Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI - 10.26479/2017.0304.08**

THE COMPUTATIONAL METHODS FOR BK POLYOMAVIRUS MICRORNA TARGET IDENTIFICATIONS ON A HUMAN GENE

Sasti Gopal Das *, Abhijit Datta

Department of Botany, Jhargram Raj College, Paschim Medinipur, Jhargram 721507, West Bengal, India

ABSTRACT: MicroRNAs (miRNAs) are small non-coding RNA molecules that serve as master regulators of cellular functions. In plants, miRNA target recognition requires extensive sequence complementarities, while in animals it requires sequence complementarities at the 5' end of the miRNAs. Many computational methods have been used for finding miRNA target interactions, but precise miRNA target gene identification is a challenging task. In this study, we employed structural properties for training a support vector machine (SVM) classifier, which was used for filtering the miRNAs targets. We achieved a sensitivity of 83.62%, a specificity of 87.92 % and an accuracy of 85.86%. This method identified new BK polyomavirus miRNA on the human gene as an efficient and accurate identification of potential miRNA targets.

KEYWORDS: MicroRNA; target prediction; seed match; Support Vector Machines (SVM)

***Corresponding Author**: Dr.Sasti Gopal Das Ph.D

Department of Botany, Jhargram Raj College, Paschim Medinipur, Jhargram 721507, West Bengal, India. * Email Address: shasti_gopaldas05@yahoo.co.in

1.INTRODUCTION

MicroRNAs (miRNAs) are small non-coding (20-24 nucleotide long) [31] RNA molecules that serve as post-transcriptional regulators of plant and animal gene expression. miRNA target recognition in plants requires extensive sequence complementarities whereas, in animals, miRNAs show partial complementarities to their targets for recognition [1], [4]. miRNAs are involved in the regulation of an array of gene expressions such as embryonic development, cellular functions and biological expressions [2], [3]. Target recognition occurs via ~7nt consecutive Watson-Crick base pairs on CDS [37] and the 3' Untranslated Regions (UTRs) [32] of mRNAs. The miRNA-target paired ~7nt bases at the 5' ends of miRNAs, is called the seed region, important for target recognition [5], [6], [8], [9],

[10]. Recent work in the identification of miRNA targets has been done with the help of bioinformatics and molecular techniques. Identification of potential miRNA target genes is a great challenging task and can be predicted using bioinformatics approaches and Machine Learning techniques [3], [7], [10], [11], [12], [13], [14]. Several computational methods already predict miRNA targets, but the complexity of miRNA targeting deprives the standardization of specific thumb rules, aggravating the problem of false positive results. Many authors support the use of Machine Learning methods for improving target prediction by training with biologically relevant data [11], [12], [13], [14], [15], [16]. The publicly available miRNA-target databases - miRecords, Tarbase and mirtarbase contain experimentally validated miRNA-target interactions [24], [25], [26]. Information from such databases can help to train Support Vector Machines (SVM) and predict miRNA targets, based on experimentally validated miRNA-targets [11], [12], [13], [14], [15], [16]. Experimental identification of miRNA targets is a time-consuming and labor-intensive process, making the development of computational approaches for predicting potential miRNA-targets, a necessary one. Computational methods have been developed and widely used in miRNA target prediction such as HomoTarget, miTarget, MiRTif, MTar, TargetMiner, MultiMiTar, miRanda, TargetScan and RNAhybrid, where the predictions are performed considering miRNA-target sequence complementarities and thermodynamics parameters [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Several experimental studies support that the first 2-7 nucleotides at the 5' end of the miRNA are particularly important for target recognition [5], [6], [8], [9], [10], [19], [20]. The miRNA-target interaction occurs via near-perfect complementarities of base-pairing in the seed region of the miRNA-target.

In this study, we employed a dynamic programming algorithm for searching the miRNA-target duplexes, which were subsequently classified as true or false targets using Support Vector Machines. In the machine learning approach using SVM, experimentally validated miRNA-target interactions were used for training the SVM model [22], [23], [33], [34]. The miRTarBase (2016,) database provides information on 366,181 experimentally validated miRNA-target interactions. This database contains pre-classified Functional MTIs (Functional miRNA-target interactions) and NON-Functional MTIs (NON-Functional miRNA-target interactions) [26]. Our method detects functional miRNA target sites in the CDS genome sequence based on the Smith-Waterman dynamic programming algorithm and subsequent post processing filtering of the predicted targets using support vector machine (SVMlight) [33]. The SVM classifier was previously trained with the experimentally validated miRNA-target interactions found in miRTarBase. For model training, each experimentally validated miRNA-mRNA target interaction from miRTarBase was mapped onto a Support Vector Machine (SVM) model, and the trained model was used to classify the predicted targets. Support Vector Machines (SVMs) were first introduced by Vapnik and co-workers [34]. The trained model was used to classify unlabeled inputs into two classes: positive target and negative target. Our model achieved a sensitivity of 83.62, specificity 87.92, and an accuracy of 85.86.

2. MATERIALS AND METHODS

Dataset

Experimentally verified miRNA-mRNA target datasets were retrieved from the miRTarBase. 464 Positive datasets (Functional miRNA-target interactions, MTIs) carrying strong evidence of interaction using reporter assay and Western blots, while the 505 NON-Functional miRNA-target interactions (NON-MTIs) were considered as the negative dataset. miRNA-target hybrid duplexes of *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio,* and *Caenorhabditis elegans*, were retrieved and duplicate entries removed [26], [27], [28].

Motif search for miRNA target sites recognition

At first, the miRNA sequence stretches between positions 2-7nt at the 5' end of the miRNAs were converted to their complementary sequences. For motif recognition (7mer) [5], [6], [7], [8], [9], [10], (Supplementary 1) complementary RNA sequences were searched against target sites in the mRNA sequences retrieved from the NCBI FTP human genome database. Once the motif position was determined, the entire sequences of the miRNAs were aligned against the target mRNA sequences using the Smith-Waterman dynamic programming algorithm [22], [23].

Dynamic Programming Algorithm scoring for local alignment

The dynamic programming alignment scoring was based on sequence complementarities and not on sequence identities. Wobble base pairs were also taken into consideration. Sequence alignment scores were assigned as: Watson Crick base pair(A:U =5, G:C=5), mismatch= (A:A=-3, G:G=-3, C:C=-3, U:U=-3, A:G=-3, U:C=-3, A:C=-3), gap-opening=-8 (A- ,U- ,G- ,C-) gap-extension=-2 and wobble pair G:U =2 (19).

Feature selection

The 464 experimentally verified positive (MITs) and 505 negatives (NON-MITs) miRNA-target hybrids were analyzed by Fisher score and the most significant 71 SVM features were determined. Identification of SVM features was based on Position specific structural feature, entire structural feature and consecutive Watson-Crick base pairing patterns [12], [13], [15].

Position specific structural feature

1st type: In the position specific scoring scheme, we focused on 14 type of base pairings in a miRNA-target duplex: Watson-Crick base pair (A: U, G: C), wobble base pair (G: U), gap (A-, U-, G-, C-) and mismatch (AA, UU, GG, CC, AG, CU, AC). For each miRNA-target interaction, we calculated the F-score from the Position Specific Scoring Scheme (p1.p2p3…..p22) and selected the 39 most important features out of 308 SVM features (Table 1 and Supplementary 3).

Table 1

Position Specific	Selected Features	Total Features	Selected Features
1	AU, GC, GU, UU, CU, U-, G-, C-, A-.	14	9
2	AU, GC, GU, UU, AG, U-.	14	6
3	GC, GU, UU, GG, AC, CU.	14	6
4	GC, GU, CC, AG, AC, CU.	14	6
5	AU, GU, AA, AG, AC.	14	5
6	AU, GU.	14	2
7	GU, UU.	14	2
8	GU,	14	1
10	U-	14	1
15	GC.	14	1
	(14 Types of Base Pairing) x (22 Positions)= Total Features	308	39

[Table 1: Base Pairing types = (AU, GC, GU, AA, UU, GG, CC, AG, CU, AC, A-, U-, G, C-).]

2nd type: 3 types of pairing were defined in a miRNA-target duplex: Watson-crick base pair (WC = AU, GC), the gap (A-, U-, G-, C-) and mismatch (M = AA, UU, GG, CC, AG, CU, AC). For each miRNA-target interaction, we determined the F-score from the computed Position Specific Scoring Scheme (p1.p2p3…..p22) and selected the 16 most important features out of the 66 SVM features (Table 2 and Supplementary 4).

Table 2

Position Specific	Selected Features	Total Features	Selected Features
1	WC, M, Gap	3	3
2	WC, M.	3	2
3	WC, M	3	2
4	WC, M,	3	2
5	WC, M,	3	2
6	WC, M,	3	2
7	M,	3	2
8	Gap	3	1
	(3 Type of Base Pairing) x (22 Position)= Total Features	66	16

[Table 2: WC= (AU,GC); Mismatch=M (AA,UU,GG,CC,AG,CU,AC), Gap=(A-,U-,G-,C-).]

Entire structural 14 base pairing

In the 1st type structure based feature selection, we defined 14 types of base pairing in a duplex: Watson-Crick base pair (A: U, G: C), wobble base pair (G: U), gaps (A-, U-, G-, C-) and mismatches (AA, UU, GG, CC, AG, CU, AC). For each miRNA: target interaction, we computed the entire length of miRNA duplex target *F*-score and found the 4 most important features of the 14 SVM features (Table 3,Supplementary 5).

Table 3

Entire length	Selected Features	Total Features	Selected Features
1-22	AU, UU, GG, GU	14	4
	(14 base pairing type) x (1-22 total Alignment)= Total Features	14	4

[Table 3: Base Pairing type= (AU, GC, GU, AA, UU, GG, CC, AG, CU, AC, A-, U-, G, C-).]

In the 2nd type, we defined 3 types of pairing in a duplex: Watson-Crick base (WC: AU, GC), gaps (A-, U-, G-, C-) and mismatches (AA, UU, GG, CC, AG, CU, AC). For each miRNA: target interaction, we computed the F-score from the entire length of the miRNA-target duplex and identified the 1 most significant feature out of 3 SVM features (Table 4, Supplementary 6).

Table 4

Entire length	Selected Features	Total Features	Selected Features
1-22	WC.	3	1
	(3 base pairing type) x (1-22 total Alignment)= Total Features	3	1

[Table 4: WC= (AU,GC), Mismatch=M (AA,UU,GG,CC,AG,CU,AC) and Gap=(A-,U-,G-,C-).]

Consecutive Watson-Crick base pairing patterns

We employed a novel approach for finding SVM features, in which we identified (6, 7 or 8) consecutive Watson-Crick base pairing pattern in the seed region and 4, 5, 6, 7, 8 consecutive Watson-Crick base pairing patterns in the non-seed region, and selected the 11 most important SVM features (Table 5, Figure 1 and supplementary-2).

Table 5

<table>
<tr><td rowspan="2">Number of consecutive Watson-Crick base pairs (the number of base pairs indicated by | vertical bars)</td><td colspan="4">F scores (F)</td></tr>
<tr><td colspan="3">Seed region(Positions 0-8)</td><td>Non-Seed region (Positions 8-22)</td></tr>
<tr><td>4 (||||)</td><td colspan="3">None found</td><td>F= 0.0946</td></tr>
<tr><td>5 (|||||)</td><td colspan="3">None found</td><td>F= 0.1110</td></tr>
<tr><td>6 (||||||)</td><td>F= 0.3372 (position 1-6)</td><td>F= 0.5637 (position 2 -7)</td><td>F= 0.4887 (position 3-8)</td><td>F= 0.1060</td></tr>
<tr><td>7 (|||||||)</td><td colspan="2">F= 0.3455 (position 1-7)</td><td colspan="2">F= 0.5452 (position 2-8)</td><td>F= 0.1076</td></tr>
<tr><td>8 (||||||||)</td><td colspan="3">F= 0.3009 (position 1-8)</td><td>F= 0.1157</td></tr>
</table>

[Table 5: WC= (AU,GC), Mismatch=M (AA,UU,GG,CC,AG,CU,AC) and Gap=(A-,U-,G-,C-).]

SVM models training

For the proper training of SVM models, selection of several hyper-parameters is an essential prerequisite; their values determine the function of SVM model and it has a significant impact on the performance of the trained SVM model classifiers. In this work, the optimal threshold value selected based on 10-fold cross-validation on the entire dataset. The widely used radial basis function (RBF) kernel was chosen for the training of the model [12], [13], [15].

Pseudo-code of our algorithm

Our algorithm

Motif search 6-8 seed (miRNA sequence, mRNA sequence);

{

Alignments= Dynamic Programming local alignment (miRNA, mRNA);

svm data =(ExtractFeatures);

Inputs=svm Classifier (svm data);

Outputs=SVM_clasiffy(inputs);

Return outputs;

}

3.RESULT AND DISCUSSION

The performance of the final SVM classification is measured by the quantity of true positives (*TP*), true negatives (*TN*), false positives (*FP*), false negatives (*FN*), sensitivity (*SE*), specificity (*SP*) and overall accuracy (*Q) [13],[15],[16]*. Using currently known experimentally verified miRNA-mRNA target duplexes, 464 Positive datasets (Functional miRNA-target interactions, MTIs) and 505 NON-

Functional miRNA-target interactions (NON-MTIs), the SVM classifier was tested.

Overall prediction accuracy

The sensitivity, specificity, overall accuracy and AUC (area under the curve) of the 10-fold cross-validation are 83.62%, 87.92%, 85.86% and 91.69% respectively (Figure 1). The prediction methods usually produce hundreds of targets for a given miRNA, meaning a large number of them could be false positives. Using the known miRNA-target interaction pairs, our trained SVM classifier successfully predicted most of the false positives: 444 were correctly predicted out of 505 from negative samples (NON-MTIs) with known false positive interactions. Of the 464 positive samples with known experimental evidence, our model could predict 388 correctly.

TP=388, FP =76, TN=444, FN=61, Sensitivity = (83.62), Specificity = (87.92), Accuracy = (85.86) AUC=0.9169.

Sensitivity (SE) =TP/TP+FN.

Specificity (SP) =TN/TN+FP.

Accuracy (Q) =TP+TN/TP+TN+FP+FN.

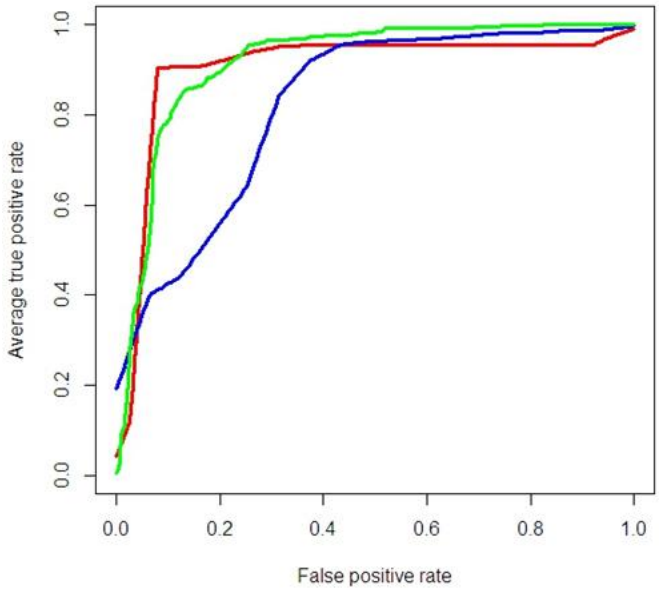


Figure 1: The ROC curve denotes the performance of TARFi and with other the tools; TARFi (Green line), MiRTif (Red line) and TargetMiner (Blue line). The Y-axis represents the true positive rate (sensitivity) and X-axis shows the false positive rate (1—specificity), (ROC)is created by R Programming 'ROCR' package.

Figure 1 shows the plot of the true positive rate versus the false positive rate on the completely independent test data set. The TargetMiner independent test data set was retrieved from TargetMiner reference miRNA::mRNA duplexes and MiRTif independent test data set was retrieved from MiRTif (supplementary data set). These data sets were submitted to our algorithm. The plot compares the balance between sensitivity and specificity of the proposed method, with other existing methods. To evaluate the prediction Performance of this tool with other method is our algorithm (AUC=0.9169),.

MiRTif(AUC= 0.9028) and TargetMiner(AUC=0.8229).The true positive rate of independent test data set of TargetScan, RNAhybrid, and miRanda. The compares the sensitivity of the proposed method of our method with other existing independent test data sets - TargetScan, RNAhybrid, and miRanda. To evaluate the prediction sensitivity of method, it was tested with TargetScan, RNAhybrid and miRanda datasets and sensitivity values of 0.8686, 0.6569 and 0.7980 were respectively achieved [17], [18], [19], [20], [21], [29], [30].

4. CONCLUSION

We propose a novel computational method for miRNA target prediction from CDS sequences, which can identify all known miRNA targets. Scanning the motif region in CDS sequences, first ~7 nt at the 5' end of the miRNA is particularly important for target site recognition. The dynamic programming alignment aligns the two sequences and the resulting miRNA::mRNA duplex is further classified (with SVMlight) for a post-processing filtering of the targets [12, 13] to reduce false positive miRNA target prediction. The sensitivity, specificity, overall accuracy and AUC indicate that the overall performance of the algorithm is good for miRNA target prediction. The results of the ROC curve indicate the true positive rate of this tool. In this work, we have employed an algorithm for the prediction of miRNA targets. The novelty of our approach lies in the use of an initial motif search (7mer) (supplementary-2, table 5), which reduces computational time. The algorithm uses a three-step strategy for finding potential miRNA targets in genomic sequences based on motif search, Dynamic Programming algorithm, and a trained SVM classifier. Our algorithm compares favorably to other algorithms, both in terms of overall performance and when making highly specific predictions. We believe that our algorithm will be an important algorithm for finding the target sites of known miRNAs. We performed the same feature filtering and SVM parameter optimization steps. 71 feature parameters were used for training the SVM classifier (miRNA::mRNA) complementarities [13], [15], [16], [35], [36]. BK polyomavirus Viral miRNAs can target to the host cell environment is achieved by targeting cellular network for favorable to completion of the viral life cycle. Viruses can mimic human cellular miRNAs target to control existing molecular regulatory pathways for influence directly target mammalian RNA virus genomes.The function of BK polyomavirus viral miRNAs target remains unclear to host cellular molecular regulatory pathways. In this study, we investigated the role of the BK polyomavirus miRNA in the human gene (supplementary-7 & 8).

Conflict of Interest

No Conflict of Interest

ACKNOWLEDGEMENTS

The authors remain grateful to the DBT, BIF-BTBI scheme, GoI for providing the infrastructure. The authors would like to thank Aditi Gangopadhyay (DBT Centre for Bioinformatics, Presidency University, Kolkata) for her help during manuscript preparation.

REFERENCES

(1) David P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.

(2) *V. Narry Kim*, MicroRNA biogenesis: coordinated cropping and dicing, Nat. Rev. Mol. Cell Biol. 6 (2005) 376–385.

(3) Natascha Bushati and Stephen M. Cohen, microRNA Functions, Annu. Rev. Cell Dev. Biol. 23 (2007) 175–205

(4) MatthewW. Jones-Rhoades, David P. Bartel, and Bonnie Bartel, MicroRNAs and Their Regulatory Roles in Plants, Annu. Rev. Plant Biol. 57 (2006) 19–53

(5) Witold Filipowicz, Suvendra N. Bhattacharyya, and Nahum Sonenberg, Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9 (2006) 102–114

(6) Andrew Grimson, Kyle Kai-How Farh, Wendy K. Johnston, Philip Garrett-Engele, Lee P. Lim, and David P. Bartel, MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing, Mol. Cell. 27 (2007) 91–105

(7) David P. Bartel, MicroRNAs: Target Recognition and Regulatory Functions, Cell Rev. 136 (2009) 215–233

(8) Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA–target recognition. PLoS Biol 3(3): e85. (2005)404–418

(9) Robin C. Friedman, Kyle Kai-How Farh, Christopher B. Burge, Most mammalian mRNAs are conserved targets of microRNAs, *Genome Res.*19 (2009) 92-105

(10) Hui Liu, Dong Yue, Yidong Chen, Shou-Jiang Gao, Yufei Huang, Improving performance of mammalian microRNA target prediction. BMC Bioinformatics 11(2010) ,476

(11) Hamed Ahmadi, Ali Ahmadi, Sadegh Azimzadeh-Jamalkandi, Mahdi Aliyari Shoorehdeli, Ali Salehzadeh-Yazdi, Gholamreza Bidkhori, Ali Masoudi-Nejad, HomoTarget: A new algorithm for prediction of microRNA targets in Homo sapiens, Elsevier. Genomics 101 (2012) 94–100

(12) Sung-Kyu Kim, Jin-Wu Nam, Je-Keun Rhee, Wha-Jin Lee and Byoung-Tak Zhang, miTarget: microRNA target gene prediction using a support vector machine, BMC Bioinformatics 7 (2006) 441

(13) Yuchen Yang, Yu-Ping Wang and Kuo-Bin Li, MiRTif: a support vector machine-based microRNA-target interaction filter, *BMC Bioinformatics*, 9(2008) S4

(14) Vinod Chandra, Reshmi Girijadevi, Achuthsankar S Nair, Sreenadhan S Pillai and Radhakrishna M Pillai, MTar: a computational microRNA target prediction architecture for human transcriptome, BMC Bioinformatics 11(2010) S2

(15) Sanghamitra Bandyopadhyay and Ramkrishna Mitra, TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples, *Bioinformatics*. *25* (*2009*) *2625–2631*

(16) Mitra R, Bandyopadhyay S MultiMiTar, A Novel Multi-Objective Optimization based miRNA-Target Prediction Method. PLoS ONE 6((2011)): e24583

(17) Benjamin P. Lewis, Christopher B. Burge, and David P. Bartel1, Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell 120 (2005): 15–20

(18) Bino John, Anton J. Enright, Alexei Aravin, Thomas Tuschl, Chris Sander, Debora S. Marks, Human MicroRNA Targets. PLoS BIOLOGY 2(2004): e363

(19) Anton J Enright, Bino John, Ulrike Gaul, Thomas Tuschl, Chris Sander and Debora S Marks, MicroRNA targets in Drosophila. Genome Biology 2003, 5: R1

(20) Benjamin P. Lewis, I-hung Shih Matthew W. Jones-Rhoades, David P. Bartel and Christopher B. Burge, Prediction of Mammalian MicroRNA Targets. Cell 115(2003): 787–798

(21) T.F. Smith and M.S.Waterman, Identification of common molecular subsequences. j.Mol 147 (1981): 195-197

(22) Sean R Eddy, What is dynamic programming?. Nature Biotechnology 22(2004): 909–910

(23) Feifei Xiao, Zhixiang Zuo, Guoshuai Cai, Shuli Kang, Xiaolian Gao and Tongbin Li, miRecords: an integrated resource for microRNA–target interactions, Nucleic Acids Research 37 (2009) 105–110

(24) Thanasis Vergoulis, Ioannis S. Vlachos, Panagiotis Alexiou, George Georgakilas, Manolis Maragkakis, Martin Reczko, Stefanos Gerangelos, Nectarios Koziris, Theodore Dalamagas and Artemis G. Hatzigeorgiou, TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support Nucleic Acids Research 40 (2011) 222–229

(25) Sheng-Da Hsu, Yu-Ting Tseng, Sirjana Shrestha, Yu-Ling Lin, Anas Khaleel,Chih-Hung Chou, Chao-Fang Chu, Hsi-Yuan Huang, Ching-Min Lin, Shu-Yi Ho,Ting-Yan Jian, Feng-Mao Lin, Tzu-Hao Chang, Shun-Long Weng,Kuang-Wen Liao, I-En Liao, Chun-Chi Liu, and Hsien-Da Huang, miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Research 6(2013): 78–85

(26) Paula H. Reyes-Herrera, Elisa Ficarra, One Decade of Development and Evolution of microRNA Target Prediction Algorithms. Genomics Proteomics Bioinformatics, 10 (2012) 254–263

(27) Wan J. Hsieh, Feng-Mao Lin, Hsien-Da Huang, Hsiuying Wang, Investigating microRNA-Target Interaction-Supported Tissues in Human Cancer Tissues Based on miRNA and Target Gene Expression Profiling. PLOS ONE, 4(2014) e95697

(28) Chia-Huei Chu, Ann-Ping Tsou, Shu-Jen Chen, Hua-Chien Chen, Paul Wei-Che Hsu, Yung-Hao Wong, Yi-Hsuan Chen, Gian-Hung Chen3and Hsien-Da Huang, miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. Nucleic Acids Research,36(2008) D165–D169

(29) Doron Betel, Manda Wilson, Aaron Gabow, Debora S. Marks and Chris Sander, The microRNA.org resource: targets and expression. Nucleic Acids Research,36(2008) D149–D153

(30) Jan Kru ger and Marc Rehmsmeier, RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Research,34(2006) W451–W454

(31) Sam Griffiths-Jones, Harpreet Kaur Saini, Stijn van Dongen and Anton J. Enright, miRBase: tools for microRNA genomics. Nucleic Acids Research,36(2008) D154–D158

(32) Sam Griffiths-Jones, Harpreet Kaur Saini, Stijn van Dongen and Anton J. Enright, UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Research,38(2010) D75–D80

(33) Thorsten Joachims: SVMlight: Support Vector Machine. University of Dortmund 2008.

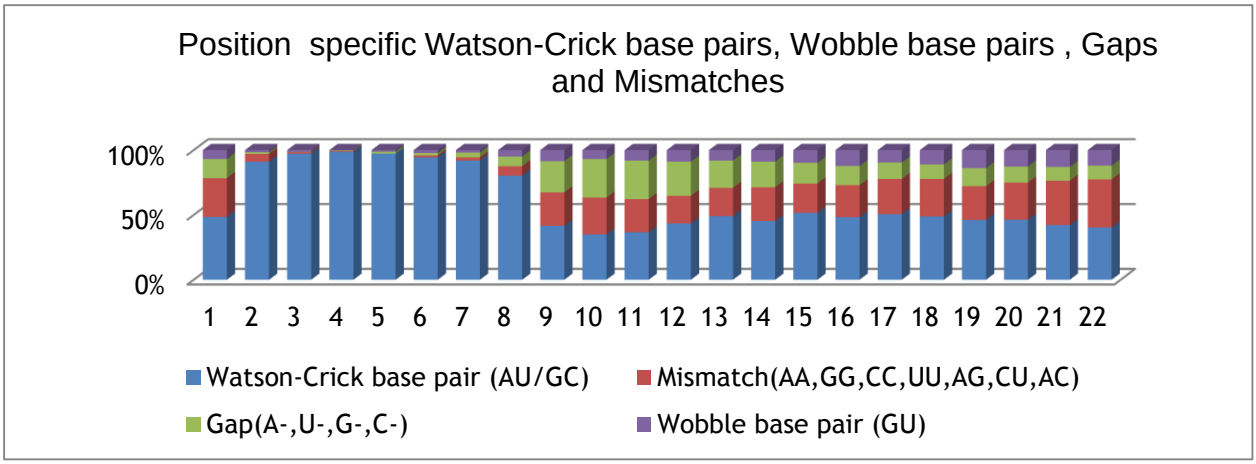
(34) Vladimir N. Vapnik, Statistical Learning Theory, *Wiley-Interscience (*1998)

(35) Package ROCR ,https://rocr.bioinf.mpi-sb.mpg.de/

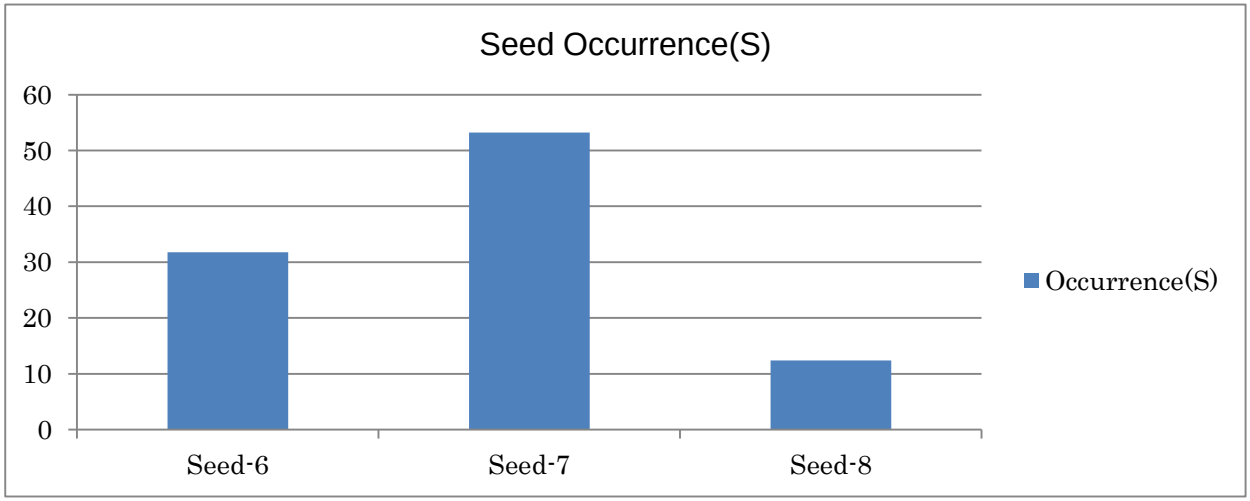
(36) R-3.2.2 , https://cran.r-project.org/bin/windows/base/

(37) Wenlong Xu, Anthony San Lucas, Zixing Wang, Yin Liu, Identifying microRNA targets in different gene regions, BMC Bioinformatics 2014, 15(Suppl 7): S4

Supplementary



Supplementary-1, position specific base pairings



Supplementary-2, Seed (6mer, 7mer, and 8mer) in 464 experimentally verified positive miRNA-mRNA target interactions