0.663158) and sensitivity (*Sn*=0.82888). Thus, our proposed algorithm shows good sensitivity and specificity so it could be considered as a better prediction method in consideration of both sensitivity and specificity.

The Matthews correlation coefficient (MCC) is a statistical method that is considered more reliable. It yields a high score only when the prediction accurately reflects the size of both positive and negative elements in the dataset, and only in each of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives). Table 1 also displays that RNA22 achieves lowest MCC score (0.10). MCC scores of PITA and miRanda are 0.31 and 0.26 respectively whereas microTarget achieves highest MCC score (0.51). PITA, miRanda, and RNA22 have low MCC scores, which suggests that they are biased in predicting all input data as positive examples, leading to a large false positive rate.

3.5 Evaluation of Prediction Methods

The performances of each prediction algorithm were also compared to microTarget using ROC analysis and F score to see if any improvement was obtained with our prediction method. Fig. 3 shows the ROC curves of the algorithms used in this paper. It is clear that AUC (area under the curve) of microTarget is highest (0.90) whereas AUC of miRanda, Pita and RNA22 are 0.82, 0.77 and 0.74 respectively. It can be easily verified that microTarget shows better performance than all the three algorithms. Specifically, the microTarget shows the lowest false positive rate (FPR) compared to the other methods at a constant true positive rate (TPR), as well as a higher TPR than all three methods' TPRs at a constant FPR. Fig. 4 shows the F-measures of all four techniques. This discovery indicates that PITA and microTarget excel, while miRanda and RNA22 seem to underperform. MicroTarget outperforms the other options by a large margin, with a success rate of 77%.

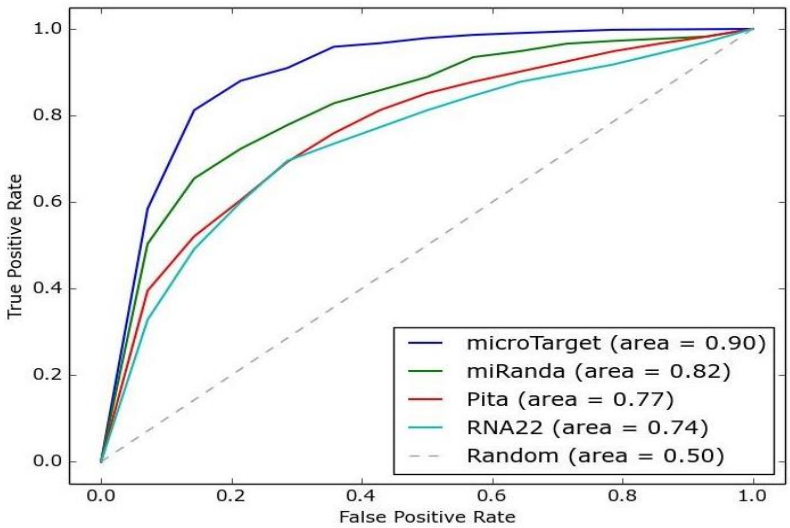

Fig. 3. The ROC curve of different algorithms (AUC shown in the bracket)

Table 1. Sensitivity, specificity and MCC obtained by different target prediction algorithms

	miRanda	Rna22	PITA	microTarget
Sn	0.67204	0.30645	0.90323	0.82888
Sp	0.584211	0.784211	0.352632	0.663158
MCC	0.26	0.10	0.31	0.51

Table 2. of miRNA-target interactions predicted by all four methods in *D. melanogaster*

	Total Hits	Single Hit	No.of Hits 2	Average number of targets pr miRNA (TPM)
RNA22	998419	330552	667867	4267
miRanda	346465	262495	83970	1481
PITA	843274	480079	363195	3604
microTarget	330358	263879	66479	1412

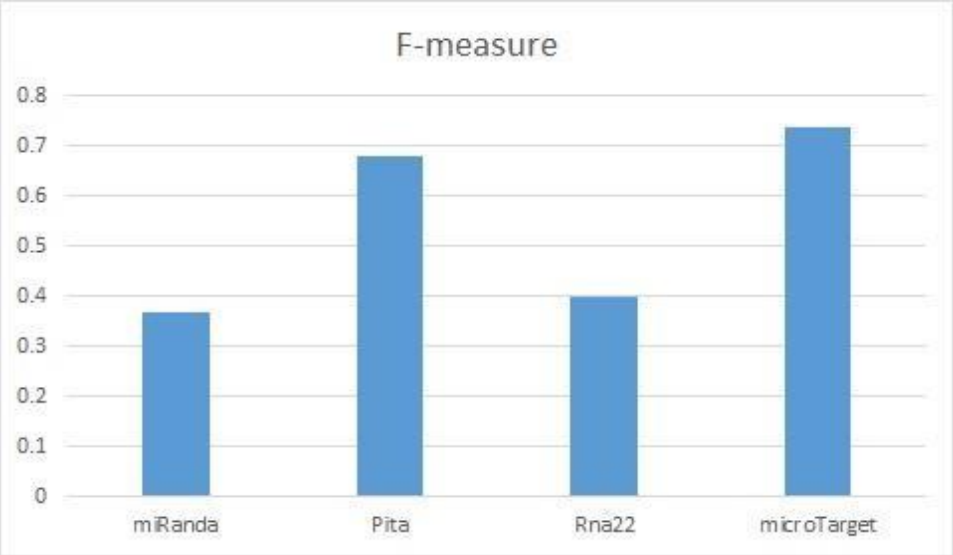

Fig. 4. F-measure of all algorithms

3.6 Prediction of miRNA-target Interactions in *D. melanogaster*

We used all four approaches (i.e. microTarget, miRanda, PITA and RNA22) to predict miRNA-target interactions in *D. melanogaster.* We used 234 miRNA genes with 19001 3'-UTR of target gene sequences. Out of 19001, 11379 3'-UTR of target gene sequences are unique and remaining have multiple copies. Executable of miRanda and PITA were used to find the miRNA target interactions whereas results of miRNA-target genes by RNA22 were obtained from theRNA22dataset (https://cm.jefferson.edu/data-tools-downloads/rna22-full-sets-of-predictions/).
Table 2 shows the number of miRNA-target interactions of 234 miRNAs and 11379 3'-UTR in *D. melanogaster.* It has been observed that the average number of targets predicted per miRNA (TPM) for microTarget is lowest (1412) as compared with several other methods. RNA22 and PITA showed the high average number of targets predicted per miRNA (TPM) as those methods are biased to predict all the miRNA-target interactions as valid miRNA-target interactions due to their sensitivity and specificity [21].

4. CONCLUSION

In this article, we have proposed a new miRNA-target prediction algorithm microTarget which is sequence-based method rather using orthogonal non-sequence based data. Current popular computational approaches for target predictions either use evolutionary conservation which is not always possible to compare or non-sequence based data (like feature based multiple instance Learning (MIL) methods), but our proposed method microTarget does use neither evolutionary conservation nor feature based MIL model. So it can be used any genome even if evolutionary conservation of this genome is not known. We have collected 187 negative miRNA-target interactions from (Bandyopadhyay and Mitra 2009) and 187 miRNA transcript pairs (positive examples) extracted from them Records database (Xiao *et al*. 2009) to validate and compare the performance of microTarget with of miRanda, Pita and RNA22. F-measure, MCC score and ROC curve are calculated and results showed that microTarget comprehensively outperforms than other miRNA-target prediction methods with a high margin. microTarget were applied in the

genome *D. melanogaster* to predict miRNA-target interactions and it is shown that an average number of targets predicted per miRNA (TPM) for microTarget is lowest than other methods.

Based on results and performance of microTarget, it can be concluded that microTarget will make a valuable impact on future laboratory experiments for finding out miRNA-target interactions. Although our proposed miRNA target algorithm integrates many important features to predict microRNA-target interactions, due to biological complexity of microRNA-target interactions, all possible characteristics of miRNA-mRNA interactions are not included in our methods as these features are still unknown. So new experimental results from high-throughput miRNA-mRNA interactions could improve the success efficiency of computational approaches of target prediction and help to untangle the biology of regulation by miRNA-mRNA interaction.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

SUPPLEMENTARY MATERIALS

Supplementary material is available in the following link:

https://ikprress.org/media/documents/Supplementary_2024_PCBMB_12232.pdf

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Pasquinelli AE. MicroRNAs: Deviants no longer. Trends Genet. 2002;18:171-173.
2. Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing Cell. 2003;113 673-676.
3. Carrington JC and Ambros V. 2003 Role of microRNAs in plant and animal development. Science. 2003;301:336 -338.
4. Lin SY, Johnson SM, Abraham M, Vella MC, Pasquinelli A, Gamberi C, Gottlieb E and Slack FJ. The Celeganshunchback Homolog hbl-1 controls temporal patterning and is a probable microRNA target. Dev. Cell 2003;4:639-650.
5. Chu YW, Chang KP, Chen CW. miRgo: Integrating various off-the-shelf tools for identification of microRNA–target interactions by heterogeneous features and a novel evaluation indicator. Sci Rep. 2000;10:1466.
6. Ridley M Genome: The autobiography of a species in 23 chapters, 1st ed.; Topeka bindery: New York; 2006.
7. Bartel D. MicroRNAs: genomics, biogenesis, mechanism, and function. 2004;116 281–97.
8. Zhao T, Li G, Mi S, Hannon GJ, Wang XJ and Qi Y. A complex system of small RNAs in the unicellular green alga Chlamydomonasreinhardtii. Genes Dev 2007;21(10):1190–2003.
9. He L, Hannon G. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet. 2004;5(7):522–531.
10. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2004;5 (1):R1-14.
11. Kertesz M, Iovino N, Unnerstall U, Gaul U and Segal E. The role of site accessibility in microRNA target recognition. Nat. Genet. 2007;39:1278–1284.
12. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim BR. A tern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. 2006;126(6):1203-1217.
13. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA–target recognition. PLoS Biol. 2005;3(3):e85.
14. Quillet A, Saad C, Ferry G, Anouar Y, Vergne N, Lecroq T and Dubessy C. Improving bioinformatics prediction of microRNA targets by ranks aggregation. Front. Genet. 2020;10:1330.
15. Wang X, Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling. Nucleic Acids Res. 2006;34:1646–1652.
16. 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(1):91-105.

17. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, Tsai TR, Ho SY, Jian TY, Wu HY, Chen PR, Lin NC, Huang HT, Yang TL, Pai CY, Tai CS, Chen WL, Huang CY, Liu CC, Weng SL, Liao KW, Hsu WL and Huang HD. miRTarBase: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 2016;44(D1):239-47.
18. Giorgio G, Antonio T, Flavio L, Flavio M, Sabino L, Sandro B, Vincenzo AG, David SH, Giulio P, Ernesto P and Graziano P. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res. 2010;38(Database issue):D75–D80.
19. Das S, Mandal D. MicroRNA target gene prediction and validation in different Genomes. Rsearch Journal of Life Science, Bioinformatics, Pharmaceutical and Chemical Science. 2016;2(2):10 – 28.
20. Wuchty S, Fontana W, Hofacker IL, Schuster P. Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers. 1999;49:145-165.
21. Bandyopadhyay S, Mitra R. TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. Bioinformatics. 2009;25:2625–2631.
22. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008;455: 58–63.
23. 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–D110.
24. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005;433:769–773.
25. Xiaowei Wang. Composition of seed sequence is a major determinant of microRNA targeting patterns. Bioinformatics. 2014;30 (10):1377–1383.

Peer-review history:
The peer review history for this paper can be accessed here:
https://prh.ikprress.org/review-history/12232