Biochemical and Biophysical Research Communications 413 (2011) 111–115

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

A probabilistic approach to microRNA-target binding

Hasan Oğul [a,*], Sinan U. Umu [b,c], Y. Yener Tuncel [c], Mahinur S. Akkaya [b]

[a] Department of Computer Engineering, Başkent University, Bağlıca TR-06810, Ankara, Turkey
[b] Department of Chemistry, Middle East Technical University, Çankaya TR-06800, Ankara, Turkey
[c] Bioinformatics Program, Informatics Institute, Middle East Technical University, Çankaya TR-06800, Ankara, Turkey

ARTICLE INFO

Article history:
Received 5 August 2011
Available online 22 August 2011

Keywords:
MicroRNA–target duplex
Variable Length Markov Chains
Probabilistic sequence model

ABSTRACT

Elucidation of microRNA activity is a crucial step in understanding gene regulation. One key problem in this effort is how to model the pairwise interactions of microRNAs with their targets. As this interaction is strongly mediated by their sequences, it is desired to set-up a probabilistic model to explain the binding preferences between a microRNA sequence and the sequence of a putative target. To this end, we introduce a new model of microRNA-target binding, which transforms an aligned duplex to a new sequence and defines the likelihood of this sequence using a Variable Length Markov Chain. It offers a complementary representation of microRNA–mRNA pairs for microRNA target prediction tools or other probabilistic frameworks of integrative gene regulation analysis. The performance of present model is evaluated by its ability to predict microRNA–target mRNA interaction given a mature microRNA sequence and a putative mRNA binding site. In regard to classification accuracy, it outperforms two recent methods based on thermodynamic stability and sequence complementarity. The experiments can also unveil the effects of base pairing types and non-seed region in duplex formation.

1. Introduction

MicroRNAs (miRNAs) are tiny regulators of gene expression in post-transcriptional level. Many evidences suggest that they abundantly exist in many organisms and play pivotal roles in regulation of several biological processes [1,2]. Last decade of bioinformatics research has been strongly influenced by the discovery of miRNAs. Recent findings about the complex behavior of miRNAs have triggered the application of computational methods to analyze miRNA functions and activities. Relevant computational research has been enriched in several directions like predicting miRNA targets [3–6], inferring regulatory modules or networks comprising miRNAs [7,8], identifying disease-related miRNAs [9,10] or finding functionally related miRNAs [11,12]. The studies have been resulted with various online tools or standalone programs developed for storage, retrieval or analysis of miRNA-related data.

Despite the enormous number miRNA target prediction programs, current systems biology can benefit too little from their outcome [13]. The reasons for the limitation of available target prediction tools are twofold. First, it is hard to see a consensus in their prediction results, which make them unreliable to use in generation of further hypotheses. Second, a conventional tool for target prediction can only reveal a one-to-one relationship between a miRNA and its predicted target under consideration. However, we know that the picture is more complex and larger: one miRNA can regulate many genes and one gene can be regulated by multiple miRNAs in cooperation, even with other regulatory factors [14]. An obvious solution to these problems is to integrate multiple data sources to infer more reliable and native explanations. The lack of sufficient data samples will also motivate the use of probabilistic methods to construct integrative frameworks to analyze the multi-source data. Since it is well-known that a miRNA can target an mRNA in a sequence-specific manner [2], an integrative model, whatever its practical purpose is, should incorporate the sequence information. Furthermore, such a model of sequence-based miRNA–mRNA binding should be defined in a probabilistic way so that it could represent the degree of this interaction and be easily incorporated into other data. Recent attempts for integrating sequence with functional data have largely ignored this issue and considered the sequence-directed miRNA–mRNA interaction in a binary way, which is usually obtained from other target predictors [15].

In this study, we aim to provide a means of modeling miRNA and mRNA regulatory relationship by a probabilistic description over the sequential content of a putative miRNA–mRNA duplex. In this respect, our model first performs a complementary alignment between the mature miRNA sequence and a putative binding site from present mRNA. Resulting alignment is represented by a new sequence over an alphabet of possible matches or mismatches, where different base paring rules are taken into account by distinct alphabet symbols. The probability of new sequence of miRNA–mRNA duplex is then analyzed using a Variable Length Markov Chain (VLMC)

* Corresponding author. Fax: +90 312 2341051.
E-mail address: hogul@baskent.edu.tr (H. Oğul).

0006-291X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2011.08.065

approach. VLMC [16] is a flexible yet powerful model to analyze a sequential content based on the order of local arrangements by quantifying the probability of the occurrence of a specific symbol after a certain sub-sequence with varying length less than a predefined maximum. This enables us to calculate the likelihood of whole sequence by simply multiplying local probabilities. For miRNA–mRNA duplex case, the independence of model from global position information allows the evaluation of significant local regions which might be enriched in any part of miRNA–target duplex formation. Therefore, the order of distinct base pairs and mismatches are taken into account in addition to their frequency of appearances. Two VLMC-based likelihoods of new sequence which are obtained from positive and negative training sets can reveal the degree of a potential interaction between two entities.

2. Materials and methods

2.1. Data set

For a comparative analysis with previous works, we used a common data set provided by Yang et al. [17]. The repository contains 233 miRNA-binding site pairs from drosophila, *Caenorhabditis elegans*, human, mouse, rat and zebrafish, where 195 of them are positive interactions and 38 are negatives. We refer this data set as DS1. To make deeper analysis of sequential effects in our model, we extended the data set by adding more recent experimentally supported positive and negative pairs. The latest release of miRecords [18] was used to collect additional positive samples with experimentally verified binding sites. Additional negative samples were obtained from the latest update of TarBase [19]. Final dataset comprises 283 positive and 115 negative mRNA–mRNA site pairs. This larger dataset is referred as DS2. In both datasets, negative samples still contain at least six perfect base pairs in seed region for reliable evaluation of the algorithms. Note that a positive interaction is referred as targeting relationship between miRNA and mRNA under consideration in this context.

2.2. miRNA–mRNA duplex sequence

miRNA–mRNA duplex is constructed by complementary alignment of mature miRNA sequence with mRNA binding site. Resulting alignment is transformed into a new sequence defined over an alphabet of symbols representing distinct nucleotide pair types including mismatch and space in any other site (Fig. 1). The alignment is performed using a dynamic programing algorithm with a penalty of -1 for both mismatches and gaps [20].

2.3. Markov chain model

Markov chains are used to model sequential data in terms of the order of individual symbols but regardless of their global positions. In its simplest form, called zero-order Markov chain, the likelihood of a sequence S_1^N is given by the probability that is obtained by multiplying the probabilities of each symbols contained, i.e.,

$$P(S_1^N) = \prod_{j=1}^{N} P(S_j = s_j)$$

where $P(.)$ refers to probability, S_j is the random variable representing the letter at position j with s_j as its realization. In this simple model, it is assumed that each position is independent from the others; the context information is not taken into account at all. Since the complex structure of biological sequences is poorly reflected with this approach, the use of higher order models has been suggested [21]. In a generalized form, called L-order Markov chain, the likelihood of the sequence is defined over the symbol probabilities which are now calculated depending on the preceding subsequence of a fixed length $L < N$. Although it is able to efficiently model rich sources, defining a fixed size on the length of the preceding subsequence has critical drawbacks for practical use; large number of model parameters and disregard of the domain specific nature of biological sequences [22]. Even with an optimized selection of L value, the model may not be effective in the detection of significant cut-off locations in the duplex chain. A more flexible version of higher order Markov models allows a variable order, or length, that depends on the preceding subsequence to given position such that the order of the model becomes a function the context at each position [16]. The final model is called as Variable Length Markov Chain (VLMC) and previously found applications in biological sequence analysis domain [23,24]. We further extend the model to take into account the succeeding subsequence and define the sequence likelihood as

$$P(S_1^N) = \prod_{j=1}^{N} P\left(S_j = s_j | S_{j-L_j}^{j-1} = s_{j-L_j}^{j-1}\right) \cdot P\left(S_j = s_j | S_{j+L_j'}^{j+1} = s_{j+L_j'}^{j+1}\right)$$

Where L_j and L_j' are the optimal lengths for preceding and succeeding subsequences respectively and $s_{j-L_j}^{j-1}$ and $s_{j+L_j'}^{j+1}$ are those sub-sequences. The last modification allows us to consider the context surrounding the symbol and provides a better generalization of the model. Since the lengths of the duplex sequences may slightly vary, we normalize the sum of log-likelihoods over the sequence length.

2.4. Interaction prediction

We consider the miRNA–mRNA interaction prediction problem as a binary classification task where a miRNA–mRNA duplex sequence is desired to be assigned to one of the positive ($C = 1$) or negative ($C \neq 1$) classes. Given a sequence $S_1^N = s_1 s_2 \ldots s_N$, which resides between the position 1 and the position N, where s_i is drawn from the finite alphabet of nucleotide pair symbols, we employ a classification rule based on the likelihood of the S_1^N that outputs a positive (target interaction) label if $P(S_1^N|C = 1) > P(S_1^N|C \neq 1)$ and a negative (no interaction) label otherwise.

2.5. Availability

An updated C implementation of Bejerano and Yona's VLMC algorithm [22] was used to learn the conditional probabilities from corresponding training sets [25]. A Linux-compatible Perl script to perform duplex alignment, VLMC runs and interaction prediction can be obtained from our web site at www.baskent.edu.tr/~hogul/mirna. The datasets (DS1 and DS2) are also available in the same page.

3. Results

We evaluated our model by its ability to predict miRNA–target interaction when we are given a miRNA sequence and a putative binding site. We compared it with two major approaches; thermodynamic stability and sequence complementarity. RNAHybrid is one of the earliest methods to model miRNA–mRNA binding in terms of thermodynamic properties of resulting duplex [26]. It calculates the minimum free energy to form the duplex from predicted structure information. Several target prediction tools have utilized RNAHybrid methodology in their algorithms in addition

```
5' GGGUGUUAAGACUUGACACAGUACCUCG 3'
   .. |.| | .||.| ||    ||||||.
3' UUGAUA-UGUUGGA-UG----AUGGAGU 5'
   eegbebqageadfbqadqqqqbaccbce
```

Fig. 1. Alignment of miRNA sequence (middle) and mRNA binding site (top) is transformed into a new sequence of duplex formation (bottom).

to other factors [5]. miRTif is an advanced sequence-based model for miRNA-target binding [17]. The algorithm counts the frequencies of some predefined motifs in several lengths, defined over the complementary/noncomplementary pairs in the duplex. These frequencies are used to feed a powerful machine learning method, called Support Vector Machines (SVM), which is able to learn an optimal hyperplane to separate positive and negative examples based on these features.

miRNA target prediction tools have often used validated data together with some predicted data either to develop or assess their algorithms in order not to suffer from small sample size. When an evaluation of target prediction ability for given miRNAs and genes is in question, it is reasonable to ignore the lack of verified binding site information since they only deal with the binary result. However, a correct and reliable evaluation of a computational binding model requires a set of experimentally verified binding sites, both for positive and negative cases. Yang et al., the developers of miRTif, collected a number of experimentally validated positive and negative miRNA–mRNA duplex sets (DS1) and tested their method in this benchmark set to predict miRNA–target interactions [17]. They reported the sensitivity (percentage of actual positives correctly predicted as such), specificity (percentage of actual negatives correctly predicted as such), accuracy (percentage of all correctly predicted samples) and AUC (area under Receiver Operating Characteristics Curve) of their predictions based on tenfold cross-validation tests, where, at each fold, 1/9 of the samples are predicted based on the training over other 9/10 samples. RNAHybrid does not require a training stage, instead, it checks the predicted energy whether it is above or below a predefined threshold. We compiled the RNAHybrid algorithm in this way, but set the energy threshold for classification such that it optimizes the trade-off between sensitivity and specificity at each fold of cross-validation.

To make a fair comparison, we also compiled our method in a tenfold cross-validation set-up on the same data set. During this experiment, we treated the miRNA–mRNA duplex as a sequence over a 8-letter alphabet, where each directed base pair, including wobbles, i.e. A-U, U-A, G-C, C-G, G-U and U-G, was represented by a distinct letter and all mismatches were unified into one letter. An additional letter is used for a space in any other site. Maximum length for Markov chains was set to 5 since longer motifs are quite rare, which may not contribute to the result, but several significant motifs could be observed up to 5-letter length. We ignored the single appearance of any sub-string in positive or negative set to eliminate the data set bias. Results of the experiments are shown in Table 1 in comparison with RNAHybrid and miRTif.

The table demonstrates that new method can outperform RNAHybrid and miRTif in all evaluation criteria except specificity, where it performs equally with miRTif. The superiority of present method is quite significant in terms of AUC statistics. This shows that our new method is not only successful in separation of positive and negative examples but also able to successfully quantize the level of certainty in its prediction. This result is consistent with our intention to release this model; enabling the sequence information to be easily integrated with other data with the degree of its influence.

Table 1
Comparison of present model with RNAHybrid and miRTif on miRNA–mRNA interaction prediction.

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
RNAHybrid	64.1	60.5	63.7	0.71
miRTif	83.6	73.7	81.9	0.89
Present model	86.7	73.7	84.6	0.95

Relevance of seed complementarity to miRNA target selection has been approved in several publications [6,14]. However, the contribution of other parts of mature miRNA is still elusive. We evaluated the effect of non-seed region by replicating the experiments in three ways. First, the experiment was performed using only seed region of duplex. Second, the same was repeated for only non-seed region. And finally, we considered the independent effects of both regions over a weighted sum of their log-likelihoods. Here, we used a manual parameter, α, which adjusts the weight of seed region, e.g., $\alpha = 0.5$ means that they contribute equally or, when $\alpha = 0.8$, seed information is more effective whereas non-seed region says little. Fig. 2 depicts the effect of seed and non-seed region by number of incorrect predictions for changing values of α.

Although seed region itself can predict interactions to some degree, and even be more successful in rejection of false positives ($\alpha = 1.0$), total number of incorrect predictions was minimized when the other parts were allowed to contribute to the result in fair ($\alpha = 0.5$). Nevertheless, best results were obtained when they were not treated independently but considered as a whole sequence. In this case, the number of incorrect predictions was lowered to 36 (experiment in Table 1), while it was 39 in their independent evaluation (Fig. 2).

Finally, we assessed our choice of duplex alphabet. Our set-up has a biochemical intuition. Argonaute (AGO) protein, which is a major component of RNA-induced silencing complex (RISC), prefers miRNA but not miRNA* strand in the machinery of duplex formation. Selection of correct strand is related to thermodynamic properties of miRNA duplex [2,27]. Recent experiments also revealed that strand selection is guided by asymmetric thermostability of duplex termini [28,29]. It means that the direction of any base pair between two RNAs may lead to different thermodynamics, therefore, may affect the stability of resulting duplex. Since this is usually invalid for mismatches, we prefer to unify all mismatches into one letter, but use a different letter to represent each directional base pair type. We also represent G-U and U-G wobbles by distinct letters since wobble pairs have shown to be comparable to other Watson–Crick pairs in terms of their thermodynamic stability and as effective as them in RNA-level bonding [30]. In the past, several target prediction tools treated G-U wobbles differentially as opposed to mismatches and the other base pairs in their algorithms [6,17,30].

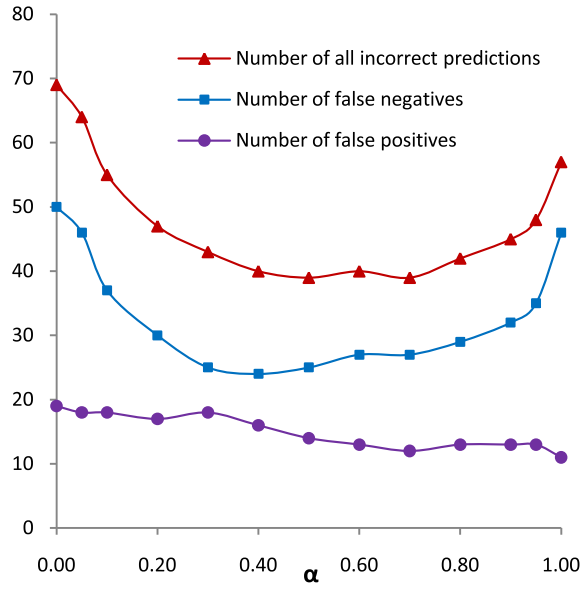

Fig. 2. Effects of seed and non-seed regions on interaction prediction results. A higher value of α refers to higher contribution of seed.

To support our hypothesis, we repeated the experiments in both datasets (DS1 and DS2) with several different configurations of the alphabet. A hierarchical representation of the alphabets is shown in Fig. 3. Starting from widest alphabet with 18 letters; 16 for all possible pairs of nucleotides and two for spaces, we gradually reduced the number of letters in each level by combining the related ones. Table 2 depicts the results with each level of alphabet. Best accuracy was obtained with level-2 alphabet, which corresponds to our choice based on the structure and thermodynamic property of certain nucleotide pairs. When we ignored the direction and type of the base pairs we observed a serious decline (about 30%) in the prediction accuracy. Moreover, 8-letter alphabet is slightly better than 18-letter alphabet where each possible pair type including mismatches is represented by a different letter. This difference is more apparent in smaller dataset due to the sparsity of mismatch types in the training set.

4. Discussion

We have introduced a probabilistic method to model the miRNA–target duplex formation and evaluated its performance on prediction of the interaction when miRNA sequence and a putative binding site are given. The accurate results that we obtained from the experiments suggest that present model is able to capture compositional properties of a duplex sequence by additionally considering the effect of different base pairings, mismatches and gaps with their arrangements inside the duplex. The model does not rely on any explicit use of known or hypothetical rule based on sequential structures, such as seed complementarity, thermodynamics or central loops, but rather implicitly accommodate the effects of any types of relevant motifs, either known or unknown, provided that they are present in the training set. Therefore, increasing number of experimentally validated data will promise higher prediction accuracy.

The experiments have confirmed the contribution of non-seed region in target selection. Another notable result is that the accuracy can be increased when each base paring type, including G-U wobble, with its direction is considered as a distinct match, instead of interpreting it simply as any base pair. Type of mismatches does not have any significant influence. We believe that this finding will be beneficial in other studies, because existing works have considered the issue by only appearances of perfect base pairs and mismatches, although some already emphasized the effect of G-U wobbles.

The model which we proposed may find applications in several platforms. First, it can be used as a post-processing filter for other miRNA target prediction tools. Many of available algorithms con-

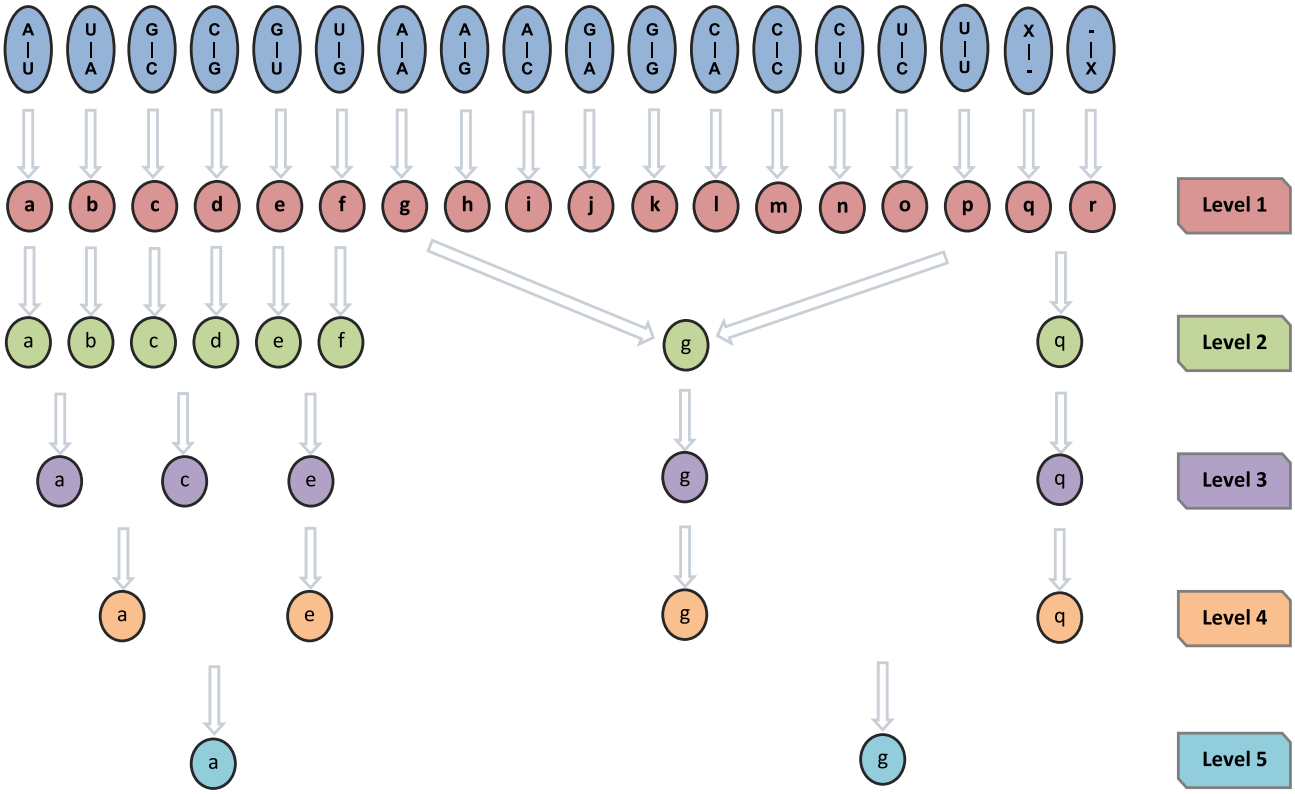

Fig. 3. A hierarchy of miRNA–mRNA duplex alphabet.

Table 2
Effects of different duplex alphabets (Fig. 3) in interaction prediction performance over the datasets DS1 and DS2.

Alphabet	DS1			DS2		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Level-1	75.5	50.0	80.5	**82.6**	71.3	79.3
Level-2	**86.7**	**73.7**	**84.6**	81.9	**76.5**	**80.3**
Level-3	70.2	50.0	66.9	70.6	58.2	67.0
Level-4	70.2	39.5	65.2	62.0	35.6	54.4
Level-5	53.8	57.9	54.5	57.8	35.6	51.4

sider the seed match as a strong evidence for target identification. Since it is possible to observe random mRNA matches to seed region without any interaction, this decision criterion can mislead the algorithm to produce excessive number of false positives. Present method is able to reject miRNA–nontarget duplexes despite a high seed complementarity, thus it may help to reduce the number of false positives by reanalysing the binding site predicted by former tool. Concurrent use of seed and non-seed region is also proven to increase the reliability of predictions. Second, it may serve complementary information which can be deployed in target prediction algorithms. Conventional methods perform a window-based linear scan over the mRNA sequence to identify a putative binding site which may attain a large binding score based on a weighted sum of predefined criteria. Output of proposed model is an obvious complement to other determinants such as structure, site accessibility or cross-species conservation in this scoring scheme. Third, the model enables the researchers to integrate sequence data directly with other behavioral data such as gene expression profiles over a probabilistic framework. An integrated framework can provide a comprehensive analysis of miRNA functions associated with other entities, conditions or diseases. Machine learning research has been competing in two directions for intelligent analysis of heterogeneous data: black-box kernel methods such as Support Vector Machines and probabilistic graphical models such as Bayesian Networks. Latter requires a probabilistic representation of each contributor in the model. Present scheme can fill a gap in this respect.

It is anticipated that the participation of computational models into miRNA research will increasingly continue in coming years. We believe that integration of multi-source heterogeneous data will be a focal point in this research. Our study does not yield a standalone tool in this context; however, it provides a different view of miRNA–target interactions from which future research can definitely benefit.

Acknowledgment

This study was supported by the Scientific and Technological Research Council of Turkey (TUBİTAK) under the Project 110E160.

References

[1] R.C. Lee, R.L. Feinbaum, V. Ambros, The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14, Cell 75 (1993) 843–854.
[2] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[3] P. Alexiou, M. Maragkakis, G.L. Papadopoulos, et al., Lost in translation: an assessment and perspective for computational microRNA target identification, Bioinformatics 25 (2009) 3049–3055.
[4] P.M. Szabo, Z. Tombol, V. Molnar, et al., MicroRNA target prediction: problems and possible solutions, Curr. Bioinf. 5 (2010) 81–88.
[5] M. Hammell, Computational methods to identify miRNA targets, Semin. Cell Dev. Biol. 21 (2010) 738–744.
[6] D.P. Bartel, MicroRNAs: target recognition and regulatory functions, Cell 136 (2009) 215–233.
[7] S. Yoon, G. Micheli, Prediction of regulatory modules comprising microRNAs and target genes, Bioinformatics 21 (2005) i93–i100.
[8] B. Liu, L. Liu, A. Tsykin, et al., Identifying functional miRNA–mRNA regulatory modules with correspondence latent dirichlet allocation, Bioinformatics 26 (2010) 3105–3111.
[9] J. Lu, G. Getz, E.A. Miska, et al., MicroRNA expression profiles classify human cancers, Nature 435 (2005) 834–838.
[10] S.F. Madden, S.B. Carpenter, I.B. Jeffery, et al., Detecting microRNA activity from gene expression data, BMC Bioinf. 11 (2010) 257.
[11] G. Yu, Q. He, Functional similarity analysis of human virus-encoded miRNAs, J. Clin. Bioinf. 1 (2011) 1–15.
[12] D. Wang, J. Wang, M. Lu, et al., Bioinformatics 26 (2010) 1644–1650.
[13] C. Barbato, I. Arisi, M.E. Frizzo, et al., Computational challenges in miRNA target predictions: to be or not to be a true target?, J Biomed. Biotechnol. 2009 (2009) 803069.
[14] A. Krek, D. Grun, M.N. Poy, et al., Combinatorial microRNA target predictions, Nat. Genet. 37 (2005) 495–500.
[15] J.C. Huang, Q.D. Morris, B.J. Frey, Bayesian inference of microRNA targets from sequence and expression data, J. Comput. Biol. 14 (2007) 550–563.
[16] D. Ron, Y. Singer, N. Tishby, The power of amnesia: learning probabilistic automata with variable memory length, Mach. Learn. 25 (1996) 117–149.
[17] Y. Yang, Y.P. Wang, K.B. Li, miRTif: a support vector machine-based microRNA target interaction filter, BMC Bioinf. 9 (2008) S4.
[18] F. Xiao, Z. Zuo, G. Cai, et al., miRecords: an integrated resource for microRNA–target interactions, Nucleic Acids Res. 37 (2009) D105–D110.
[19] G.L. Papadopoulos, M. Reczko, V.A. Simossis, et al., The database of experimentally supported targets: a functional update of tarbase, Nucleic Acids Res. 37 (2009) D155–D158.
[20] T. Smith, M. Waterman, Identification of common molecular subsequences, J. Mol. Biol. 147 (1981) 195–197.
[21] G. Thijs, M. Lescot, K. Marchal, et al., A higher order background model improves the detection of promoter regulatory elements by Gibbs sampling, Bioinformatics 17 (2001) 1113–1122.
[22] G. Bejerano, G. Yona, Variations on probabilistic suffix trees: statistical modeling and prediction of protein families, Bioinformatics 17 (2001) 23–43.
[23] H. Ogul, Variable context Markov chains for HIV protease cleavage site prediction, Biosystems 96 (2009) 246–250.
[24] I. Ben-Gal, A. Shani, A. Gohr, et al., Identification of transcription factor binding sites with variable-order Bayesian networks, Bioinformatics 21 (2004) 2657–2666.
[25] F.G. Leonardi, A generalization of the pst algorithm: modeling the sparse nature of protein sequences, Bioinformatics 22 (2006) 1302–1307.
[26] M. Rehmsmeier, P. Steffen, M. Hochmann, et al., Fast and effective prediction of microRNA/target duplexes, RNA 10 (2004) 1507–1517.
[27] X. Chen, MicroRNA metabolism in plants, Curr. Top. Microbiol. 320 (2008) 117–136.
[28] H.Y. Hu, Z. Yan, Y. Xu, et al., Sequence features associated with microRNA strand selection in humans and flies, BMC Genomics 10 (2009) 413.
[29] A.L. Eamens, N.A. Smith, S.J. Curtin, et al., The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes, RNA 15 (2009) 2219–2235.
[30] P. Brodersen, O. Voinnet, Revisiting the principles of microRNA target recognition and mode of action, Nat. Rev. Mol. Cell Biol. 10 (2009) 141–148.