Original Article

Computational Detection of piRNA in Human Using Support Vector Machine

Atefeh Seyeddokht [1], Ali Asghar Aslaminejad [1*], Ali Masoudi-Nejad [2], Mohammadreza Nassiri [3], Javad Zahiri [4], and Balal Sadeghi [5]

1. Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
2. Laboratory of System Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
3. Department of Animal Science, Research Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran
4. Bioinformatics and Computational Omics. LAB (BioCOOL), Faculty of Biological Sciences, Tarbiat Modares University (TMU), Tehran, Iran
5. Department of Food Hygiene and Public Health, Faculty of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman, Iran

Abstract

Background: Piwi-interacting RNAs (piRNAs) are small non-coding RNAs (ncRNAs), with a length of about 24-32 nucleotides, which have been discovered recently. These ncRNAs play an important role in germline development, transposon silencing, epigenetic regulation, protecting the genome from invasive transposable elements, and the pathophysiology of diseases such as cancer. piRNA identification is challenging due to the lack of conserved piRNA sequences and structural elements.

Methods: To detect piRNAs, an appropriate feature set, including 8 diverse feature groups to encode each RNA was applied. In addition, a Support Vector Machine (SVM) classifier was used with optimized parameters for RNA classification. According to the obtained results, the classification performance using the optimized feature subsets was much higher than the one in previously published studies.

Results: Our results revealed 98% accuracy, Mathew' correlation coefficient of 98% and 99% specificity in discriminating piRNAs from the other RNAs. Also, the obtained results show that the proposed method outperforms its competitors.

Conclusion: In this paper, a prediction method was proposed to identify piRNA in human. Also, 48 heterogeneous features (sequence and structural features) were used to encode RNAs. To assess the performance of the method, a benchmark dataset containing 515 piRNAs and 1206 types of other RNAs was constructed. Our method reached the accuracy of 99% on the benchmark dataset. Also, our analysis revealed that the structural features are the most contributing features in piRNA prediction.

Avicenna J Med Biotech 2016; 8(1): 36-41

Keywords: Piwi-interacting RNAs (piRNAs), RNA, Support Vector Machines (SVM)

*** Corresponding author:**
Ali Asghar Aslaminejad, Ph.D.,
Department of Animal Science,
Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
Tel: +98 9155161513
Fax: +98 51 38804633
E-mail:
agr764@yahoo.co.uk
Received: 22 Jun 2015
Accepted: 15 Oct 2015

Introduction

In recent years, numerous confirmations from high-throughput genomic programs show that albeit less than 2% of the mammalian genome translates proteins, a major segment can be transcribed into diverse mixed members of non-coding RNAs (ncRNAs) [1-3]. The Encyclopedia of DNA Elements (ENCODE) and associated projects indicated that the majority of eukaryotic transcripts are ncRNAs [4]. There are more than twenty thousand protein-coding genes in the human genome, which correspond to roughly two percent of human genome [5]. The remaining regions in the human genome are non-coding RNAs, which were previously named "dark substance" or "junk DNA" [6]. Recently, ncRNAs have attracted significant attentions with respect to their various biological roles, highlighting the biological importance of previously "overlooked" RNA reservoir [7]. ncRNAs are complicated elements with different significant biological functions in the cell, including the control of chromosome dynamics, RNA splicing, RNA excision, translational inhibition and mRNA demolition [8]. ncRNAs can be coarsely categorized into small ncRNAs (such as small nucleolar RNAs (snoRNAs), short-interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs), microRNAs (miRNAs), and short hairpin RNAs (shRNAs)) or long ncRNAs (lncRNAs), based on the transcript size [9-12]. The range of

ncRNAs is growing rapidly as new ncRNAs remain to be discovered by high-throughput sequencing methods. Nevertheless, a large number of ncRNAs presumably cannot be recognized [8,13]. Therefore, the identification and explanation of ncRNAs is a considerable step for the explanation of different regulatory mechanisms in the cell. piRNAs are about 19 to 33 nucleotides long and most of their sequences fall in the range of 25-33 nucleotides. These ncRNAs are similar to siRNA and miRNA which also have a strong preference for the 5'uridine. Besides, piRNA molecules are located in clusters of length 20-100 *kb*. The density of piRNA clusters ranges from 40 to 4000 [14-18]. piRNAs are the most diverse and the most expressed small ncRNAs in animals [19,20]. They are involved in epigenetic and post-transcriptional regulation of retrotransposons [21]. Lately, numerous researches have started to discover the hitherto unknown pathways of piRNA synthesis [22]. Until now, a large number of piRNA sequences have been identified in human [23], mouse [16], rat [17], zebra fish [24], and fruit fly [25]. Computational identification approaches can supply experimental approaches to identify ncRNAs quickly in novel genomes, specifically the ncRNAs that are transcribed under particular conditions in specific cell types. Many computational methods have been suggested for ncRNAs prediction, consisting of comparative [26-29] and non-comparative methods [30-37].

In recent years, some studies were devoted to analyze computational identification of piRNAs [38-41]. Brayet *et al* [38] integrated machine learning method based on multiple kernels and a Support Vector Machine (SVM) classifier to identify the human and *Drosophila* piRNAs. Their method combined previously identified features and a new telomere/centromere neighborhood feature. The results from their SPG-GMKL method were better than the ones reported by Zhang *et al* [41] (>0.8 in almost all measurements for both Human and *Drosophila)*. Wang *et al* [39] performed transposon interaction and a SVM for piRNAs prediction. They used SVM to predict human, mouse and rat piRNAs, and they achieved 90.6% accuracy. They developed Piano program to predict piRNAs for the rice stem borer, *Chilo suppressalis*. They achieved an accuracy and sensitivity of 95%, and 96%, respectively. Betel *et al* [40] trained a SVM classifier to distinguish between 5'-RNA and all other uridin positions for mouse piRNA sequences. In this way, they could identify mouse piRNAs with a precision of 61-72 percent. But their method could not effectively predict those piRNA derived from the 3'-UTR of mRNA which are produced by Ping Pong model. Also, Zhang *et al* [41] used Fisher separator algorithm by setting different cutoffs for piRNA identification in five model species including mice, humans, rats, fruit fly, and nematode. Their approach reached a precision of over 90% and a sensitivity of over 60%. But these studies are computationally intensive or they did not show a satisfactory prediction performance. In this study, to find a set of effective descriptors for discriminating piRNAs from other ncRNAs, heterogeneous features of ncRNAs were extracted [42,43]. Then, the SVM classifier with different parameters was performed to detect an optimized feature subset. The results show that different feature types have different discriminative power in piRNA prediction. Also, according to the results (our method achieved accuracy of 98% and specificity of 99%), the proposed method can be used effectively for piRNA detection.

Materials and Methods

Dataset

In order to acquire the appropriate piRNA dataset, 515 piRNAs from piRNABank [44] (http://pirnabank.ibab.ac.in/) as positive instances were extracted.

To construct negative dataset, three groups of RNAs were used. The first one was composed of precursor miRNA sequences (http://www.mirbase.org/, version 21), the second group consisted of non-piRNA sequences and they were derived from different databases including 1200 sequences of various types:
sequences of lncRNA, extracted from NCBI (http://www.ncbi.nlm.nih.gov).
precursor miRNAs, extracted from miRBase (http://www.mirbase.org/, version 21); and
human mRNA sequences ,downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov).
And the third group was composed of:
sequences of snoRNA, collected from snoRNA-LBME-db database (https://www-snorna.biotoul.fr);
precursor miRNA sequences, extracted from miRBase (http://www.mirbase.org/, version 21); and
sequences of tRNA, extracted from genomic tRNA database (http://gtrnadb.ucsc.edu/). As a result, 1206 sequences of the third group were selected as negative instances.

Extracted features

Sequence-based features: Sequence-based features have showed a discriminatory power to predict biological functions of macromolecules [45,46]. With respect to the sequence features, the frequency of two neighboring bases (*e.g*., %AA), 15 sequence motifs [43] and the content of G and C (%G+C) formed the sequence-based feature sets.

Structural features

Generally, structural attributes are significant for the identification of human piRNAs. Thus, sequence and structural features were incorporated to recognize human piRNAs. RNAfold program [47] was used with the default parameters to calculate structural features based on the RNA secondary structures. Since the Minimum Free Energy (MFE) is an index that assesses the stability of the secondary structure of non-coding RNAs, several structural features, including MFE, as the structural feature sets were selected. Table 1 shows the total 48 features.

Table 1. The final 48 features used for building our model

No. of features	Feature	Description
1	ANAA	Motif
2	CNTG	Motif
3	CNTA	Motif
4	CTNT	Motif
5	CNTC	Motif
6	CAC	Motif
7	CNTNT	Motif
8	GNCA	Motif
9	ATA	Motif
10	ANTT	Motif
11	TGNNT	Motif
12	GNAC	Motif
13	CTT	Motif
14	ANNCT	Motif
15	AGNG	Motif
16	%G+C	GC content
17-32	%XY	Frequency of dinucleotide XY (A,T,C,G)
33	MFEI1	Index 1 based on the minimum free energy
34	MFEI2	Index 2 based on the minimum free energy
35	MFEI3	Index 3 based on the minimum free energy
36	MFEI4	Index 4 based on the minimum free energy
37	dG	Normalized minimum free energy
38	dp	Normalized base-pairing propensity
39	NEFE	Normalized ensemble free energy
40	Freq	Frequency of the MFE structure
41	Diff	Structural diversity
42	|A-U|/L	Normalized base pair counts
43	|G-C|/L	Normalized base pair counts
44	|G-U|/L	Normalized base pair counts
45	ABS	Average base pairs per stem
46	%(A-U)/s	Based on the average base pairs per stem
47	%(G-U)/s	Based on the average base pairs per stem
48	%(G-C)/s	Based on the average base pairs per stem

Support vector machine (SVM)

SVM is an efficient machine learning technique, responding to classification problems in bioinformatics and computational biology [48]. The SVM is able to convert low-dimensional non-linear matters into high-dimensional linear problems, resolving non-linear classification issues by reducing the linear classification problems [49]. In this study, two models were trained and tested separately using one- and two-class SVM for human dataset. In addition, for optimizing the one-class SVM model, the radial basis function (RBF) kernel parameter nu (γ) were adjusted by the grid search strategy in MATLAB. Figure 1 illustrates the pipeline for piRNA identification.

Performance evaluation

A 10-fold cross validation procedure was used to assess the performance of the proposed model using four measures: sensitivity (SE), specificity (SPC), accuracy (ACC), and Matthews's correlation coefficient (MCC). The formulas of these four measures are as follows:

$$SE=TP/(TP+FN)$$

$$SPC=TN/(TN+FP)$$

$$ACC=(TP+TN)/(TN+FP+TP+FN)$$

$$MCC=\frac{(TP\times TN-FP\times FN)}{\sqrt{(TP+FP)(TN+FN)(TP+FN)(TN+FP)}}$$

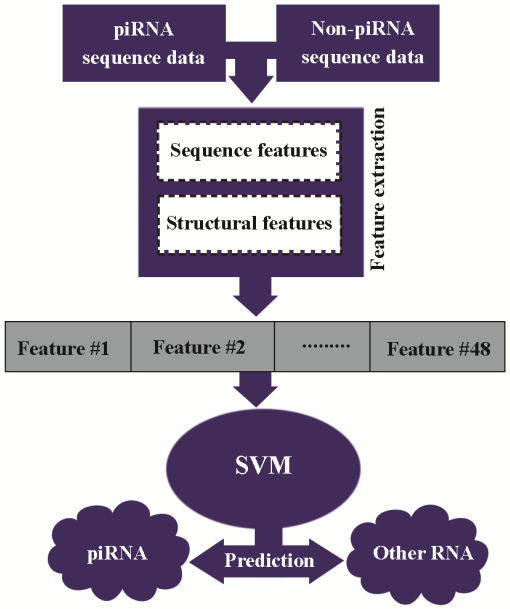

Figure 1. Flowchart describing the pipeline for piRNA identification.

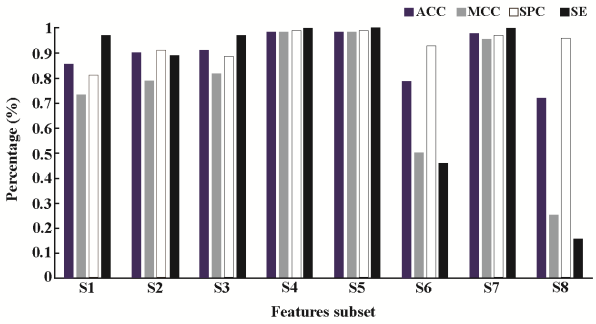

Figure 2. Different performance measures when different subsets of features were used.

TP and TN are the number of piRNAs and non-piRNAs, respectively, that predicted correctly. Also, FN and FP are the number of piRNAs and non-piRNAs that predicted wrongly [38].

Results

Kernel function selection

One of the most important parameters in SVM classifier is the kernel function [50]. In this study, four different kernel functions were used to obtain the best SVM model for piRNA prediction: Linear, Polynomial, Radial Basis Function (RBF) and Sigmoid. As figure 2 shows, the SVM with RBF kernel performed better than the others.

Optimizing the nu (γ) parameter

Selecting the appropriate value for the nu (γ) parameter is a critical step in SVM model with RBF kernel function[51]. To optimize this parameter, different values of γ were tried (Figure 3). According to the obtained results, the best values for γ were 0.06 and 0.08 in the training dataset.

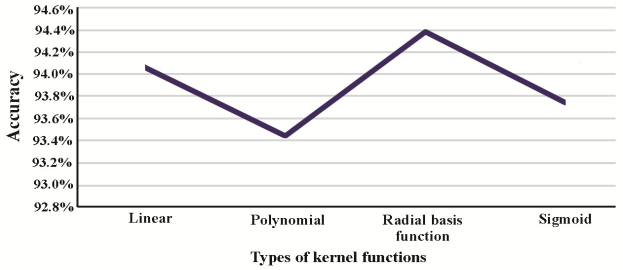

Figure 3. Accuracy of the SVM model with different kernel functions.

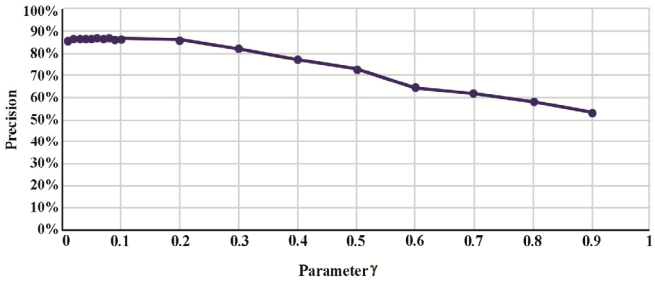

Figure 4. Different values of the parameter γ in SVM model.

Table 2. Performance of the SVM using different subset of features

Description of groups with different features	Features' subset	ACC	MCC	SPC
15 motif features (Features 1-15)	S1	0.85	0.73	0.81
10 features corresponding to the sequence (Features 16-25)	S2	0.90	0.78	0.91
17 features corresponding to the sequence (Features 16-32)	S3	0.91	0.81	0.88
9 structural features (Features 33-41)	S4	0.98	0.98	0.99
7 features corresponding to the base pair (Features 42-48)	S5	0.98	0.98	0.99
22 features corresponding to the motif and base pair (Features 1-15 and 42-48)	S6	0.78	0.50	0.92
16 features corresponding to structure and base pair (Features 34-48)	S7	0.97	0.95	0.97
31 features corresponding to structure, base pair, and motif (Features 1-15, and 33-48)	S8	0.71	0.25	0.96

Table 3. Comparison with other methods

Method	SPC (%)	SE (%)	ACC (%)
Zhang *et al* [41]	98	52	75
Liu *et al* [43]	89	91	90
Brayet *et al* [38]	82	30	58
Our method	99	99	98

Assessing the prediction performance

The 48 features were divided into eight overlapping groups and in each group the optimized SVM was run. These eight groups were 15 motif features (Features 1-15), 10 features corresponding to the sequence (Features 16-25), 17 features corresponding to the sequence (Features 16-32), 9 structural features (Features 33-41), 7 features corresponding to the base pair (Features 42-48), 22 features corresponding to the motif and base pair (Features 1-15 and 42-48), 16 features corresponding to structure and base pair (Features 34-48) and finally, 31 features corresponding to structure, base pair, and motif (Features 1-15, and 33-48). Table 2 shows the performance of the SVM using different subset of features. As table 2 shows, the best performances were achieved for the fourth and fifth groups. Figure 4 provides a better illustration of the performance when different subsets of features were used.

Thus, such structural and base pair features (Table 2) that were previously used to classify real and pseudo miRNAs [42] propose the similarity of these structure-base pair features for all types of small ncRNAs [39].

Discussion

In this paper, SVM was exploited to predict piRNA in human. After examining various kernel functions, RBF kernel function was used for the SVM. Choosing appropriate descriptors are of considerable importance to encode RNAs for building an accurate model. In this study, 48 various descriptors were used to build feature vectors. These features can roughly be categorized into eight overlapping groups (Table 2).

To assess the contribution of different feature types in piRNA prediction, performance of the SVM model was computed using different feature subsets. Our results showed that structural features (Features 33-41) and 7 features corresponding to the base pair (Features 42-48) are the most contributing features. These features had near perfect performance (accuracy of 98% and sensitivity of 99%).

Also, an attempt was made to compare the proposed method with the three recently published methods of Zhang *et al* [41], Lakshmi *et al* [43] and Brayet *et al* [38]. Zhang *et al* [41] used k-mer schema to identify piRNA sequence in five model species. Liu *et al* [43] developed a method for piRNA identification based on motif discovery using SVM classifier, named Pibomd. Brayet *et al* [38] proposed an algorithm, named piRPred, to identify piRNAs. They used a multiple kernel fusion and an SVM-based approach which allow using heterogeneous features. To have a fair comparison, these methods were run on our dataset. As table 3 shows, our method outperformed the other methods in three different performance measures.

The biological significance and the functions of piRNA molecules are currently the subject of intensive study and, numerous researches have started to uncover the hitherto unknown biological mechanisms of piRNA [22,52,53]. Therefore, computational identification of piRNA has been at the forefront of research for understanding the mechanisms that maintain germline integ-

rity. According to the results, it can be stated that feature subsets 4 and 5 are the most contributing features in piRNAs prediction. So, secondary structure features and base pairing information, which can be computed by appropriate tools, can effectively help the biologists to design and discriminate piRNAs from other RNAs [47,54].

Conclusion

In this paper, a prediction method was proposed to identify piRNA in human. Also, 48 heterogeneous features (sequence and structural features) were used to encode RNAs. To assess the performance of the method, a benchmark dataset containing 515 piRNAs and 1206 types of other RNAs was constructed. Our method reached the accuracy of 99% on the benchmark dataset. Also, our analysis revealed that the structural features are the most contributing features in piRNA prediction. Finally, three recently published computational studies were used in piRNA detection. The obtained results show that the proposed method outperforms its competitors.

Acknowledgement

We are very grateful to Dr. Ahmadi and Dr. Azimzadeh, for their helpful suggestions.

References

1. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science 2005;309(5740):1559-1563.
2. Kadri S, Hinman V, Benos PV. HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models. BMC Bioinformatics 2009;10 Suppl 1: S35.
3. Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, et al. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res 2005;15(7):987-997.
4. Maher B. ENCODE: The human encyclopaedia. Nature 2012;489(7414):46-48.
5. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291(5507):1304-1351.
6. Ponting CP, Belgard TG. Transcribed dark matter: meaning or myth? Hum Mol Genet 2010;19(R2):R162-168.
7. Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long noncoding RNAs. Nucleic Acids Res 2012;40(14):6391-6400.
8. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet 2006;15 Spec No 1:R17-29.
9. Brosnan CA, Voinnet O. The long and the short of noncoding RNAs. Curr Opin Cell Biol 2009;21(3):416-425.
10. Managadze D, Rogozin IB, Chernikova D, Shabalina SA, Koonin EV. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. Genome Biol Evol 2011;3:1390-1404.
11. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10 (3):155-159.
12. Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet 2011;12(2):136-149.
13. Marques AC, Ponting CP. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biol 2009;10(11):R124.
14. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 2006;442(7099):199-202.
15. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 2006;442 (7099):203-207.
16. Grivna ST, Beyret E, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20(13):1709-1714.
17. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science 2006;313(5785):363-367.
18. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, et al. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev 2006;20(13):1732-1743.
19. Seto AG, Kingston RE, Lau NC. The coming of age for Piwi proteins. Mol Cell 2007;26(5):603-609.
20. Mani SR, Juliano CE. Untangling the web: the diverse functions of the PIWI/piRNA pathway. Mol Reprod Dev 2013;80(8):632-664.
21. Luteijn MJ, Ketting RF. PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat Rev Genet 2013;14(8):523-534.
22. Weick EM, Miska EA. piRNAs: from biogenesis to function. Development 2014;141(18):3458-3471.
23. Lukic S, Chen K. Human piRNAs are under selection in Africans and repress transposable elements. Mol Biol Evol 2011;28(11):3061-3067.
24. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 2007;129(1):69-82.
25. Yin H, Lin, H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. Nature 2007;450(7167):304-308.
26. Coventry A, Kleitman DJ, Berger B. MSARI: multiple sequence alignments for statistical detection of RNA secondary structure. Proc Natl Acad Sci USA 2004;101 (33):12102-12107.
27. Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, et al. Identification and

classification of conserved RNA secondary structures in the human genome. PLoS Comput Biol 2006;2(4):e33.

28. Rivas E, Eddy SR. Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinformatics 2001;2:8.
29. Washietl S, Hofacker IL, Stadler PF. Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci USA 2005;102(7):2454-2459.
30. Bao M, Cervantes Cervantes M, Zhong L, Wang JT. Searching for non-coding RNAs in genomic sequences using ncRNAscout. Genomics Proteomics Bioinformatics 2012;10(2):114-121.
31. Lertampaiporn S, Thammarongtham C, Nukoolkit C, Kaewkamnerdpong B, Ruengjitchatchawalya M. Heterogeneous ensemble approach with discriminative features and modified-SMOTEbagging for pre-miRNA classification. Nucleic Acids Res 2013;41(1):e21.
32. Raasch P, Schmitz U, Patenge N, Vera J, Kreikemeyer B, Wolkenhauer O. Non-coding RNA detection methods combined to improve usability, reproducibility and precision. BMC Bioinformatics 2010;11:491.
33. Saetrom P, Sneve R, Kristiansen KI, Snøve O Jr, Grünfeld T, Rognes T, et al. Predicting non-coding RNA genes in Escherichia coli with boosted genetic programming. Nucleic Acids Res 2005;33(10):3263-3270.
34. Salari R, Aksay C, Karakoc E, Unrau PJ, Hajirasouliha I, Sahinalp SC. SmyRNA: a novel Ab initio ncRNA gene finder. PLoS One 2009;4(5):e5433.
35. Tran TT, Zhou F, Marshburn S, Stead M, Kushner SR, Xu Y. De novo computational prediction of non-coding RNA genes in prokaryotic genomes. Bioinformatics 2009;25(22):2897-2905.
36. Wang C, Ding C, Meraz RF, Holbrook SR. PSoL: a positive sample only learning algorithm for finding non-coding RNA genes. Bioinformatics 2006;22(21):2590-2596.
37. Washietl S, Findeiss S, Muller SA, Kalkhof S, von Bergen M, Hofacker IL, et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. RNA 2011;17(4):578-594.
38. Brayet J, Zehraoui F, Jeanson-Leh L, Israeli D, Tahi F. Towards a piRNA prediction using multiple kernel fusion and support vector machine. Bioinformatics 2014;30 (17):i364-370.
39. Wang K, Liang C, Liu J, Xiao H, Huang S, Xu J, et al. Prediction of piRNAs using transposon interaction and a support vector machine. BMC Bioinformatics 2014;15 (1):419.
40. Betel D, Sheridan R, Marks DS, Sander C. Computational analysis of mouse piRNA sequence and biogenesis. PLoS Comput Biol 2007;3(11):e222.
41. Zhang Y, Wang X, Kang L. A k-mer scheme to predict piRNAs and characterize locust piRNAs. Bioinformatics 2011;27(6):771-776.
42. Sinha S, Vasulu T, De RK. Performance and evaluation of MicroRNA gene identification tools. J Proteomics Bioinform 2009;2:336-343.
43. Liu X, Ding J, Gong F. PiRNA identification based on motif discovery. Mol Biosyst 2014;10(12):3075-3080.
44. Sai Lakshmi S, Agrawal S. PiRNABank: a web resource on classified and clustered Piwi-interacting RNAs. Nucleic Acids Res 2008;36(Database issue):D173-177.
45. Zahiri J, Yaghoubi O, Mohammad-Noori M, Ebrahimpour R, Masoudi-Nejad A. PPIevo: protein-protein interaction prediction from PSSM based evolutionary information. Genomics 2013;102(4):237-242.
46. Zahiri J, Mohammad-Noori M, Ebrahimpour R, Saadat S, Bozorgmehr JH, Goldberg T, et al. LocFuse: human protein-protein interaction prediction via classifier fusion using protein localization information. Genomics 2014; 104(6 Pt B):496-503.
47. Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Res 2003;31(13):3429-3431.
48. Zahiri J, Bozorgmehr JH, Masoudi-Nejad A. Computational Prediction of Protein-Protein Interaction Networks: Algo-rithms and Resources. Curr Genomics 2013;14(6): 397-414.
49. Li L, Jiang W, Li X, Moser KL, Guo Z, Du L, et al. A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset. Genomics 2005;85(1):16-23.
50. Gönen M, Alpaydın E. Multiple kernel learning algorithms. J Machine Learning Res 2011;12:2211-2268.
51. Wang JY. Application of support vector machines in bioinformatics. [master's thesis]. Taiwan: National Taiwan University; 2002. 65 p.
52. Wedeles CJ, Wu MZ, Claycomb JM. Protection of germline gene expression by the C. elegans Argonaute CSR-1. Dev Cell 2013;27(6):664-671.
53. Weick EM, Sarkies P, Silva N, Chen RA, Moss SM, Cording AC, et al. PRDE-1 is a nuclear factor essential for the biogenesis of Ruby motif-dependent piRNAs in C. elegans. Genes Dev 2014;28(7):783-796.
54. Wang Y, Chen X, Jiang W, Li L, Li W, Yang L, et al. Predicting human microRNA precursors based on an optimized feature subset generated by GA-SVM. Genomics 2011;98(2):73-78.