Structural Bioinformatics

De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures

Kwang Loong Stanley Ng[1,2] , Santosh K. Mishra[1,2]

[1]Bioinformatics Institute, Singapore 138671 and [2]NUS Graduate School for Integrative Sciences & Engineering, Centre for Life Sciences, Singapore 117456

1   BIOGENESIS OF MATURE MICRORNAS

In the prevailing biogenesis model of miRNAs maturation (Bartel 2004; Kim 2005), >1000-nts primary transcripts (*pri-miRs*) originate from the intergenic or intronic regions of annotated (non-)protein-coding transcription units (Rodriguez *et al.*, 2004). The *pri-miRs* are cleaved by the nuclear RNase III endonuclease Drosha/Pasha complex, yielding ~70–120-nts precursor transcripts (*pre-miRs*) with 5' phosphate and ~2-nts 3' overhang. These *pre-miRs* exhibiting characteristic imperfect and extended RNA stem-loop structures are actively exported by the cargo transporter Exportin-5 in a Ran-GTP dependent manner into the cytoplasm (Bohnsack *et al.*, 2004; Yi *et al.*, 2003; Zeng and Cullen 2004). The *pre-miRs* are further excised by another RNase III endonuclease Dicer into an intermediate duplex *miR:miR**, a ~21–23-nts asymmetric mature miRNA duplex. The *miR:miR** is recruited by a ribonucleoprotein RNA-Induced Silencing Complex (RISC) (Gregory *et al.*, 2005; Maniataki and Mourelatos 2005; Rivas *et al.*, 2005; Tang 2005). The strand *miR* with weaker hydrogen binding survives as the mature miRNA, which is preferentially loaded into RISC. miRNA-directed posttranscriptional silencing of target genes occurs by mRNA degradation (Brennecke *et al.*, 2005), or translational arrest of protein synthesis (Doench and Sharp 2004), or mRNA deadenylation (Wu *et al.*, 2006).

2   MATERIALS AND METHODS

2.1   Biologically relevant datasets

8,494 pseudo hairpin sequences. We analyzed 8,494 pseudo hairpin sequences that were extracted from the protein-coding regions (CDSs) according to the UCSC refGene annotation tables (Karolchik *et al.*, 2003) and human RefSeq genes (Pruitt and Maglott 2001). As wrongly assumed 'negative samples' can distort the decision boundary of SVM in an unpredictable and/or significant manner, special requirements were imposed on the selection of genomic inverted repeats. First, they must originate from genomic regions that do not undergo any known experimentally validated alternative splicing (AS) events, as described previously (Xue *et al.*, 2005). This criterion ensures that they do not encode genuine human *pre-miRs*. Second, they are analogous to genuine human *pre-miRs* by displaying similar distribution in terms of their length ~90-nts, hairpin structures with stem ≥8-bps including the GU wobble pairs, and MFEs ≤ -15 kcal/mol. In addition, they fold without multiple loops in their RNA structures as verified by the RNAfold program (Hofacker 2003).

2,241 non-redundant pre-miR sequences. We retrieved 4,028 annotated *pre-miR* sequences spanning across 45 species from miRBase Registry Database release 8.2, July 2006 (Griffiths-Jones *et al.*, 2006). As strong sequence homologies existed among *pre-miRs* both within a single and between different specie(s), the homologs of the training *pre-miRs* must be excluded from the testing and inde-

pendent evaluation sets. The original dataset was filtered to 90% identity using a greedy incremental clustering algorithm (Li and Godzik 2006). Briefly, all the sequences were first sorted in order of decreasing length and the longest one became the representative of the first cluster. Each remaining sequence was compared with the existing representatives and grouped into their cluster if the similarity with any representative was above a given threshold (default value is 0.9), else that sequence became the representative of a new cluster. Consequently, 2,241 non-redundant *pre-miRs* spanning 41 species (categorized into arthropoda, nematoda, verterbrata, viridiplantae, and viruses) served as the reference positive set; none of the sequences from *Gorilla gorilla*, *Macaca nemestrina*, *Pan paniscus*, and *Pongo pygmaeus* was retained.

12,387 functional prokaryotic and eukaryotic ncRNA sequences. We retrieved all curated seed ncRNA sequences from Rfam repository release 7.0, March 2005 (Griffiths-Jones *et al.*, 2005). After removing 46 types of *pre-miRs*, 12,387 functional prokaryotic and eukaryotic ncRNAs spanning 457 types (categorized into 16 classes) served as the negative non *pre-miR* dataset. Functional ncRNAs have similar length distribution to the known *pre-miRs*, and can fold with hairpin(s) or stem-loop(s) (Eddy 2001; Storz 2002; Svoboda and Cara 2006).

31 mRNA sequences. We investigated 31 mRNA sequences that tend to fold into complex RNA structures with extremely negative MFEs (Freyhult *et al.*, 2005). They were randomly selected from GeneBank DNA database (Benson *et al.*, 2005) with the following GenBank accession numbers: NM_001005151.1, NM_001003967.1, NM_177233.4, AY675236.1, NM_001004202.1, NM_178539.2, AB164385.1, AY555511.1, AB189435.1, NM_178307.2, NM_001003-966.1, NM_205498.1, NM_013564.3, Z81556.1, NM_131070.2, X56279.1, AK045412.1, AF452886.1, BC049701.1, BC050086.1, NM_172343.1, AY182163.1, BC072691.1, CV127341.1, NC_004671.1, X00910.1, AY226143.1, AJ621386, CV122154.1, X68284, and CV199185.1.

2.2   Feature vector

Adjusted base pairing propensity, *dP* measures the total number of base pairs present in the RNA secondary structure *S* (Schultes *et al.*, 1999) divided by the length *L* in nucleotides. It removes the bias that a long sequence tends to have more base pairs. *dP* ranges [0.0, 0.5], 0.0 for no base pair interactions and 0.5 for maximum of *L/2* base pairs.

Adjusted minimum free energy of folding, *dG* measures the thermodynamic stability of RNA structure *S* i.e., the lowest MFE for the most favorable conformation, divided by the length *L* in nucleotides (Freyhult *et al.*, 2005). It removes the bias that a long sequence tends to have lower negative MFE (Seffens and Digby 1999).

MFE Index 1, $MFEI_1$ is the ratio of *dG* and %*G+C* content (Zhang *et al.*, 2006).

Adjusted shannon entropy, *dQ* in Eq. (1), characterizes the base pairing probability distribution (BPPD) in a RNA structure *S* as a chaotic dynamical system (Freyhult *et al.*, 2005; Huynen *et al.*, 1997; Schultes *et al.*, 1999). Low values of *dQ* correspond to BPPD that are dominated by single, a few, or by the absence of base pairings. These bases are better predicted than those having multiple alternative states.

$$dQ = -\frac{1}{L}\sum_{i<j} p_{ij}\log_2(p_{ij}), \quad p_{ij} = \sum_{S_a \in S(\mathbf{x})} P(S_a)d_{ij}^a. \quad (1)$$

Here, p_{ij} denotes the probability of bases i and j pair, computed from the McCaskill's algorithm (J.S.McCaskill 1990); $d_{ij}^a = 1$ if i and j pair, 0 otherwise. RNA molecules exist *in vivo* as an ensemble of secondary structures $S_\alpha \in S(\mathbf{x})$ with the Boltzmann distribution probability $P(S_\alpha)$ (Mathews 2004).

Adjusted base pair distance, dD in Eq. (2), is the base pair distance for all pairs of structures S_α and S_β inferred from sequence $\mathbf{s}$ (Freyhult *et al.*, 2005; Moulton *et al.*, 2000).

$$dD = \frac{1}{2L}\sum_{S_a, S_b \in \mathbf{S}(s)} P(S_a)P(S_b)d_{BP}(S_a, S_b) = \frac{1}{L}\sum_{i<j} p_{ij}(1-p_{ij}). \quad (2)$$

Here, the number of base pairs not shared by them is given by $d_{BP}(S_a, S_\beta) = |S_\alpha \cup S_\beta| - |S_\alpha \cap S_\beta| = \sum_{i<j}(d_{ij}^a + d_{ij}^b - 2d_{ij}^a d_{ij}^b)$. The number of base pairs in S_α is $|S_\alpha| = \sum_{i<j} d_{ij}^a$. Definitions of p_{ij} and d_{ij}^a follow those of dQ in Eq. (1).

Second (or the Fiedler) eigenvalue, dF in Eq. (3), measures the compactness of a tree-graph $G = (V, E)$ (Fera *et al.*, 2004; Gan *et al.*, 2004). At the coarsest scale, each vertex $v \in V$ represents a bulge loop, hairpin loop, internal loop, the 5' and 3' unpaired termini, or the multi-branch loop; each edge $e \in E$ denotes a RNA stem. dF is computed from the Laplacian matrix $\mathbf{L}(G)$, a mathematical representation of the tree-graph G. dF can be used as a similarity measure among a collection of RNA secondary structures.

$$\mathbf{L}(G)\mathbf{X} = l\,\mathbf{X} \Leftrightarrow dF = FidlerEigen[\mathbf{L}(G)]. \quad (3)$$

MFE Index 2, $MFEI_2$ is the ratio of dG and the number of stems S, which are structural motifs containing more than three contiguous base pairs.

Normalized feature vectors. The Z-score $Z(\mathbf{s}_n)$ in Eq. (4) normalizes the feature $S(\mathbf{s}_n)$ of nth native RNA sequence $\mathbf{s}_n$ in terms of the number of standard deviations by which $S(\mathbf{s}_n)$ differs from the mean of inferred $R = 10^4$ random RNA sequences $\mathbf{r}_n$. $S(\mathbf{s}_n)$ can be dG, dP, dQ, dD, and dF; corresponding Z-scores are denoted as zG, zP, zQ, zD, and zF.

$$Z(\mathbf{s}_n) = \frac{S(\mathbf{s}_n) - m_n}{s_n}, \quad s_n^2 = \frac{1}{R-1}\sum_{i=1}^{R}[S_i(\mathbf{r}_n) - m_n]^2. \quad (4)$$

Here, $S_i(\mathbf{r}_n)$ is the computed feature for the ith random sequence of $\mathbf{r}_n$, μ_n and σ_n are the sample mean and the standard deviation of the feature $S(\mathbf{s}_n)$. The entire set of R random sequences $\mathbf{r}_n$ is synthesized by the "Altschul-Erikson algorithm" (Altschul and Erickson 1985), a form of dinucleotide shuffling. Briefly, it shuffles $\mathbf{s}_n$ while preserving exactly both the mono- and di-nucleotide frequencies. The $\mathbf{r}_n$ shares the same first and last nucleotides as $\mathbf{s}_n$. The order of the shuffled nucleotides is 'less random' due to fewer possible dinucleotide-preserving permutations.

2.3 Statistical tests and performance evaluation

F-scores of features. The "quality" of the ith feature is described commonly by the F-scores F1 (Dror *et al.*, 2005) and F2 (Chen and Lin 2006) in Eq. (5). The larger their values for the ith feature, the more likely this feature possesses discriminative importance/power.

$$F1 = \frac{|m_i^+ - m_i^-|}{|s_i^+ + s_i^-|}, \quad F2 = \frac{(m_i^+ - \bar{m}_i)^2 + (m_i^- - \bar{m}_i)^2}{(s_i^+)^2 + (s_i^-)^2}. \quad (5)$$

Here m_i^+/m_i^- and s_i^+/s_i^- denote the means and standard deviations of the positive (+) and negative (−) training datasets, correspondingly. The numerator and denominator describe the discrimination between the two classes, and that within each of the two classes.

Performance measures. SE (Sensitivity or recall), SP (Specificity), ACC (Accuracy), Fm (F-measure) (Liu *et al.*, 2006), and MCC (Matthew's Correlation Coefficient) (Bhasin *et al.*, 2006) are defined in Eq. (6). All metrics (except MCC) range [0.0, 1.0]; closer to 1.0 implies better scores, and *vice-versa*. MCC ranges [-1.0, 1.0]; -1.0, 0.0, and 1.0 indicate worst possible, perfectly random, and best possible classification, respectively. Unlike ACC, Fm and MCC account for the unbalanced datasets.

$$SE = \frac{TP}{TP+FN}, \quad SP = \frac{TN}{TN+FP}, \quad ACC = \frac{TP+TN}{TP+TN+FP+FN},$$

$$Fm = \frac{2(SP \bullet PPV)}{SP+PPV} \text{ where } PPV \text{ (Positive Predictive Value)} = \frac{TP}{TP+FP}, \quad (6)$$

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

Here *TP*, *FN*, *FP*, and *TN* denote the number of true/false *pre-miRs* detected/missed, correspondingly. The "quality" of a binary classification, denoted simply as ROC score, is measured by the normalized area under the Receiver Operating Characteristic curve (ROC) that plots the trade-off between the *SE* and *FPR* (False-positive rate = 1 − *SP*) across all possible classification thresholds (Lasko *et al.*, 2005). ROC score ranges [0.5, 1.0]; closer to 0.5 (about the upward diagonal) and to 1.0 (along the left-top boundary) signify a totally random and a perfect classifier (Lasko *et al.*, 2005). In this work, the ROC score was computed by the trapezoidal rule (Kestler 2001).

Benchmarking miPred. Both *3SVM* (Xue *et al.*, 2005) and Naïve Bayesian Classifier (*NBC*) served as independent baseline models to benchmark the performance improvements or deterioration (if any) of *miPred*. The original *3SVM* was previously trained on 163 human *pre-miRs* and 168 pseudo hairpins using the older libSVM 2.36 with the "-b 1" option disabled. Here, *3SVM* was trained on randomly selected 200 human *pre-miRs* and 400 pseudo hairpins using the latest libSVM 2.82 (the "-b 1" option is enabled) and the optimal hyperparameter pair (C, γ). *3SVM* was applied to the testing and independent evaluation datasets with "*svm-predict* -b 1". The Bayes Classifier Induction (*bci*) version 2.14, a free implementation of *NBC* (http://fuzzy.cs.unimagdeburg.de/~borgelt/bayes.html), was used for training and testing with the exact samples and attributes employed by *3SVM* and *miPred*; denoted as *3SVM-NBC* and *miPred-NBC*. For training, "*bci* -L1" yielded better classification results than the default "-L0". In theory, *NBC* seeks to maximize the probability $P(X|C) = P(f_1, f_2, \dots f_n)|C)$ such that the sample X belongs to one of the binary classes $C = (T, F)$.

REFERENCES

Altschul,S.F. and Erickson,B.W. (1985) Significance of nucleotide sequence alignments: a method for random sequence permutation that preserves dinucleotide and codon usage. *Mol. Biol. Evol.*, **2**, 526-538.

Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.

Benson,D.A. *et al.* (2005) GenBank. *Nucleic Acids Res.*, **33**, D34-D38.

Bhasin,M. *et al.* (2006) Recognition and Classification of Histones Using Support Vector Machine. *J. Comp. Biol.*, **13**, 102-112.

Bohnsack,M.T. *et al.* (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185-191.

Brennecke,J. *et al.* (2005) Principles of MicroRNA-Target Recognition. *PLoS Biol.*, **3**, e85.

Chen,Y.-W. and Lin,C.-J. (2006) Combining SVMs with various feature selection strategies. In Guyon,I., Gunn,S., Nikravesh,M.and Zadeh,L. (eds), *Feature extraction, foundations and applications.* Springer, pp. 315-323.

Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504-511.

Dror,G. *et al.* (2005) Accurate identification of alternatively spliced exons using support vector machine. *Bioinformatics*, **21**, 897-901.

Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919-929.

Fera,D. *et al.* (2004) RAG: RNA-As-Graphs web resource. *BMC Bioinformatics*, **5**, 88.

Freyhult,E. *et al.* (2005) A comparison of RNA folding measures. *BMC Bioinformatics*, **6**, 241.

Gan,H.H. *et al.* (2004) RAG: RNA-As-Graphs database--concepts, analysis, and features. *Bioinformatics*, **20**, 1285-1291.

Gregory,R.I. *et al.* (2005) Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell*, **123**, 631-640.

Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140-D144.

Griffiths-Jones,S. *et al.* (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**, D121-D124.

Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429-3431.

Huynen,M. *et al.* (1997) Assessing the reliability of RNA folding using statistical mechanics. *J. Mol. Biol.*, **267**, 1104-1112.

J.S.McCaskill (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105-1119.

Karolchik,D. *et al*. (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51-54.

Kestler,H.A. (2001) ROC with confidence - a Perl program for receiver operator characteristic curves. *Comput. Methods Programs Biomed.*, **64**, 133-136.

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

Lasko,T.A. *et al*. (2005) The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inform.*, **38**, 404-415.

Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.

Liu,J. *et al*. (2006) Distinguishing Protein-Coding from Non-Coding RNAs through Support Vector Machines. *PLoS Genet.*, **2**, e29.

Maniataki,E. and Mourelatos,Z. (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.*, **19**, 2979-2990.

Mathews,D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**, 1178-1190.

Moulton,V. *et al*. (2000) Metrics on RNA Secondary Structures. *J. Comp. Biol.*, **7**, 277-292.

Pruitt,K.D. and Maglott,D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137-140.

Rivas,F.V. *et al*. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.*, **12**, 340-349.

Rodriguez,A. *et al*. (2004) Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Res.*, **14**, 1902-1910.

Schultes,E.A. *et al*. (1999) Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.*, **49**, 76-83.

Seffens,W. and Digby,D. (1999) mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res.*, **27**, 1578-1584.

Storz,G. (2002) An Expanding Universe of Noncoding RNAs. *Science*, **296**, 1260-1263.

Svoboda,P. and Cara,A.D. (2006) Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.*, **63**, 901-908.

Tang,G. (2005) siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.*, **30**, 106-114.

Wu,L. *et al*. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci. USA*, **103**, 4034-4039.

Xue,C. *et al*. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics*, **6**, 310.

Yi,R. *et al*. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.*, **17**, 3011-3016.

Zeng,Y. and Cullen,B.R. (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.*, **32**, 4776-4785.

Zhang,B. *et al*. (2006) Evidence that miRNAs are different from other RNAs. *Cell. Mol. Life Sci.*, **63**, 246-254.

Table S1. The prediction performances of *miPred*, *miPred-NBC*, *3SVM*, and *3SVM-NBC* evaluated on the *pre-miR* datasets TR-H (200 human *pre-miRs* and 400 pseudo hairpins), TE-H (remaining 123 human *pre-miRs* and 246 pseudo hairpins), and IE-NH (1,918 *pre-miRs* across 40 non-human species and 3,836 pseudo hairpins).

miPred

Species	Genus	TP	FN	P	FP	TN	N	%SE	%SP	%FPR	%ACC
Homo sapiens	*Vertebrata*	176	24	200	10	390	400	88.00	97.50	2.50	94.33
Homo sapiens	*Vertebrata*	104	19	123	5	241	246	84.55	97.97	2.03	93.50
Anopheles gambiae	*Arthropoda*	37	1	38	1	75	76	97.37	98.68	1.32	98.25
Apis mellifera	*Arthropoda*	25	0	25	2	48	50	100	96	4	97.33
Arabidopsis thaliana	*Viridiplantae*	101	7	108	2	214	216	93.52	99.07	0.93	97.22
Ateles geoffroyi	*Vertebrata*	2	0	2	0	4	4	100	100	0	100
Bos taurus	*Vertebrata*	8	3	11	2	20	22	72.73	90.91	9.09	84.85
Caenorhabditis briggsae	*Nematoda*	72	4	76	1	151	152	94.74	99.34	0.66	97.81
Caenorhabditis elegans	*Nematoda*	96	17	113	7	219	226	84.96	96.9	3.1	92.92
Canis familiaris	*Vertebrata*	3	0	3	0	6	6	100	100	0	100
Danio rerio	*Vertebrata*	235	11	246	19	473	492	95.53	96.14	3.86	95.94
Drosophila melanogaster	*Arthropoda*	67	6	73	4	142	146	91.78	97.26	2.74	95.43
Drosophila pseudoobscura	*Arthropoda*	32	3	35	1	69	70	91.43	98.57	1.43	96.19
Epstein Barr (EBV)	*Viruses*	22	0	22	2	42	44	100	95.45	4.55	96.97
Fugu rubripes	*Vertebrata*	68	2	70	2	138	140	97.14	98.57	1.43	98.1
Gallus gallus	*Vertebrata*	87	5	92	4	180	184	94.57	97.83	2.17	96.74
Glycine max	*Viridiplantae*	20	1	21	0	42	42	95.24	100	0	98.41
Herpes Simplex (HSV)	*Viruses*	1	0	1	0	2	2	100	100	0	100
Human cytomegalovirus (HCMV)	*Viruses*	11	0	11	1	21	22	100	95.45	4.55	96.97
Kaposi sarcoma-associated herpesvirus (KSHV)	*Viruses*	11	1	12	0	24	24	91.67	100	0	97.22
Lagothrix lagotricha	*Vertebrata*	1	1	2	0	4	4	50	100	0	83.33
Lemur catta	*Vertebrata*	2	1	3	0	6	6	66.67	100	0	88.89
Macaca mulatta	*Vertebrata*	1	1	2	0	4	4	50	100	0	83.33
Medicago truncatula	*Viridiplantae*	17	1	18	0	36	36	94.44	100	0	98.15
Mouse γ-herpesvirus (MGHV68)	*Viruses*	8	1	9	1	17	18	88.89	94.44	5.56	92.59
Mus musculus	*Vertebrata*	166	33	199	9	389	398	83.42	97.74	2.26	92.96
Oryza sativa	*Viridiplantae*	140	12	152	4	300	304	92.11	98.68	1.32	96.49
Ovis aries	*Vertebrata*	2	0	2	0	4	4	100	100	0	100
Pan troglodytes	*Vertebrata*	2	1	3	1	5	6	66.67	83.33	16.67	77.78
Physcomitrella patens	*Viridiplantae*	17	0	17	0	34	34	100	100	0	100
Populus trichocarpa	*Viridiplantae*	144	13	157	13	301	314	91.72	95.86	4.14	94.48
Rattus norvegicus	*Vertebrata*	56	12	68	10	126	136	82.35	92.65	7.35	89.22
Rhesus lymphocryptovirus	*Viruses*	16	0	16	2	30	32	100	93.75	6.25	95.83
Saccharum officinarum	*Viridiplantae*	3	1	4	0	8	8	75	100	0	91.67
Saguinus labiatus	*Vertebrata*	1	1	2	0	4	4	50	100	0	83.33
Simian virus (SV40)	*Viruses*	1	0	1	0	2	2	100	100	0	100
Sorghum bicolor	*Viridiplantae*	48	2	50	2	98	100	96	98	2	97.33
Sus scrofa	*Vertebrata*	1	1	2	0	4	4	50	100	0	83.33
Tetraodon nigroviridis	*Vertebrata*	40	3	43	0	86	86	93.02	100	0	97.67
Xenopus laevis	*Vertebrata*	4	1	5	0	10	10	80	100	0	93.33
Xenopus tropicalis	*Vertebrata*	119	6	125	4	246	250	95.2	98.4	1.6	97.33
Zea mays	*Viridiplantae*	79	0	79	5	153	158	100	96.84	3.16	97.89
Total samples		2046	195	2241	114	4368	4482				

(*Species*) Row 1 (TR-H), row 2 (TE-H), and the remaining rows 3−43 (IE-NH). *TP* (real *pre-miRs* detected), *FN* (real *pre-miRs* missed), *P* (real *pre-miRs*), *FP* (pseudo hairpins detected), *TN* (pseudo hairpins missed), *N* (pseudo hairpins), *%SE* (Sensitivity), *%SP* (Specificity), *%FPR* (False-positive rate), and *%ACC* (Accuracy).

miPred-NBC

Species	Genus	TP	FN	P	FP	TN	N	%SE	%SP	%FPR	%ACC
Homo sapiens	*Vertebrata*	200	0	200	0	400	400	100.00	100.00	0.00	100.00
Homo sapiens	*Vertebrata*	46	77	123	36	210	246	37.40	85.37	14.63	69.38
Anopheles gambiae	*Arthropoda*	12	26	38	7	69	76	31.58	90.79	9.21	71.05
Apis mellifera	*Arthropoda*	6	19	25	6	44	50	24.00	88.00	12.00	66.67
Arabidopsis thaliana	*Viridiplantae*	20	88	108	27	189	216	18.52	87.50	12.50	64.51
Ateles geoffroyi	*Vertebrata*	0	2	2	1	3	4	0.00	75.00	25.00	50.00
Bos taurus	*Vertebrata*	1	10	11	2	20	22	9.09	90.91	9.09	63.64
Caenorhabditis briggsae	*Nematoda*	27	49	76	20	132	152	35.53	86.84	13.16	69.74
Caenorhabditis elegans	*Nematoda*	51	62	113	26	200	226	45.13	88.50	11.50	74.04
Canis familiaris	*Vertebrata*	0	3	3	1	5	6	0.00	83.33	16.67	55.56
Danio rerio	*Vertebrata*	71	175	246	62	430	492	28.86	87.40	12.60	67.89
Drosophila melanogaster	*Arthropoda*	21	52	73	17	129	146	28.77	88.36	11.64	68.49
Drosophila pseudoobscura	*Arthropoda*	12	23	35	5	65	70	34.29	92.86	7.14	73.33
Epstein Barr (EBV)	*Viruses*	6	16	22	4	40	44	27.27	90.91	9.09	69.70
Fugu rubripes	*Vertebrata*	10	60	70	18	122	140	14.29	87.14	12.86	62.86
Gallus gallus	*Vertebrata*	24	68	92	22	162	184	26.09	88.04	11.96	67.39
Glycine max	*Viridiplantae*	2	19	21	3	39	42	9.52	92.86	7.14	65.08
Herpes Simplex (HSV)	*Viruses*	0	1	1	1	1	2	0.00	50.00	50.00	33.33
Human cytomegalovirus (HCMV)	*Viruses*	0	11	11	5	17	22	0.00	77.27	22.73	51.52
Kaposi sarcoma-associated herpesvirus (KSHV)	*Viruses*	1	11	12	5	19	24	8.33	79.17	20.83	55.56
Lagothrix lagotricha	*Vertebrata*	0	2	2	0	4	4	0.00	100.00	0.00	66.67
Lemur catta	*Vertebrata*	0	3	3	2	4	6	0.00	66.67	33.33	44.44
Macaca mulatta	*Vertebrata*	0	2	2	0	4	4	0.00	100.00	0.00	66.67
Medicago truncatula	*Viridiplantae*	4	14	18	4	32	36	22.22	88.89	11.11	66.67
Mouse γ-herpesvirus (MGHV68)	*Viruses*	2	7	9	3	15	18	22.22	83.33	16.67	62.96
Mus musculus	*Vertebrata*	37	162	199	52	346	398	18.59	86.93	13.07	64.15
Oryza sativa	*Viridiplantae*	35	117	152	37	267	304	23.03	87.83	12.17	66.23
Ovis aries	*Vertebrata*	0	2	2	0	4	4	0.00	100.00	0.00	66.67
Pan troglodytes	*Vertebrata*	2	1	3	1	5	6	66.67	83.33	16.67	77.78
Physcomitrella patens	*Viridiplantae*	3	14	17	3	31	34	17.65	91.18	8.82	66.67
Populus trichocarpa	*Viridiplantae*	33	124	157	41	273	314	21.02	86.94	13.06	64.97
Rattus norvegicus	*Vertebrata*	23	45	68	11	125	136	33.82	91.91	8.09	72.55
Rhesus lymphocryptovirus	*Viruses*	5	11	16	2	30	32	31.25	93.75	6.25	72.92
Saccharum officinarum	*Viridiplantae*	1	3	4	0	8	8	25.00	100.00	0.00	75.00
Saguinus labiatus	*Vertebrata*	0	2	2	1	3	4	0.00	75.00	25.00	50.00
Simian virus (SV40)	*Viruses*	0	1	1	1	1	2	0.00	50.00	50.00	33.33
Sorghum bicolor	*Viridiplantae*	7	43	50	13	87	100	14.00	87.00	13.00	62.67
Sus scrofa	*Vertebrata*	0	2	2	1	3	4	0.00	75.00	25.00	50.00
Tetraodon nigroviridis	*Vertebrata*	9	34	43	9	77	86	20.93	89.53	10.47	66.67
Xenopus laevis	*Vertebrata*	2	3	5	4	6	10	40.00	60.00	40.00	53.33
Xenopus tropicalis	*Vertebrata*	35	90	125	33	217	250	28.00	86.80	13.20	67.20
Zea mays	*Viridiplantae*	16	63	79	18	140	158	20.25	88.61	11.39	65.82
Total samples		724	1517	2241	504	3978	4482				

(*Species*) Row 1 (TR-H), row 2 (TE-H), and the remaining rows 3−43 (IE-NH). *TP* (real *pre-miRs* detected), *FN* (real *pre-miRs* missed), *P* (real *pre-miRs*), *FP* (pseudo hairpins detected), *TN* (pseudo hairpins missed), *N* (pseudo hairpins), *%SE* (Sensitivity), *%SP* (Specificity), *%FPR* (False-positive rate), and *%ACC* (Accuracy).

3SVM[†]

Species	Genus	TP	FN	P	FP	TN	N	%SE	%SP	%FPR	%ACC
Homo sapiens	*Vertebrata*	172	28	200	12	388	400	86.00	97.00	3.00	93.33
Homo sapiens	*Vertebrata*	79	29	108	10	206	216	73.15	95.37	4.63	87.96
Anopheles gambiae	*Arthropoda*	33	4	37	1	73	74	89.19	98.65	1.35	95.50
Apis mellifera	*Arthropoda*	23	2	25	1	49	50	92.00	98.00	2.00	96.00
Arabidopsis thaliana	*Viridiplantae*	69	2	71	5	137	142	97.18	96.48	3.52	96.71
Ateles geoffroyi	*Vertebrata*	2	0	2	0	4	4	100.00	100.00	0.00	100.00
Bos taurus	*Vertebrata*	7	1	8	3	13	16	87.50	81.25	18.75	83.33
Caenorhabditis briggsae	*Nematoda*	68	2	70	6	134	140	97.14	95.71	4.29	96.19
Caenorhabditis elegans	*Nematoda*	94	13	107	4	210	214	87.85	98.13	1.87	94.70
Canis familiaris	*Vertebrata*	3	0	3	1	5	6	100.00	83.33	16.67	88.89
Danio rerio	*Vertebrata*	201	32	233	30	436	466	86.27	93.56	6.44	91.13
Drosophila melanogaster	*Arthropoda*	57	9	66	7	125	132	86.36	94.70	5.30	91.92
Drosophila pseudoobscura	*Arthropoda*	28	7	35	1	69	70	80.00	98.57	1.43	92.38
Epstein Barr (EBV)	*Viruses*	19	3	22	0	44	44	86.36	100.00	0.00	95.45
Fugu rubripes	*Vertebrata*	48	16	64	5	123	128	75.00	96.09	3.91	89.06
Gallus gallus	*Vertebrata*	73	14	87	4	170	174	83.91	97.70	2.30	93.10
Glycine max	*Viridiplantae*	16	0	16	0	32	32	100.00	100.00	0.00	100.00
Herpes Simplex (HSV)	*Viruses*	0	1	1	0	2	2	0.00	100.00	0.00	66.67
Human cytomegalovirus (HCMV)	*Viruses*	8	3	11	0	22	22	72.73	100.00	0.00	90.91
Kaposi sarcoma-associated herpesvirus (KSHV)	*Viruses*	4	8	12	0	24	24	33.33	100.00	0.00	77.78
Lagothrix lagotricha	*Vertebrata*	1	0	1	0	2	2	100.00	100.00	0.00	100.00
Lemur catta	*Vertebrata*	2	0	2	0	4	4	100.00	100.00	0.00	100.00
Macaca mulatta	*Vertebrata*	1	0	1	0	2	2	100.00	100.00	0.00	100.00
Medicago truncatula	*Viridiplantae*	15	0	15	2	28	30	100.00	93.33	6.67	95.56
Mouse γ-herpesvirus (MGHV68)	*Viruses*	5	4	9	1	17	18	55.56	94.44	5.56	81.48
Mus musculus	*Vertebrata*	145	41	186	5	367	372	77.96	98.66	1.34	91.76
Oryza sativa	*Viridiplantae*	106	9	115	11	219	230	92.17	95.22	4.78	94.20
Ovis aries	*Vertebrata*	1	0	1	0	2	2	100.00	100.00	0.00	100.00
Pan troglodytes	*Vertebrata*	2	1	3	0	6	6	66.67	100.00	0.00	88.89
Physcomitrella patens	*Viridiplantae*	14	0	14	0	28	28	100.00	100.00	0.00	100.00
Populus trichocarpa	*Viridiplantae*	106	15	121	12	230	242	87.60	95.04	4.96	92.56
Rattus norvegicus	*Vertebrata*	50	12	62	5	119	124	80.65	95.97	4.03	90.86
Rhesus lymphocryptovirus	*Viruses*	16	0	16	1	31	32	100.00	96.88	3.13	97.92
Saccharum officinarum	*Viridiplantae*	0	0	0	0	0	0	NaN	NaN	NaN	NaN
Saguinus labiatus	*Vertebrata*	0	1	1	0	2	2	0.00	100.00	0.00	66.67
Simian virus (SV40)	*Viruses*	1	0	1	0	2	2	100.00	100.00	0.00	100.00
Sorghum bicolor	*Viridiplantae*	33	2	35	2	68	70	94.29	97.14	2.86	96.19
Sus scrofa	*Vertebrata*	0	2	2	0	4	4	0.00	100.00	0.00	66.67
Tetraodon nigroviridis	*Vertebrata*	39	2	41	3	79	82	95.12	96.34	3.66	95.94
Xenopus laevis	*Vertebrata*	2	3	5	1	9	10	40.00	90.00	10.00	73.33
Xenopus tropicalis	*Vertebrata*	101	21	122	7	237	244	82.79	97.13	2.87	92.35
Zea mays	*Viridiplantae*	50	2	52	7	97	104	96.15	93.27	6.73	94.23
Total samples		1694	289	1983	147	3819	3966				

†, *3SVM* model was trained on 200 human *pre-miRs* and 400 pseudo hairpins randomly selected using the latest libSVM 2.82 (the "-b 1" option was enabled) and the optimal hyperparameter pair (*C*, *γ*). (*Species*) Row 1 (TR-H), row 2 (TE-H), and the remaining rows 3−43 (IE-NH). *TP* (real *pre-miRs* detected), *FN* (real *pre-miRs* missed), *P* (real *pre-miRs*), *FP* (pseudo hairpins detected), *TN* (pseudo hairpins missed), *N* (pseudo hairpins), *%SE* (Sensitivity), *%SP* (Specificity), *%FPR* (False-positive rate), and *%ACC* (Accuracy).

3SVM-NBC

Species	Genus	TP	FN	P	FP	TN	N	%SE	%SP	%FPR	%ACC
Homo sapiens	*Vertebrata*	196	4	200	13	387	400	98.00	96.75	3.25	97.17
Homo sapiens	*Vertebrata*	71	37	108	51	165	216	65.74	76.39	23.61	72.84
Anopheles gambiae	*Arthropoda*	27	10	37	18	56	74	72.97	75.68	24.32	74.77
Apis mellifera	*Arthropoda*	20	5	25	9	41	50	80.00	82.00	18.00	81.33
Arabidopsis thaliana	*Viridiplantae*	44	27	71	30	112	142	61.97	78.87	21.13	73.24
Ateles geoffroyi	*Vertebrata*	1	1	2	1	3	4	50.00	75.00	25.00	66.67
Bos taurus	*Vertebrata*	4	4	8	4	12	16	50.00	75.00	25.00	66.67
Caenorhabditis briggsae	*Nematoda*	52	18	70	23	117	140	74.29	83.57	16.43	80.48
Caenorhabditis elegans	*Nematoda*	87	20	107	39	175	214	81.31	81.78	18.22	81.62
Canis familiaris	*Vertebrata*	3	0	3	2	4	6	100.00	66.67	33.33	77.78
Danio rerio	*Vertebrata*	140	93	233	112	354	466	60.09	75.97	24.03	70.67
Drosophila melanogaster	*Arthropoda*	38	28	66	31	101	132	57.58	76.52	23.48	70.20
Drosophila pseudoobscura	*Arthropoda*	20	15	35	15	55	70	57.14	78.57	21.43	71.43
Epstein Barr (EBV)	*Viruses*	12	10	22	9	35	44	54.55	79.55	20.45	71.21
Fugu rubripes	*Vertebrata*	31	33	64	33	95	128	48.44	74.22	25.78	65.63
Gallus gallus	*Vertebrata*	48	39	87	44	130	174	55.17	74.71	25.29	68.20
Glycine max	*Viridiplantae*	5	11	16	5	27	32	31.25	84.38	15.63	66.67
Herpes Simplex (HSV)	*Viruses*	0	1	1	0	2	2	0.00	100.00	0.00	66.67
Human cytomegalovirus (HCMV)	*Viruses*	3	8	11	1	21	22	27.27	95.45	4.55	72.73
Kaposi sarcoma-associated herpesvirus (KSHV)	*Viruses*	2	10	12	5	19	24	16.67	79.17	20.83	58.33
Lagothrix lagotricha	*Vertebrata*	1	0	1	0	2	2	100.00	100.00	0.00	100.00
Lemur catta	*Vertebrata*	2	0	2	0	4	4	100.00	100.00	0.00	100.00
Macaca mulatta	*Vertebrata*	1	0	1	1	1	2	100.00	50.00	50.00	66.67
Medicago truncatula	*Viridiplantae*	8	7	15	6	24	30	53.33	80.00	20.00	71.11
Mouse γ-herpesvirus (MGHV68)	*Viruses*	4	5	9	7	11	18	44.44	61.11	38.89	55.56
Mus musculus	*Vertebrata*	110	76	186	83	289	372	59.14	77.69	22.31	71.51
Oryza sativa	*Viridiplantae*	73	42	115	56	174	230	63.48	75.65	24.35	71.59
Ovis aries	*Vertebrata*	1	0	1	1	1	2	100.00	50.00	50.00	66.67
Pan troglodytes	*Vertebrata*	2	1	3	1	5	6	66.67	83.33	16.67	77.78
Physcomitrella patens	*Viridiplantae*	7	7	14	3	25	28	50.00	89.29	10.71	76.19
Populus trichocarpa	*Viridiplantae*	73	48	121	52	190	242	60.33	78.51	21.49	72.45
Rattus norvegicus	*Vertebrata*	39	23	62	24	100	124	62.90	80.65	19.35	74.73
Rhesus lymphocryptovirus	*Viruses*	10	6	16	6	26	32	62.50	81.25	18.75	75.00
Saccharum officinarum	*Viridiplantae*	0	0	0	0	0	0	NaN	NaN	NaN	NaN
Saguinus labiatus	*Vertebrata*	0	1	1	0	2	2	0.00	100.00	0.00	66.67
Simian virus (SV40)	*Viruses*	1	0	1	0	2	2	100.00	100.00	0.00	100.00
Sorghum bicolor	*Viridiplantae*	19	16	35	14	56	70	54.29	80.00	20.00	71.43
Sus scrofa	*Vertebrata*	1	1	2	0	4	4	50.00	100.00	0.00	83.33
Tetraodon nigroviridis	*Vertebrata*	18	23	41	17	65	82	43.90	79.27	20.73	67.48
Xenopus laevis	*Vertebrata*	0	5	5	1	9	10	0.00	90.00	10.00	60.00
Xenopus tropicalis	*Vertebrata*	68	54	122	49	195	244	55.74	79.92	20.08	71.86
Zea mays	*Viridiplantae*	18	34	52	30	74	104	34.62	71.15	28.85	58.97
Total samples		1260	723	1983	796	3170	3966				

(*Species*) Row 1 (TR-H), row 2 (TE-H), and the remaining rows 3−43 (IE-NH). *TP* (real *pre-miRs* detected), *FN* (real *pre-miRs* missed), *P* (real *pre-miRs*), *FP* (pseudo hairpins detected), *TN* (pseudo hairpins missed), *N* (pseudo hairpins), *%SE* (Sensitivity), *%SP* (Specificity), *%FPR* (False-positive rate), and *%ACC* (Accuracy).

Table S2. The mean sensitivity and specificity of *miPred*, *miPred-NBC*, *3SVM*, and *3SVM-NBC* evaluated on the non-human *pre-miR* dataset IE-NH (1,918 *pre-miRs* across 40 non-human species and 3,836 pseudo hairpins) categorized by genus of *pre-miRs*.

Genus	No. of species	miPred		miPred-NBC		No. of excluded species	3SVM[‡]		3SVM-NBC	
		%SE	%SP	%SE	%SP		%SE	%SP	%SE	%SP
Arthropoda	4	95.14 ± 2.11	97.63 ± 0.63	29.66 ± 2.20	90.00 ± 1.14	0	86.89 ± 2.57	97.48 ± 0.94	66.92 ± 5.71	78.19 ± 1.41
Viridiplantae	9	93.11 ± 2.47	98.72 ± 0.51	19.02 ± 1.60	90.09 ± 1.40	1	95.92 ± 1.57	96.31 ± 0.93	51.16 ± 4.31	79.73 ± 1.92
Vertebrata[†]	18	79.29 ± 4.56	97.53 ± 1.05	15.91 ± 4.43	84.83 ± 2.60	0	76.44 ± 7.48	96.11 ± 1.35	61.23 ± 7.22	79.58 ± 3.53
Nematoda	2	89.85 ± 4.89	98.12 ± 1.22	40.33 ± 4.80	87.67 ± 0.83	0	92.50 ± 4.65	96.92 ± 1.21	77.80 ± 3.51	82.68 ± 0.90
Viruses	7	97.22 ± 1.81	97.01 ± 1.08	12.72 ± 5.23	74.92 ± 6.81	0	64.00 ± 14.04	98.76 ± 0.84	43.63 ± 12.49	85.22 ± 5.36

†, *Homo sapiens* is excluded. ‡, *3SVM* model was trained on 200 human *pre-miRs* and 400 pseudo hairpins randomly selected using the latest libSVM 2.82 (the "-b 1" option was enabled) and the optimal hyperparameter pair (C, γ). *%SE* (Sensitivity) and *%SP* (Specificity). Values are expressed as mean ± standard error.

Table S3. The prediction performances of *miPred*, *miPred-NBC*, *3SVM*, and *3SVM-NBC* evaluated on the non *pre-miR* datasets IE-NC (12,387 functional ncRNAs) and IE-M (31 mRNAs).

Accession	Type†	Class		miPred		miPred-NBC			3SVM‡		3SVM-NBC	
			N	TN	%SP	TN	%SP	N	TN	%SP	TN	%SP
RF00001	5S ribosomal RNA	Gene/rRNA	589	409	69.44	517	87.78	2	2	100.00	1	50.00
RF00002	5.8S ribosomal RNA	Gene/rRNA	63	59	93.65	59	93.65	1	1	100.00	1	100.00
RF00003	U1 spliceosomal RNA	Gene/snRNA/splicing	54	38	70.37	45	83.33	0	0	NaN	0	NaN
RF00004	U2 spliceosomal RNA	Gene/snRNA/splicing	73	8	10.96	53	72.60	0	0	NaN	0	NaN
RF00005	tRNA	Gene/tRNA	1114	953	85.55	969	86.98	158	150	94.94	142	89.87
RF00006	Vault RNA	Gene	9	5	55.56	8	88.89	3	3	100.00	1	33.33
RF00007	U12 minor spliceosomal RNA	Gene/snRNA/splicing	7	4	57.14	7	100.00	0	0	NaN	0	NaN
RF00008	Hammerhead ribozyme (type III)	Gene/ribozyme	84	61	72.62	68	80.95	1	1	100.00	1	100.00
RF00009	Nuclear RNase P	Gene/ribozyme	53	16	30.19	50	94.34	0	0	NaN	0	NaN
RF00010	Bacterial RNase P class A	Gene/ribozyme	236	77	32.63	203	86.02	0	0	NaN	0	NaN
RF00011	Bacterial RNase P class B	Gene/ribozyme	30	12	40.00	28	93.33	0	0	NaN	0	NaN
RF00012	U3 small nucleolar RNA	Gene/snRNA/guide/C/D-box	21	10	47.62	18	85.71	0	0	NaN	0	NaN
RF00013	6S / SsrS RNA	Gene	7	1	14.29	6	85.71	2	0	0.00	1	50.00
RF00014	DsrA RNA	Gene/sRNA	3	0	0.00	2	66.67	0	0	NaN	0	NaN
RF00015	U4 spliceosomal RNA	Gene/snRNA/splicing	25	21	84.00	25	100.00	1	1	100.00	1	100.00
RF00016	U14 small nucleolar RNA	Gene/snRNA/guide/C/D-box	18	17	94.44	16	88.89	2	2	100.00	2	100.00
RF00017	Eukaryotic type signal recognition particle RNA	Gene	70	3	4.29	61	87.14	0	0	NaN	0	NaN
RF00018	CsrB/RsmB RNA family	Gene/sRNA	9	9	100.00	8	88.89	0	0	NaN	0	NaN
RF00019	Y RNA	Gene	15	9	60.00	12	80.00	5	5	100.00	2	40.00
RF00020	U5 spliceosomal RNA	Gene/snRNA/splicing	32	12	37.50	26	81.25	0	0	NaN	0	NaN
RF00021	Spot 42 RNA	Gene/sRNA	8	0	0.00	8	100.00	0	0	NaN	0	NaN
RF00022	GcvB RNA	Gene/sRNA	5	3	60.00	5	100.00	0	0	NaN	0	NaN
RF00023	tmRNA	Gene	87	53	60.92	79	90.80	0	0	NaN	0	NaN
RF00024	Vertebrate telomerase RNA	Gene	35	10	28.57	31	88.57	0	0	NaN	0	NaN
RF00025	Ciliate telomerase RNA	Gene	16	13	81.25	12	75.00	0	0	NaN	0	NaN
RF00026	U6 spliceosomal RNA	Gene/snRNA/splicing	53	52	98.11	48	90.57	0	0	NaN	0	NaN
RF00028	Group I catalytic intron	Intron	30	15	50.00	29	96.67	0	0	NaN	0	NaN
RF00029	Group II catalytic intron	Intron	116	37	31.90	89	76.72	0	0	NaN	0	NaN
RF00030	RNase MRP	Gene/ribozyme	26	9	34.62	25	96.15	0	0	NaN	0	NaN
RF00031	Selenocysteine insertion sequence	Cis-reg	64	52	81.25	50	78.13	56	56	100.00	50	89.29
RF00032	Histone 3' UTR stem-loop	Cis-reg	64	64	100.00	57	89.06	26	26	100.00	26	100.00
RF00033	MicF RNA	Gene/antisense	9	8	88.89	6	66.67	0	0	NaN	0	NaN
RF00034	RprA RNA	Gene/sRNA	9	7	77.78	9	100.00	0	0	NaN	0	NaN
RF00035	OxyS RNA	Gene/sRNA	6	4	66.67	6	100.00	0	0	NaN	0	NaN
RF00036	HIV Rev response element	Cis-reg	65	0	0.00	39	60.00	0	0	NaN	0	NaN
RF00037	Iron response element	Cis-reg	39	39	100.00	33	84.62	0	0	NaN	0	NaN
RF00038	PrfA thermoregulator UTR	Cis-reg/thermoregulator	11	11	100.00	11	100.00	5	5	100.00	5	100.00
RF00039	DicF RNA	Gene/antisense	5	5	100.00	5	100.00	2	2	100.00	2	100.00
RF00040	RNase E 5' UTR element	Cis-reg	7	5	71.43	7	100.00	0	0	NaN	0	NaN
RF00041	Enteroviral 3' UTR element	Cis-reg	60	49	81.67	45	75.00	0	0	NaN	0	NaN
RF00042	CopA-like RNA	Gene/antisense	17	0	0.00	11	64.71	0	0	NaN	0	NaN
RF00043	R1162-like plasmid antisense RNA	Gene/antisense	6	6	100.00	5	83.33	0	0	NaN	0	NaN
RF00044	Bacteriophage pRNA	Gene	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00045	U17/E1 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	23	16	69.57	18	78.26	0	0	NaN	0	NaN
RF00046	Small nucleolar RNA R30/Z108	Gene/snRNA/guide/C/D-box	6	6	100.00	2	33.33	0	0	NaN	0	NaN
RF00048	Enterovirus cis-acting replication element	Cis-reg	56	31	55.36	35	62.50	56	30	53.57	23	41.07
RF00049	U36/R47/Z100 small nucleolar RNA	Gene/snRNA/guide/C/D-box	20	20	100.00	19	95.00	3	3	100.00	2	66.67
RF00050	FMN riboswitch (RFN element)	Cis-reg/riboswitch	48	41	85.42	45	93.75	0	0	NaN	0	NaN
RF00054	U25 small nucleolar RNA	Gene/snRNA/guide/C/D-box	8	8	100.00	7	87.50	2	2	100.00	1	50.00
RF00055	Small nucleolar RNA Z37	Gene/snRNA/guide/C/D-box	8	8	100.00	5	62.50	0	0	NaN	0	NaN
RF00056	U71 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	15	10	66.67	11	73.33	0	0	NaN	0	NaN
RF00057	RyhB RNA	Gene/sRNA	9	9	100.00	6	66.67	0	0	NaN	0	NaN
RF00058	HgcF RNA	Gene	4	0	0.00	4	100.00	0	0	NaN	0	NaN
RF00059	TPP riboswitch (THI element)	Cis-reg/riboswitch	236	223	94.49	201	85.17	4	4	100.00	4	100.00

				miPred		miPred-NBC		3SVM‡			3SVM-NBC	
Accession	Type†	Class	N	TN	%SP	TN	%SP	N	TN	%SP	TN	%SP
RF00060	HgcE RNA	Gene	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00061	Hepatitis C virus IRES	Cis-reg/IRES	786	658	83.72	674	85.75	1	0	0.00	0	0.00
RF00062	HgcC family RNA	Gene	22	7	31.82	22	100.00	0	0	NaN	0	NaN
RF00063	SscA RNA	Gene	5	5	100.00	3	60.00	0	0	NaN	0	NaN
RF00064	HgcG RNA	Gene	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00065	snoR9 / snoR19 family	Gene/snRNA/guide/C/D-box	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00066	U7 small nuclear RNA	Gene/snRNA	28	24	85.71	24	85.71	7	7	100.00	7	100.00
RF00067	U15 small nucleolar RNA	Gene/snRNA/guide/C/D-box	18	16	88.89	15	83.33	2	1	50.00	2	100.00
RF00068	U21 small nucleolar RNA	Gene/snRNA/guide/C/D-box	5	5	100.00	4	80.00	3	3	100.00	1	33.33
RF00069	U24/Z20/U76 small nucleolar RNA	Gene/snRNA/guide/C/D-box	14	14	100.00	10	71.43	3	3	100.00	3	100.00
RF00070	Small nucleolar RNA U29	Gene/snRNA/guide/C/D-box	10	10	100.00	6	60.00	2	2	100.00	2	100.00
RF00071	U73 small nucleolar RNA	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	2	2	100.00	2	100.00
RF00072	U23 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	6	2	33.33	3	50.00	0	0	NaN	0	NaN
RF00077	SraB RNA	Gene/sRNA	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00078	SraD RNA	Gene/sRNA	5	5	100.00	4	80.00	0	0	NaN	0	NaN
RF00079	SraE/RygA/RygB family RNA	Gene/sRNA	6	5	83.33	4	66.67	0	0	NaN	0	NaN
RF00080	yybP-ykoY element	Cis-reg/riboswitch	74	52	70.27	70	94.59	2	2	100.00	1	50.00
RF00081	SraH RNA	Gene/sRNA	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00082	SraG RNA	Gene/sRNA	5	4	80.00	5	100.00	0	0	NaN	0	NaN
RF00083	SraJ RNA	Gene/sRNA	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00084	CsrC RNA family	Gene/sRNA	5	1	20.00	4	80.00	0	0	NaN	0	NaN
RF00085	U28 small nucleolar RNA	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	2	2	100.00	1	50.00
RF00086	U27/Z191/snR74/Z4 small nucleolar RNA	Gene/snRNA/guide/C/D-box	10	10	100.00	7	70.00	0	0	NaN	0	NaN
RF00087	U26 small nucleolar RNA	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	2	0	0.00	2	100.00
RF00088	U30 small nucleolar RNA	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	2	1	50.00	1	50.00
RF00089	U31 small nucleolar RNA	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	4	4	100.00	2	50.00
RF00090	U19 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	3	0	0.00	2	66.67	0	0	NaN	0	NaN
RF00091	Small nucleolar RNA E2/ACA6/M2/MBI-136	Gene/snRNA/guide/H/ACA-box	10	2	20.00	8	80.00	0	0	NaN	0	NaN
RF00092	E3 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	9	4	44.44	9	100.00	0	0	NaN	0	NaN
RF00093	U18 small nucleolar RNA	Gene/snRNA/guide/C/D-box	16	16	100.00	14	87.50	10	9	90.00	8	80.00
RF00094	Hepatitis delta virus ribozyme	Gene/ribozyme	15	14	93.33	15	100.00	0	0	NaN	0	NaN
RF00095	Pyrococcus C/D box small nucleolar RNA	Gene/snRNA/guide/C/D-box	38	38	100.00	37	97.37	18	18	100.00	17	94.44
RF00096	U8 small nucleolar RNA	Gene/snRNA/guide/C/D-box	5	2	40.00	3	60.00	0	0	NaN	0	NaN
RF00097	Plant small nucleolar RNA R71	Gene/snRNA/guide/C/D-box	21	18	85.71	21	100.00	0	0	NaN	0	NaN
RF00098	Snake H/ACA box small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	22	22	100.00	20	90.91	0	0	NaN	0	NaN
RF00099	U22 small nucleolar RNA	Gene/snRNA/guide/C/D-box	3	2	66.67	2	66.67	0	0	NaN	0	NaN
RF00100	7SK RNA	Gene	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00101	SraC/RyeA RNA	Gene/sRNA	7	3	42.86	7	100.00	0	0	NaN	0	NaN
RF00102	VA RNA	Gene	23	0	0.00	22	95.65	0	0	NaN	0	NaN
RF00105	HBII-52 small nucleolar RNA	Gene/snRNA/guide/C/D-box	23	23	100.00	14	60.87	1	1	100.00	1	100.00
RF00106	RNAI	Gene/antisense	10	0	0.00	6	60.00	0	0	NaN	0	NaN
RF00107	FinP	Gene	6	0	0.00	6	100.00	0	0	NaN	0	NaN
RF00108	HBII-85 small nucleolar RNA	Gene/snRNA/guide/C/D-box	7	7	100.00	7	100.00	0	0	NaN	0	NaN
RF00109	Vimentin 3' UTR protein-binding region	Cis-reg	12	12	100.00	11	91.67	2	2	100.00	2	100.00
RF00110	RybB RNA	Gene/sRNA	4	2	50.00	4	100.00	2	2	100.00	0	0.00
RF00111	RyeB RNA	Gene/sRNA	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00112	RyeE RNA	Gene/sRNA	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00113	QUAD RNA	Gene/sRNA	15	6	40.00	15	100.00	0	0	NaN	0	NaN
RF00114	Ribosomal S15 leader	Cis-reg	11	11	100.00	7	63.64	0	0	NaN	0	NaN
RF00115	IS061 RNA	Gene/sRNA	5	5	100.00	2	40.00	0	0	NaN	0	NaN
RF00116	C0465 RNA	Gene/sRNA	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00117	C0719 RNA	Gene/sRNA	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00118	rydB RNA	Gene/sRNA	5	5	100.00	4	80.00	5	5	100.00	0	0.00
RF00119	C0299 RNA	Gene/sRNA	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00120	C0343 RNA	Gene/sRNA	4	4	100.00	1	25.00	0	0	NaN	0	NaN
RF00121	MicC RNA	Gene/sRNA	4	3	75.00	3	75.00	0	0	NaN	0	NaN
RF00122	GadY	Gene/sRNA	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00124	IS102 RNA	Gene/sRNA	8	1	12.50	8	100.00	0	0	NaN	0	NaN
RF00125	IS128 RNA	Gene/sRNA	3	1	33.33	3	100.00	0	0	NaN	0	NaN

				miPred		miPred-NBC		3SVM‡			3SVM-NBC	
Accession	Type†	Class	N	TN	%SP	TN	%SP	N	TN	%SP	TN	%SP
RF00126	ryfA RNA	Gene/sRNA	6	0	0.00	6	100.00	0	0	NaN	0	NaN
RF00127	t44 RNA	Gene/sRNA	9	4	44.44	9	100.00	0	0	NaN	0	NaN
RF00128	tke1 RNA	Gene/sRNA	7	2	28.57	5	71.43	0	0	NaN	0	NaN
RF00132	Small nucleolar RNA R24	Gene/snRNA/guide/C/D-box	12	7	58.33	9	75.00	4	2	50.00	1	25.00
RF00133	Small nucleolar RNA Z195	Gene/snRNA/guide/C/D-box	8	8	100.00	8	100.00	0	0	NaN	0	NaN
RF00134	Small nucleolar RNA Z196	Gene/snRNA/guide/C/D-box	7	7	100.00	3	42.86	0	0	NaN	0	NaN
RF00135	Small nucleolar RNA Z223	Gene/snRNA/guide/C/D-box	5	4	80.00	4	80.00	2	2	100.00	0	0.00
RF00136	U81 small nucleolar RNA	Gene/snRNA/guide/C/D-box	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00137	U83/U84 small nucleolar RNA	Gene/snRNA/guide/C/D-box	7	7	100.00	7	100.00	0	0	NaN	0	NaN
RF00138	U16 small nucleolar RNA	Gene/snRNA/guide/C/D-box	5	3	60.00	5	100.00	3	2	66.67	2	66.67
RF00139	U72 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	7	4	57.14	7	100.00	0	0	NaN	0	NaN
RF00140	Alpha operon ribosome binding site	Cis-reg	9	3	33.33	8	88.89	0	0	NaN	0	NaN
RF00141	Small nucleolar RNA R39/R59	Gene/snRNA/guide/C/D-box	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00142	Small nucleolar RNA Z118/Z121/Z120	Gene/snRNA/guide/C/D-box	7	4	57.14	4	57.14	2	2	100.00	1	50.00
RF00145	Small nucleolar RNA Z105	Gene/snRNA/guide/C/D-box	5	5	100.00	5	100.00	1	1	100.00	1	100.00
RF00146	Small nucleolar RNA U33	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00147	Small nucleolar RNA U34	Gene/snRNA/guide/C/D-box	9	9	100.00	8	88.89	4	3	75.00	3	75.00
RF00149	Small nucleolar RNA Z103	Gene/snRNA/guide/C/D-box	9	9	100.00	7	77.78	0	0	NaN	0	NaN
RF00150	Small nucleolar RNA U42	Gene/snRNA/guide/C/D-box	7	7	100.00	6	85.71	0	0	NaN	0	NaN
RF00151	Small nucleolar RNA U58	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00152	Small nucleolar RNA U79/Z22	Gene/snRNA/guide/C/D-box	6	6	100.00	6	100.00	3	3	100.00	0	0.00
RF00153	Small nucleolar RNA U62	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	1	1	100.00	1	100.00
RF00154	Small nucleolar RNA U63	Gene/snRNA/guide/C/D-box	2	2	100.00	1	50.00	1	1	100.00	1	100.00
RF00155	Small nucleolar RNA U66	Gene/snRNA/guide/H/ACA-box	3	1	33.33	2	66.67	0	0	NaN	0	NaN
RF00156	Small nucleolar RNA U70	Gene/snRNA/guide/H/ACA-box	14	11	78.57	13	92.86	0	0	NaN	0	NaN
RF00157	Small nucleolar RNA U39/U55	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00158	Small nucleolar RNA U82/Z25	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	1	1	100.00	1	100.00
RF00159	Small nucleolar RNA Z168/Z174	Gene/snRNA/guide/C/D-box	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00160	Small nucleolar RNA Z159/U59	Gene/snRNA/guide/C/D-box	10	10	100.00	9	90.00	0	0	NaN	0	NaN
RF00161	Nanos 3' UTR translation control element	Cis-reg	2	1	50.00	1	50.00	0	0	NaN	0	NaN
RF00162	SAM riboswitch (S box leader)	Cis-reg/riboswitch	71	53	74.65	60	84.51	1	1	100.00	0	0.00
RF00163	Hammerhead ribozyme (type I)	Gene/ribozyme	74	72	97.30	67	90.54	39	36	92.31	26	66.67
RF00164	Coronavirus 3' stem-loop II-like motif (s2m)	Cis-reg	37	37	100.00	37	100.00	33	33	100.00	32	96.97
RF00165	Coronavirus 3' UTR pseudoknot	Cis-reg	14	14	100.00	13	92.86	0	0	NaN	0	NaN
RF00166	PrrB/RsmZ RNA family	Gene/sRNA	6	6	100.00	5	83.33	0	0	NaN	0	NaN
RF00167	Purine riboswitch	Cis-reg/riboswitch	37	36	97.30	23	62.16	0	0	NaN	0	NaN
RF00168	Lysine riboswitch	Cis-reg/riboswitch	60	37	61.67	54	90.00	0	0	NaN	0	NaN
RF00169	Bacterial signal recognition particle RNA	Gene	70	43	61.43	52	74.29	58	55	94.83	32	55.17
RF00170	Retron msr RNA	Gene	8	7	87.50	6	75.00	2	2	100.00	1	50.00
RF00171	Tombusvirus 5' UTR	Cis-reg	9	9	100.00	9	100.00	1	0	0.00	1	100.00
RF00172	ctgf/hcs24 CAESAR	Cis-reg	9	9	100.00	8	88.89	0	0	NaN	0	NaN
RF00173	Hairpin ribozyme	Gene/ribozyme	3	3	100.00	3	100.00	1	1	100.00	0	0.00
RF00174	Cobalamin riboswitch	Cis-reg/riboswitch	170	135	79.41	157	92.35	0	0	NaN	0	NaN
RF00175	Retroviral Psi packaging element	Cis-reg	168	168	100.00	156	92.86	0	0	NaN	0	NaN
RF00176	Tombusvirus 3' UTR region IV	Cis-reg	18	18	100.00	17	94.44	0	0	NaN	0	NaN
RF00177	Small subunit ribosomal RNA, 5' domain	Gene/rRNA	358	175	48.88	325	90.78	0	0	NaN	0	NaN
RF00179	GAIT element	Cis-reg	8	8	100.00	8	100.00	4	4	100.00	4	100.00
RF00180	Renin stability regulatory element (REN-SRE)	Cis-reg	13	13	100.00	13	100.00	0	0	NaN	0	NaN
RF00181	C/D box small nucleolar RNA 14q(I)/14q(II)	Gene/snRNA/guide/C/D-box	59	57	96.61	50	84.75	36	36	100.00	33	91.67
RF00182	Coronavirus packaging signal	Cis-reg	15	10	66.67	15	100.00	15	5	33.33	5	33.33
RF00183	G-CSF factor stem-loop destabilising element (SLDE)	Cis-reg	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00184	Potato virus X cis-acting regulatory element	Cis-reg	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00185	Flavivirus 3' UTR pseudoknot	Cis-reg	14	3	21.43	11	78.57	0	0	NaN	0	NaN
RF00186	Small nucleolar RNA U101	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	3	3	100.00	3	100.00
RF00187	Small nucleolar RNA U102	Gene/snRNA/guide/C/D-box	2	2	100.00	2	100.00	2	2	100.00	2	100.00
RF00188	Small nucleolar RNA U103	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00189	Small nucleolar RNA U95	Gene/snRNA/guide/C/D-box	5	5	100.00	4	80.00	5	5	100.00	5	100.00
RF00190	U98 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	3	3	100.00	2	66.67	0	0	NaN	0	NaN

				miPred		miPred-NBC		3SVM[‡]			3SVM-NBC	
Accession	*Type[†]*	*Class*	*N*	*TN*	*%SP*	*TN*	*%SP*	*N*	*TN*	*%SP*	*TN*	*%SP*
RF00191	U99 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00192	Bovine leukaemia virus RNA packaging signal	Cis-reg	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00193	Citrus tristeza virus replication signal	Cis-reg	9	9	100.00	9	100.00	0	0	NaN	0	NaN
RF00194	Rubella virus 3' cis-acting element	Cis-reg	9	9	100.00	9	100.00	0	0	NaN	0	NaN
RF00195	RsmY RNA family	Gene/sRNA	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00196	Alfalfa mosaic virus RNA 1 5' UTR stem-loop	Cis-reg	4	2	50.00	0	0.00	2	2	100.00	0	0.00
RF00197	rbcL 5' UTR RNA stabilising element	Cis-reg	3	2	66.67	3	100.00	0	0	NaN	0	NaN
RF00198	SL1 RNA	Gene	28	0	0.00	24	85.71	0	0	NaN	0	NaN
RF00199	SL2 RNA	Gene	32	10	31.25	24	75.00	0	0	NaN	0	NaN
RF00200	Small nucleolar RNA Z199	Gene/snRNA/guide/C/D-box	8	8	100.00	7	87.50	6	6	100.00	4	66.67
RF00201	Small nucleolar RNA Z278	Gene/snRNA/guide/C/D-box	7	5	71.43	7	100.00	7	7	100.00	5	71.43
RF00202	Small nucleolar RNA R66	Gene/snRNA/guide/C/D-box	6	6	100.00	6	100.00	1	1	100.00	1	100.00
RF00203	Small nucleolar RNA R160	Gene/snRNA/guide/C/D-box	9	9	100.00	9	100.00	4	4	100.00	4	100.00
RF00204	Small nucleolar RNA R12	Gene/snRNA/guide/C/D-box	9	9	100.00	8	88.89	2	2	100.00	2	100.00
RF00205	Small nucleolar RNA R41	Gene/snRNA/guide/C/D-box	7	7	100.00	6	85.71	7	7	100.00	1	14.29
RF00206	Small nucleolar RNA U54	Gene/snRNA/guide/C/D-box	13	13	100.00	11	84.62	1	1	100.00	1	100.00
RF00207	K10 transport/localisation element (TLS)	Cis-reg	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00208	Small nucleolar RNA R72	Gene/snRNA/guide/C/D-box	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00209	Pestivirus IRES	Cis-reg/IRES	25	1	4.00	20	80.00	0	0	NaN	0	NaN
RF00210	Aphthovirus IRES	Cis-reg/IRES	32	2	6.25	29	90.63	0	0	NaN	0	NaN
RF00211	Small nucleolar RNA U35	Gene/snRNA/guide/C/D-box	8	8	100.00	5	62.50	1	1	100.00	0	0.00
RF00212	U38 small nucleolar RNA	Gene/snRNA/guide/C/D-box	7	7	100.00	6	85.71	3	3	100.00	2	66.67
RF00213	Small nucleolar RNA R38	Gene/snRNA/guide/C/D-box	12	10	83.33	11	91.67	6	6	100.00	3	50.00
RF00214	Retrovirus direct repeat 1 (dr1)	Cis-reg	25	24	96.00	21	84.00	1	0	0.00	1	100.00
RF00215	Tombus virus defective interfering (DI) RNA region 3	Cis-reg	48	48	100.00	34	70.83	6	6	100.00	6	100.00
RF00216	c-myc IRES	Cis-reg/IRES	23	23	100.00	21	91.30	0	0	NaN	0	NaN
RF00217	Small nucleolar RNA U20	Gene/snRNA/guide/C/D-box	4	4	100.00	3	75.00	4	4	100.00	3	75.00
RF00218	Small nucleolar RNA U40	Gene/snRNA/guide/C/D-box	9	9	100.00	9	100.00	8	8	100.00	4	50.00
RF00219	Small nucleolar RNA U32	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00220	Human rhinovirus internal cis-acting regulatory element	Cis-reg	12	12	100.00	12	100.00	10	10	100.00	10	100.00
RF00221	Small nucleolar RNA U43	Gene/snRNA/guide/C/D-box	6	5	83.33	3	50.00	3	2	66.67	3	100.00
RF00222	Bag-1 IRES	Cis-reg/IRES	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00223	bip IRES	Cis-reg/IRES	4	4	100.00	4	100.00	2	2	100.00	2	100.00
RF00224	FGF-2 IRES	Cis-reg/IRES	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00225	Tobamovirus IRES	Cis-reg/IRES	7	7	100.00	7	100.00	0	0	NaN	0	NaN
RF00226	n-myc IRES	Cis-reg/IRES	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00227	FIE3 (ftz instability element 3') element	Cis-reg	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00228	Hepatitis A virus IRES	Cis-reg/IRES	23	9	39.13	22	95.65	0	0	NaN	0	NaN
RF00229	Picornavirus IRES	Cis-reg/IRES	195	96	49.23	180	92.31	0	0	NaN	0	NaN
RF00230	T-box leader	Cis-reg	66	28	42.42	60	90.91	0	0	NaN	0	NaN
RF00231	U93 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00232	Spi-1 (PU.1) 5' UTR regulatory element	Cis-reg	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00233	Tymovirus/Pomovirus tRNA-like 3' UTR element	Cis-reg	27	27	100.00	23	85.19	0	0	NaN	0	NaN
RF00234	glmS glucosamine-6-phosphate activated ribozyme	Cis-reg/riboswitch	14	10	71.43	11	78.57	0	0	NaN	0	NaN
RF00235	Plasmid RNAIII	Gene	7	0	0.00	7	100.00	0	0	NaN	0	NaN
RF00236	ctRNA	Gene/antisense	17	0	0.00	16	94.12	0	0	NaN	0	NaN
RF00238	ctRNA	Gene/antisense	48	5	10.42	44	91.67	0	0	NaN	0	NaN
RF00240	RNA-OUT	Gene	7	0	0.00	3	42.86	7	2	28.57	3	42.86
RF00242	ctRNA	Gene/antisense	15	6	40.00	10	66.67	0	0	NaN	0	NaN
RF00243	traJ 5' UTR	Cis-reg	6	2	33.33	6	100.00	0	0	NaN	0	NaN
RF00250	Trans-activation response element (TAR)	Cis-reg	416	26	6.25	370	88.94	412	49	11.89	221	53.64
RF00252	Alfalfa mosaic virus coat protein binding (CPB) RNA	Cis-reg	18	2	11.11	18	100.00	0	0	NaN	0	NaN
RF00259	Interferon gamma 5' UTR regulatory element	Cis-reg	5	5	100.00	2	40.00	0	0	NaN	0	NaN
RF00260	Hepatitis C virus (HCV) cis-acting replication element	Cis-reg	52	52	100.00	52	100.00	52	52	100.00	46	88.46
RF00261	L-myc IRES	Cis-reg/IRES	2	2	100.00	2	100.00	0	0	NaN	0	NaN
RF00262	sar RNA	Gene	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00263	U68 small nucleolar RNA	Gene/snRNA/guide/H/ACA-box	4	3	75.00	3	75.00	0	0	NaN	0	NaN
RF00264	Small nucleolar RNA U64	Gene/snRNA/guide/H/ACA-box	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00265	Small nucleolar RNA U69	Gene/snRNA/guide/H/ACA-box	3	1	33.33	2	66.67	0	0	NaN	0	NaN

				miPred		miPred-NBC		3SVM‡			3SVM-NBC	
Accession	*Type†*	*Class*	*N*	*TN*	*%SP*	*TN*	*%SP*	*N*	*TN*	*%SP*	*TN*	*%SP*
RF00266	*Small nucleolar RNA Z17*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	2	50.00	0	0	NaN	0	NaN
RF00267	*Small nucleolar RNA R64*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	0	0.00	0	0	NaN	0	NaN
RF00268	*Small nucleolar RNA snoZ7/snoR77*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00270	*U61 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	2	2	100.00	2	100.00
RF00271	*U60 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00272	*U67 small nucleolar RNA*	*Gene|snRNA|guide|H/ACA-box*	10	10	100.00	8	80.00	0	0	NaN	0	NaN
RF00273	*U59 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	4	100.00	2	2	100.00	2	100.00
RF00274	*U57 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	1	1	100.00	1	100.00
RF00275	*U56 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	7	7	100.00	7	100.00	1	1	100.00	0	0.00
RF00276	*U52 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	3	75.00	3	3	100.00	2	66.67
RF00277	*U49 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	4	100.00	1	1	100.00	0	0.00
RF00278	*U50 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	6	6	100.00	6	100.00	1	1	100.00	1	100.00
RF00279	*U45 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	11	11	100.00	10	90.91	7	7	100.00	7	100.00
RF00280	*U51 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	1	0	0.00	1	100.00
RF00281	*U47 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00282	*U48 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	1	1	100.00	1	100.00
RF00283	*U91 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00284	*Z18 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	2	2	100.00	2	100.00
RF00285	*Z6 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	1	1	100.00	1	100.00
RF00286	*U92 small nucleolar RNA*	*Gene|snRNA|guide|H/ACA-box*	3	1	33.33	2	66.67	0	0	NaN	0	NaN
RF00287	*U44 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	2	66.67	1	1	100.00	1	100.00
RF00288	*Z30 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	3	75.00	4	4	100.00	1	25.00
RF00289	*Z12 small nucleolar RNA*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	2	2	100.00	2	100.00
RF00290	*Bamboo mosaic potexvirus (BaMV) CE*	*Cis-reg*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00291	*Small nucleolar RNA snoR639/H1*	*Gene|snRNA|guide|H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00292	*Small nucleolar RNA TBR5*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00293	*Small nucleolar RNA snoM1*	*Gene|snRNA|guide|H/ACA-box*	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00294	*Small nucleolar RNA TBR17*	*Gene|snRNA|guide|C/D-box*	4	3	75.00	4	100.00	0	0	NaN	0	NaN
RF00295	*Small nucleolar RNA TBR7*	*Gene|snRNA|guide|C/D-box*	6	6	100.00	5	83.33	1	1	100.00	1	100.00
RF00296	*Small nucleolar RNA R16*	*Gene|snRNA|guide|C/D-box*	6	6	100.00	5	83.33	2	2	100.00	2	100.00
RF00297	*Small nucleolar RNA Z177*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00299	*Small nucleolar RNA Z200*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00300	*Small nucleolar RNA Z221*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	2	66.67	2	2	100.00	1	50.00
RF00301	*Small nucleolar RNA Z256*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	1	33.33	0	0	NaN	0	NaN
RF00302	*Small nucleolar RNA U65*	*Gene|snRNA|guide|H/ACA-box*	4	0	0.00	4	100.00	0	0	NaN	0	NaN
RF00303	*Small nucleolar RNA snoR86*	*Gene|snRNA|guide|H/ACA-box*	3	3	100.00	1	33.33	0	0	NaN	0	NaN
RF00304	*Small nucleolar RNA Z279*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00305	*Small nucleolar RNA Z248*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00306	*Small nucleolar RNA Z178*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00307	*Small nucleolar RNA snoR98*	*Gene|snRNA|guide|H/ACA-box*	5	5	100.00	5	100.00	1	1	100.00	1	100.00
RF00308	*Small nucleolar RNA Z268*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	2	2	100.00	1	50.00
RF00309	*Small nucleolar RNA snR60/Z15/Z230/Z193/J17*	*Gene|snRNA|guide|C/D-box*	24	23	95.83	21	87.50	5	4	80.00	1	20.00
RF00310	*Small nucleolar RNA Z165*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	1	33.33	3	3	100.00	0	0.00
RF00311	*Small nucleolar RNA Z188*	*Gene|snRNA|guide|C/D-box*	4	1	25.00	4	100.00	3	0	0.00	3	100.00
RF00312	*Small nucleolar RNA Z206*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00313	*Small nucleolar RNA Z173*	*Gene|snRNA|guide|C/D-box*	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00314	*Small nucleolar RNA Z182*	*Gene|snRNA|guide|C/D-box*	7	7	100.00	7	100.00	4	4	100.00	0	0.00
RF00315	*Small nucleolar RNA J33*	*Gene|snRNA|guide|C/D-box*	5	5	100.00	3	60.00	2	2	100.00	0	0.00
RF00316	*Small nucleolar RNA R43*	*Gene|snRNA|guide|C/D-box*	16	16	100.00	16	100.00	6	6	100.00	5	83.33
RF00317	*Small nucleolar RNA Z163*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00318	*Small nucleolar RNA Z175*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00319	*Small nucleolar RNA MBI-1*	*Gene|snRNA|guide|H/ACA-box*	4	2	50.00	4	100.00	0	0	NaN	0	NaN
RF00320	*Small nucleolar RNA Z185*	*Gene|snRNA|guide|C/D-box*	3	2	66.67	2	66.67	1	1	100.00	0	0.00
RF00321	*Small nucleolar RNA Z247*	*Gene|snRNA|guide|C/D-box*	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00322	*Small nucleolar RNA MBI-161*	*Gene|snRNA|guide|H/ACA-box*	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00323	*Small nucleolar RNA R79*	*Gene|snRNA|guide|C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00324	*Small nucleolar RNA MBII-202*	*Gene|snRNA|guide|C/D-box*	5	5	100.00	4	80.00	0	0	NaN	0	NaN
RF00325	*Small nucleolar RNA U53*	*Gene|snRNA|guide|C/D-box*	3	3	100.00	3	100.00	3	3	100.00	3	100.00

				miPred		miPred-NBC		3SVM‡			3SVM-NBC	
Accession	*Type†*	*Class*	*N*	*TN*	*%SP*	*TN*	*%SP*	*N*	*TN*	*%SP*	*TN*	*%SP*
RF00326	*Small nucleolar RNA Z155*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00327	*Small nucleolar RNA Z194*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00328	*Small nucleolar RNA Z161/Z228*	*Gene/snRNA/guide/C/D-box*	7	7	100.00	5	71.43	2	2	100.00	2	100.00
RF00329	*Small nucleolar RNA Z162*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00330	*Small nucleolar RNA Z43*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	3	75.00	2	2	100.00	1	50.00
RF00331	*Small nucleolar RNA Z169*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	2	2	100.00	2	100.00
RF00332	*Small nucleolar RNA Z266*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	2	50.00	2	2	100.00	0	0.00
RF00333	*Small nucleolar RNA Z157/R69/R10*	*Gene/snRNA/guide/C/D-box*	10	8	80.00	5	50.00	2	2	100.00	0	0.00
RF00334	*Small nucleolar RNA MBI-28*	*Gene/snRNA/guide/H/ACA-box*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00335	*Small nucleolar RNA Z13/snr52*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	3	3	100.00	3	100.00
RF00336	*Small nucleolar RNA J26*	*Gene/snRNA/guide/C/D-box*	5	5	100.00	3	60.00	0	0	NaN	0	NaN
RF00337	*Small nucleolar RNA Z112*	*Gene/snRNA/guide/C/D-box*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00338	*Small nucleolar RNA snR53*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	1	0	0.00	1	100.00
RF00339	*Small nucleolar RNA snoR60*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00340	*Small nucleolar RNA snoMBI-87*	*Gene/snRNA/guide/H/ACA-box*	6	0	0.00	4	66.67	0	0	NaN	0	NaN
RF00341	*Small nucleolar RNA Z39*	*Gene/snRNA/guide/C/D-box*	5	5	100.00	5	100.00	4	4	100.00	4	100.00
RF00342	*Small nucleolar RNA Z40*	*Gene/snRNA/guide/C/D-box*	5	5	100.00	3	60.00	5	5	100.00	5	100.00
RF00343	*Small nucleolar RNA Z122*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	0	0.00	1	1	100.00	1	100.00
RF00344	*Small nucleolar RNA Z267*	*Gene/snRNA/guide/C/D-box*	5	5	100.00	3	60.00	2	2	100.00	2	100.00
RF00345	*Small nucleolar RNA snoR1*	*Gene/snRNA/guide/C/D-box*	7	7	100.00	6	85.71	3	3	100.00	1	33.33
RF00346	*Small nucleolar RNA snoZ1*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00347	*Small nucleolar RNA Z50*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	1	1	100.00	0	0.00
RF00348	*Small nucleolar RNA snoR9*	*Gene/snRNA/guide/C/D-box*	7	7	100.00	6	85.71	3	3	100.00	3	100.00
RF00349	*Small nucleolar RNA R11/Z151*	*Gene/snRNA/guide/C/D-box*	5	5	100.00	3	60.00	0	0	NaN	0	NaN
RF00350	*Small nucleolar RNA Z152/R70/R12/*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00351	*Small nucleolar RNA R20*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	2	2	100.00	0	0.00
RF00352	*Small nucleolar RNA R21*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	2	50.00	0	0	NaN	0	NaN
RF00353	*Small nucleolar RNA snoR31/Z110/Z27*	*Gene/snRNA/guide/C/D-box*	8	5	62.50	2	25.00	0	0	NaN	0	NaN
RF00355	*Small nucleolar RNA snoR28*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00356	*Small nucleolar RNA R32/R81/Z41*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	4	4	100.00	1	25.00
RF00357	*Small nucleolar RNA R44/J54*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	3	75.00	3	3	100.00	3	100.00
RF00358	*Small nucleolar RNA Z101*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	2	66.67	2	2	100.00	0	0.00
RF00359	*Small nucleolar RNA Z102/R77*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	2	2	100.00	2	100.00
RF00360	*Small nucleolar RNA Z107/R87*	*Gene/snRNA/guide/C/D-box*	6	5	83.33	6	100.00	0	0	NaN	0	NaN
RF00361	*Small nucleolar RNA Z119*	*Gene/snRNA/guide/C/D-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00362	*Pospiviroid RY motif stem loop*	*Cis-reg*	16	14	87.50	15	93.75	11	11	100.00	3	27.27
RF00368	*sroB RNA*	*Gene/sRNA*	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00369	*sroC RNA*	*Gene/sRNA*	5	0	0.00	4	80.00	0	0	NaN	0	NaN
RF00370	*sroD RNA*	*Gene/sRNA*	3	2	66.67	3	100.00	0	0	NaN	0	NaN
RF00371	*sroE RNA*	*Gene/sRNA*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00372	*sroH RNA*	*Gene/sRNA*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00373	*Archaeal RNase P*	*Gene/ribozyme*	40	16	40.00	33	82.50	0	0	NaN	0	NaN
RF00374	*Gammaretrovirus core encapsidation signal*	*Cis-reg*	23	11	47.83	23	100.00	0	0	NaN	0	NaN
RF00375	*HIV primer binding site (PBS)*	*Cis-reg*	373	265	71.05	334	89.54	0	0	NaN	0	NaN
RF00376	*HIV gag stem loop 3 (GSL3)*	*Cis-reg*	1374	1371	99.78	1200	87.34	9	9	100.00	4	44.44
RF00377	*Small nucleolar RNA U6-53/MBII-28*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00378	*Qrr RNA*	*Gene/sRNA*	14	7	50.00	9	64.29	0	0	NaN	0	NaN
RF00379	*ydaO/yuaA element*	*Cis-reg/riboswitch*	35	35	100.00	32	91.43	0	0	NaN	0	NaN
RF00380	*ykoK element*	*Cis-reg/riboswitch*	39	25	64.10	32	82.05	0	0	NaN	0	NaN
RF00381	*Antizyme RNA frameshifting stimulation element*	*Cis-reg/frameshift*	13	12	92.31	12	92.31	10	10	100.00	7	70.00
RF00382	*DnaX ribosomal frameshifting element*	*Cis-reg/frameshift*	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00383	*Insertion sequence IS1222 ribosomal frameshifting element*	*Cis-reg/frameshift*	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00384	*Poxvirus AX element late mRNA CE*	*Cis-reg*	7	7	100.00	7	100.00	0	0	NaN	0	NaN
RF00385	*Infectious bronchitis virus D-RNA*	*Cis-reg*	10	6	60.00	10	100.00	10	8	80.00	6	60.00
RF00386	*Enterovirus 5' cloverleaf cis-acting replication element*	*Cis-reg*	60	5	8.33	52	86.67	0	0	NaN	0	NaN
RF00387	*FGF-1 IRES*	*Cis-reg/IRES*	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00388	*Qa RNA*	*Gene/antisense*	5	2	40.00	3	60.00	0	0	NaN	0	NaN
RF00389	*Bamboo mosaic virus satellite RNA CE*	*Cis-reg*	42	42	100.00	41	97.62	0	0	NaN	0	NaN
RF00390	*UPSK RNA*	*Cis-reg*	4	4	100.00	4	100.00	0	0	NaN	0	NaN

				miPred		miPred-NBC		3SVM[‡]			3SVM-NBC	
Accession	*Type[†]*	*Class*	*N*	*TN*	*%SP*	*TN*	*%SP*	*N*	*TN*	*%SP*	*TN*	*%SP*
RF00391	*RtT RNA*	*Cis-reg*	19	16	84.21	18	94.74	0	0	NaN	0	NaN
RF00392	*Small nucleolar RNA ACA5*	*Gene/snRNA/guide/H/ACA-box*	6	6	100.00	4	66.67	0	0	NaN	0	NaN
RF00393	*Small nucleolar RNA ACA8*	*Gene/snRNA/guide/H/ACA-box*	5	4	80.00	4	80.00	0	0	NaN	0	NaN
RF00394	*Small nucleolar RNA ACA4*	*Gene/snRNA/guide/H/ACA-box*	7	4	57.14	7	100.00	0	0	NaN	0	NaN
RF00395	*Small nucleolar RNA ACA10*	*Gene/snRNA/guide/H/ACA-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00396	*Small nucleolar RNA ACA13*	*Gene/snRNA/guide/H/ACA-box*	3	0	0.00	1	33.33	0	0	NaN	0	NaN
RF00397	*Small nucleolar RNA ACA14*	*Gene/snRNA/guide/H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00398	*Small nucleolar RNA ACA15*	*Gene/snRNA/guide/H/ACA-box*	4	2	50.00	4	100.00	0	0	NaN	0	NaN
RF00399	*Small nucleolar RNA ACA24*	*Gene/snRNA/guide/H/ACA-box*	5	5	100.00	4	80.00	0	0	NaN	0	NaN
RF00400	*Small nucleolar RNA ACA28*	*Gene/snRNA/guide/H/ACA-box*	3	2	66.67	3	100.00	0	0	NaN	0	NaN
RF00401	*Small nucleolar RNA ACA20*	*Gene/snRNA/guide/H/ACA-box*	17	4	23.53	14	82.35	0	0	NaN	0	NaN
RF00402	*Small nucleolar RNA ACA25*	*Gene/snRNA/guide/H/ACA-box*	9	7	77.78	8	88.89	0	0	NaN	0	NaN
RF00403	*Small nucleolar RNA ACA41*	*Gene/snRNA/guide/H/ACA-box*	6	1	16.67	6	100.00	0	0	NaN	0	NaN
RF00404	*Small nucleolar RNA ACA46*	*Gene/snRNA/guide/H/ACA-box*	3	1	33.33	2	66.67	0	0	NaN	0	NaN
RF00405	*Small nucleolar RNA ACA44*	*Gene/snRNA/guide/H/ACA-box*	6	6	100.00	6	100.00	1	1	100.00	1	100.00
RF00406	*Small nucleolar RNA ACA42*	*Gene/snRNA/guide/H/ACA-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00407	*Small nucleolar RNA ACA50*	*Gene/snRNA/guide/H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00408	*Small nucleolar RNA ACA1*	*Gene/snRNA/guide/H/ACA-box*	6	5	83.33	5	83.33	0	0	NaN	0	NaN
RF00409	*Small nucleolar RNA ACA7*	*Gene/snRNA/guide/H/ACA-box*	8	8	100.00	6	75.00	1	1	100.00	1	100.00
RF00410	*Small nucleolar RNA ACA2/ACA34*	*Gene/snRNA/guide/H/ACA-box*	18	5	27.78	16	88.89	0	0	NaN	0	NaN
RF00411	*Small nucleolar RNA ACA9*	*Gene/snRNA/guide/H/ACA-box*	6	3	50.00	5	83.33	0	0	NaN	0	NaN
RF00412	*Small nucleolar RNA ACA21*	*Gene/snRNA/guide/H/ACA-box*	5	1	20.00	3	60.00	0	0	NaN	0	NaN
RF00413	*Small nucleolar RNA ACA19*	*Gene/snRNA/guide/H/ACA-box*	4	1	25.00	3	75.00	0	0	NaN	0	NaN
RF00414	*Small nucleolar RNA ACA22*	*Gene/snRNA/guide/H/ACA-box*	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00415	*Small nucleolar RNA ACA30/ACA37/MBI-26*	*Gene/snRNA/guide/H/ACA-box*	6	6	100.00	6	100.00	0	0	NaN	0	NaN
RF00416	*Small nucleolar RNA ACA43*	*Gene/snRNA/guide/H/ACA-box*	7	7	100.00	6	85.71	0	0	NaN	0	NaN
RF00417	*Small nucleolar RNA ACA56*	*Gene/snRNA/guide/H/ACA-box*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00418	*Small nucleolar RNA ACA52*	*Gene/snRNA/guide/H/ACA-box*	4	0	0.00	3	75.00	0	0	NaN	0	NaN
RF00419	*Small nucleolar RNA ACA52*	*Gene/snRNA/guide/H/ACA-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00420	*Small nucleolar RNA ACA61*	*Gene/snRNA/guide/H/ACA-box*	4	3	75.00	3	75.00	0	0	NaN	0	NaN
RF00421	*Small nucleolar RNA ACA32*	*Gene/snRNA/guide/H/ACA-box*	9	6	66.67	6	66.67	0	0	NaN	0	NaN
RF00422	*Small nucleolar RNA ACA12*	*Gene/snRNA/guide/H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00423	*Small nucleolar RNA ACA26*	*Gene/snRNA/guide/H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00424	*Small nucleolar RNA ACA47*	*Gene/snRNA/guide/H/ACA-box*	6	2	33.33	4	66.67	0	0	NaN	0	NaN
RF00425	*Small nucleolar RNA ACA18*	*Gene/snRNA/guide/H/ACA-box*	6	3	50.00	3	50.00	0	0	NaN	0	NaN
RF00426	*Small nucleolar RNA ACA45*	*Gene/snRNA/guide/H/ACA-box*	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00427	*Small nucleolar RNA ACA11*	*Gene/snRNA/guide/H/ACA-box*	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00428	*Small nucleolar RNA ACA38*	*Gene/snRNA/guide/H/ACA-box*	5	4	80.00	5	100.00	0	0	NaN	0	NaN
RF00429	*Small nucleolar RNA ACA29*	*Gene/snRNA/guide/H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00430	*Small nucleolar RNA ACA54*	*Gene/snRNA/guide/H/ACA-box*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00431	*Small nucleolar RNA ACA55*	*Gene/snRNA/guide/H/ACA-box*	3	0	0.00	3	100.00	0	0	NaN	0	NaN
RF00432	*Small nucleolar RNA ACA51*	*Gene/snRNA/guide/H/ACA-box*	9	8	88.89	9	100.00	0	0	NaN	0	NaN
RF00433	*Hsp90 CE*	*Cis-reg/thermoregulator*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00434	*Luteovirus cap-independent translation element (BTE)*	*Cis-reg*	17	17	100.00	13	76.47	0	0	NaN	0	NaN
RF00435	*Repression of heat shock gene expression (ROSE) element*	*Cis-reg/thermoregulator*	3	2	66.67	2	66.67	0	0	NaN	0	NaN
RF00436	*UnaL2 line 3' element*	*Cis-reg*	144	141	97.92	113	78.47	50	49	98.00	13	26.00
RF00437	*Hairy RNA localisation element (HLE)*	*Cis-reg*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00438	*Small nucleolar RNA ACA33*	*Gene/snRNA/guide/H/ACA-box*	5	5	100.00	4	80.00	0	0	NaN	0	NaN
RF00439	*Small nucleolar RNA U87*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	3	75.00	0	0	NaN	0	NaN
RF00440	*Small nucleolar RNA U37*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	3	3	100.00	3	100.00
RF00441	*Small nucleolar RNA Z242*	*Gene/snRNA/guide/C/D-box*	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00442	*ykkC-yxkD element*	*Cis-reg/riboswitch*	16	15	93.75	14	87.50	0	0	NaN	0	NaN
RF00443	*Small nucleolar RNA ACA27*	*Gene/snRNA/guide/H/ACA-box*	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00444	*PrrF RNA*	*Gene/sRNA*	7	2	28.57	7	100.00	0	0	NaN	0	NaN
RF00447	*Voltage-gated potassium-channel Kv1.4 IRES*	*Cis-reg/IRES*	6	5	83.33	6	100.00	0	0	NaN	0	NaN
RF00448	*Epstein-Barr virus nuclear antigen (EBNA) IRES*	*Cis-reg/IRES*	8	8	100.00	8	100.00	0	0	NaN	0	NaN
RF00449	*HIF-1 alpha IRES*	*Cis-reg/IRES*	7	7	100.00	7	100.00	0	0	NaN	0	NaN
RF00450	*Small nucleolar RNA R105/R108*	*Gene/snRNA/guide/C/D-box*	4	3	75.00	4	100.00	0	0	NaN	0	NaN

				miPred		miPred-NBC		3SVM‡			3SVM-NBC	
Accession	Type†	Class	N	TN	%SP	TN	%SP	N	TN	%SP	TN	%SP
RF00453	Cardiovirus cis-acting replication element	Cis-reg	12	11	91.67	9	75.00	2	2	100.00	2	100.00
RF00454	p27 CE	Cis-reg	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00457	Mnt IRES	Cis-reg/IRES	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00458	Cripavirus IRES	Cis-reg/IRES	7	6	85.71	6	85.71	0	0	NaN	0	NaN
RF00459	Mason-Pfizer monkey virus packaging signal	Cis-reg	4	4	100.00	4	100.00	0	0	NaN	0	NaN
RF00460	U1A polyadenylation inhibition element (PIE)	Cis-reg	6	6	100.00	6	100.00	3	3	100.00	3	100.00
RF00461	Vascular endothelial growth factor (VEGF) IRES A	Cis-reg/IRES	7	7	100.00	7	100.00	0	0	NaN	0	NaN
RF00462	APC IRES	Cis-reg/IRES	6	6	100.00	2	33.33	0	0	NaN	0	NaN
RF00463	Apolipoprotein B (apoB) 5' UTR CE	Cis-reg	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00465	Japanese encephalitis virus (JEV) hairpin structure	Cis-reg	20	19	95.00	19	95.00	12	12	100.00	5	41.67
RF00466	Agrobacterium tumefaciens ROSE element	Cis-reg/thermoregulator	3	1	33.33	3	100.00	0	0	NaN	0	NaN
RF00467	Rous sarcoma virus (RSV) primer binding site (PBS)	Cis-reg	23	1	4.35	21	91.30	22	13	59.09	18	81.82
RF00468	Hepatitis C stem-loop VII	Cis-reg	63	9	14.29	32	50.79	63	45	71.43	63	100.00
RF00469	Hepatitis C stem-loop IV	Cis-reg	109	2	1.83	109	100.00	109	109	100.00	61	55.96
RF00470	Togavirus 5' plus strand CE	Cis-reg	32	5	15.63	29	90.63	0	0	NaN	0	NaN
RF00471	Small nucleolar RNA snR48	Gene/snRNA/guide/C/D-box	6	6	100.00	5	83.33	1	1	100.00	0	0.00
RF00472	Small nucleolar RNA snR55/Z10	Gene/snRNA/guide/C/D-box	7	7	100.00	4	57.14	0	0	NaN	0	NaN
RF00473	Small nucleolar RNA snR54	Gene/snRNA/guide/C/D-box	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00474	Small nucleolar RNA snR57	Gene/snRNA/guide/C/D-box	6	6	100.00	5	83.33	2	2	100.00	0	0.00
RF00475	Small nucleolar RNA snR69	Gene/snRNA/guide/C/D-box	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00476	Small nucleolar RNA snR61/Z11	Gene/snRNA/guide/C/D-box	9	9	100.00	8	88.89	0	0	NaN	0	NaN
RF00477	Small nucleolar RNA snR66	Gene/snRNA/guide/C/D-box	5	5	100.00	5	100.00	0	0	NaN	0	NaN
RF00478	Small nucleolar RNA U88	Gene/snRNA/guide/C/D-box	4	0	0.00	3	75.00	0	0	NaN	0	NaN
RF00479	Small nucleolar RNA snR71	Gene/snRNA/guide/C/D-box	5	5	100.00	3	60.00	0	0	NaN	0	NaN
RF00480	HIV Ribosomal frameshift signal	Cis-reg/frameshift	768	152	19.79	704	91.67	765	719	93.99	107	13.99
RF00481	Hepatitis C virus 3'X element	Cis-reg	22	0	0.00	13	59.09	0	0	NaN	0	NaN
RF00482	Small nucleolar RNA F1/F2/snoR5a	Gene/snRNA/guide/H/ACA-box	8	5	62.50	6	75.00	0	0	NaN	0	NaN
RF00483	Insulin-like growth factor II IRES	Cis-reg/IRES	8	8	100.00	7	87.50	0	0	NaN	0	NaN
RF00484	Connexin-32 IRES	Cis-reg/IRES	6	6	100.00	5	83.33	0	0	NaN	0	NaN
RF00485	Potassium channel RNA editing signal	Cis-reg	85	76	89.41	69	81.18	13	10	76.92	7	53.85
RF00487	Connexin-43 IRES	Cis-reg/IRES	13	13	100.00	12	92.31	0	0	NaN	0	NaN
RF00488	Yeast U1 spliceosomal RNA	Gene/snRNA/splicing	6	0	0.00	5	83.33	0	0	NaN	0	NaN
RF00489	ctRNA	Gene/antisense	15	6	40.00	14	93.33	10	8	80.00	7	70.00
RF00490	S-element	Cis-reg	13	13	100.00	9	69.23	3	3	100.00	3	100.00
RF00491	Simian virus 40 late polyadenylation signal (SVLPA)	Cis-reg	3	3	100.00	2	66.67	0	0	NaN	0	NaN
RF00492	Small nucleolar RNA U12-22	Gene/snRNA/guide/C/D-box	7	7	100.00	6	85.71	3	3	100.00	3	100.00
RF00493	Small nucleolar RNA U2-30	Gene/snRNA/guide/C/D-box	3	3	100.00	3	100.00	0	0	NaN	0	NaN
RF00494	Small nucleolar RNA U2-19	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	1	1	100.00	1	100.00
RF00495	Heat shock protein 70 (Hsp70) IRES	Cis-reg/IRES	13	13	100.00	13	100.00	0	0	NaN	0	NaN
RF00496	Coronavirus SL-III cis-acting replication element	Cis-reg	5	5	100.00	5	100.00	3	3	100.00	1	33.33
RF00497	Dengue virus 3'-SL cis-acting replication element	Cis-reg	23	5	21.74	21	91.30	0	0	NaN	0	NaN
RF00498	Equine arteritis virus leader TRS hairpin (LTH)	Cis-reg	4	4	100.00	4	100.00	4	4	100.00	4	100.00
RF00499	Human parechovirus 1 (HPeV1) cis regulatory element	Cis-reg	5	2	40.00	5	100.00	0	0	NaN	0	NaN
RF00500	Turnip crinkle virus (TCV) repressor of minus strand synthesis H5	Cis-reg	3	2	66.67	3	100.00	3	3	100.00	2	66.67
RF00501	Rotavirus cis-acting replication element	Cis-reg	14	14	100.00	8	57.14	4	4	100.00	1	25.00
RF00502	Turnip crinkle virus (TCV) core promoter hairpin (Pr)	Cis-reg	4	4	100.00	4	100.00	4	4	100.00	2	50.00
RF00503	RNAIII	Gene	12	2	16.67	12	100.00	0	0	NaN	0	NaN
RF00504	gcvT element	Cis-reg/riboswitch	117	111	94.87	102	87.18	3	3	100.00	2	66.67
RF00505	RydC RNA	Gene/sRNA	3	3	100.00	3	100.00	2	2	100.00	2	100.00
RF00506	Threonine operon leader	Cis-reg	27	1	3.70	25	92.59	0	0	NaN	0	NaN
RF00507	Coronavirus frameshifting stimulation element	Cis-reg/frameshift	18	12	66.67	15	83.33	0	0	NaN	0	NaN
RF00509	Small nucleolar RNA snR64	Gene/snRNA/guide/C/D-box	4	4	100.00	4	100.00	0	0	NaN	0	NaN
−	mRNAs	−	31	27	87.10	27	87.10	0	0	NaN	0	NaN
Total ncRNA samples (exclude mRNAs)		−	12387	8507		10771		2404	1884		1199	

† , cis-regulatory element (CE); internal ribosome entry site (IRES). *N* (non *pre-miRs*), *TN* (non *pre-miRs* missed), and *%SP* (Specificity). ‡, *3SVM* model was trained on 200 human *pre-miRs* and 400 pseudo hairpins randomly selected using the latest libSVM 2.82 (the "-b 1" option was enabled) and the optimal hyperparameter pair (*C*, *γ*).

Table S4. The mean specificity of *miPred*, *miPred-NBC*, *3SVM*, and *3SVM-NBC* evaluated on the non *pre-miR* dataset IE-NC (12,387 functional ncRNAs) categorized by classes of ncRNAs.

Classes of ncRNAs	No. of types	*miPred* %SP	*miPred-NBC* %SP	No. of excluded types	*3SVM*[†] %SP	*3SVM-NBC* %SP
Cis-reg	77	74.91 ± 4.03	87.99 ± 2.03	46	83.36 ± 5.60	69.96 ± 5.61
Cis-reg/frameshift	5	75.75 ± 15.27	86.80 ± 5.68	3	96.99 ± 3.01	42.00 ± 28.01
Cis-reg/IRES	24	85.47 ± 6.02	91.02 ± 3.06	22	50.00 ± 50.00	50.00 ± 50.00
Cis-reg/riboswitch	12	82.28 ± 3.96	85.77 ± 2.56	8	100.00 ± 0.00	54.17 ± 20.83
Cis-reg/thermoregulator	4	75.00 ± 15.96	91.67 ± 8.33	3	100.00 ± 0.00	100.00 ± 0.00
Gene	24	34.73 ± 7.71	86.65 ± 3.03	18	70.57 ± 18.19	45.23 ± 3.26
Gene/antisense	10	41.93 ± 13.01	78.05 ± 5.03	8	90.00 ± 10.00	85.00 ± 15.00
Gene/ribozyme	9	60.08 ± 10.10	91.54 ± 2.36	6	97.44 ± 2.56	55.56 ± 29.40
Gene/rRNA	3	70.66 ± 12.94	90.74 ± 1.70	1	100.00 ± 0.00	75.00 ± 25.00
Gene/snRNA	1	85.71 ± 0.00	85.71 ± 0.00	0	100.00 ± 0.00	100.00 ± 0.00
Gene/snRNA/guide/C/D-box	165	94.61 ± 1.28	84.59 ± 1.58	72	92.78 ± 2.32	68.60 ± 4.06
Gene/snRNA/guide/H/ACA-box	71	60.97 ± 4.33	84.97 ± 2.04	68	100.00 ± 0.00	100.00 ± 0.00
Gene/snRNA/splicing	7	51.16 ± 13.89	87.30 ± 3.83	6	100.00 ± 0.00	100.00 ± 0.00
Gene/sRNA	42	65.71 ± 5.90	87.53 ± 2.81	39	100.00 ± 0.00	33.33 ± 33.33
Gene/tRNA	1	85.55 ± 0.00	86.98 ± 0.00	0	94.94 ± 0.00	89.87 ± 0.00
Intron	2	40.95 ± 9.05	86.70 ± 9.98	2	NaN	NaN

†, *3SVM* model was trained on 200 human *pre-miRs* and 400 pseudo hairpins randomly selected using the latest libSVM 2.82 (the "-b 1" option was enabled) and the optimal hyperparameter pair (*C*, *γ*). *%SP* (Specificity). Values are expressed as mean ± standard error.

Table S5. F1 and F2 scores for features of *miPred* and *3SVM*, sorted by descending F1 scores.

Rank	Features	*miPred*			Features	*3SVM*[†]		
		F1 score	F2 score	$\Delta F = F1 - F2$		F1 score	F2 score	$\Delta F = F1 - F2$
01	$MFEI_1$	1.28	1.52	-2.42E^{-01}	A(((	8.20E^{-01}	6.97E^{-01}	1.22E^{-01}
02	zG	1.27	1.48	-2.15E^{-01}	U(((	7.58E^{-01}	6.12E^{-01}	1.46E^{-01}
03	dP	1.03	1.18	-1.49E^{-01}	G...	4.57E^{-01}	2.05E^{-01}	2.52E^{-01}
04	zP	9.67E^{-01}	1.03	-6.33E^{-02}	A...	4.42E^{-01}	1.94E^{-01}	2.47E^{-01}
05	zQ	8.33E^{-01}	7.29E^{-01}	1.04E^{-01}	C...	4.31E^{-01}	1.84E^{-01}	2.47E^{-01}
06	dG	8.23E^{-01}	7.50E^{-01}	7.31E^{-02}	G.((	3.81E^{-01}	1.62E^{-01}	2.20E^{-01}
07	dQ	7.99E^{-01}	6.67E^{-01}	1.32E^{-01}	A(..	3.50E^{-01}	1.31E^{-01}	2.19E^{-01}
08	zD	7.92E^{-01}	6.70E^{-01}	1.23E^{-01}	A..(	3.28E^{-01}	1.17E^{-01}	2.11E^{-01}
09	dD	7.46E^{-01}	5.91E^{-01}	1.55E^{-01}	C((.	3.19E^{-01}	1.12E^{-01}	2.07E^{-01}
10	$MFEI_2$	4.41E^{-01}	1.53E^{-01}	2.88E^{-01}	G(..	3.07E^{-01}	9.75E^{-02}	2.10E^{-01}
11	%UA	3.87E^{-01}	1.56E^{-01}	2.31E^{-01}	U...	3.05E^{-01}	9.74E^{-02}	2.08E^{-01}
12	%G+C	3.06E^{-01}	1.04E^{-01}	2.02E^{-01}	C..(	2.97E^{-01}	9.54E^{-02}	2.02E^{-01}
13	zF	2.88E^{-01}	7.13E^{-02}	2.16E^{-01}	G(((	2.84E^{-01}	8.95E^{-02}	1.94E^{-01}
14	%UU	2.83E^{-01}	8.91E^{-02}	1.94E^{-01}	C(..	2.70E^{-01}	7.93E^{-02}	1.91E^{-01}
15	%GU	2.64E^{-01}	7.71E^{-02}	1.87E^{-01}	G((.	2.63E^{-01}	7.62E^{-02}	1.87E^{-01}
16	%GC	2.44E^{-01}	6.57E^{-02}	1.79E^{-01}	G..(	2.48E^{-01}	6.69E^{-02}	1.81E^{-01}
17	dF	2.42E^{-01}	5.16E^{-02}	1.90E^{-01}	U..(	2.19E^{-01}	5.20E^{-02}	1.67E^{-01}
18	%CC	2.04E^{-01}	4.59E^{-02}	1.58E^{-01}	C.((	1.89E^{-01}	3.92E^{-02}	1.50E^{-01}
19	%AA	1.83E^{-01}	3.73E^{-02}	1.46E^{-01}	C(((	1.87E^{-01}	3.88E^{-02}	1.48E^{-01}
20	%GG	1.82E^{-01}	3.68E^{-02}	1.45E^{-01}	G.(.	1.82E^{-01}	3.52E^{-02}	1.47E^{-01}
21	%CA	1.77E^{-01}	3.48E^{-02}	1.42E^{-01}	U.(.	1.71E^{-01}	2.88E^{-02}	1.42E^{-01}
22	%CG	1.73E^{-01}	3.30E^{-02}	1.40E^{-01}	U(..	1.56E^{-01}	2.69E^{-02}	1.30E^{-01}
23	%GA	1.41E^{-01}	2.13E^{-02}	1.19E^{-01}	U(.(	1.37E^{-01}	2.08E^{-02}	1.16E^{-01}
24	%AU	1.25E^{-01}	1.69E^{-02}	1.08E^{-01}	A.(.	1.22E^{-01}	1.52E^{-02}	1.07E^{-01}
25	%AG	1.08E^{-01}	1.28E^{-02}	9.54E^{-02}	C.(.	1.10E^{-01}	1.32E^{-02}	9.68E^{-02}
26	%UG	6.31E^{-02}	4.42E^{-03}	5.87E^{-02}	G(.(	1.02E^{-01}	1.13E^{-02}	9.05E^{-02}
27	%AC	3.71E^{-02}	1.53E^{-03}	3.55E^{-02}	C(.(	6.68E^{-02}	4.95E^{-03}	6.19E^{-02}
28	%CU	3.21E^{-02}	1.13E^{-03}	3.09E^{-02}	A(.(	6.06E^{-02}	4.06E^{-03}	5.65E^{-02}
29	%UC	2.18E^{-02}	5.21E^{-04}	2.13E^{-02}	A.((	5.90E^{-02}	3.87E^{-03}	5.52E^{-02}
30	–	–	–	–	A((.	3.21E^{-02}	1.14E^{-03}	3.10E^{-02}
31	–	–	–	–	U.((	3.28E^{-03}	1.20E^{-05}	3.26E^{-03}
32	–	–	–	–	U((.	6.80E^{-05}	0.00E^{+00}	6.80E^{-05}
		0.429 ± 0.0711	0.332 ± 0.0872	–		0.252 ± 0.0336	0.103 ± 0.0277	–

†, *3SVM* model was trained on 200 human *pre-miRs* and 400 pseudo hairpins randomly selected using the latest libSVM 2.82 (the "-b 1" option was enabled) and the optimal hyperparameter pair (C, γ).

Table S6. Effects of feature selection on *miPred*'s accuracy.

Classifiers	Human pre-miRs (TR-H and TE-H)	Non-human pre-miRs (IE-NH)	ncRNAs (IE-NC)	mRNAs (IE-M)
miPred	93.60	95.64	68.68	87.10
miPred₃	94.12	95.69	68.31	87.10
miPred₃/₅	92.67	95.36	71.20	100.00
miPred₃/₁₀	93.40	95.64	69.82	83.87
miPred₃/₁₅	93.40	95.79	60.93	80.65
miPred₃/₂₀	92.67	94.68	72.18	100.00
miPred₃/₂₁	92.67	95.29	72.01	100.00
miPred₃/₂₂	92.57	95.15	71.26	100.00
miPred₃/₂₃	92.67	95.22	70.15	100.00
miPred₃/₂₄	92.98	95.39	64.56	100.00
miPred₃/₂₅	91.64	93.52	63.16	96.77
miPredᵢ	77.30	76.35	67.53	90.32
miPredᵢᵢ	93.81	95.83	61.38	54.84
miPredᵢᵢᵢ	93.60	95.69	66.13	70.97

miPred₃ contains a subset of 26 features from *miPred* that excludes *dQ*, *dD*, and *zD*. Derived from *miPred₃*, the remaining nine variants denoted as *miPred₃/₅*, *miPred₃/₁₀*, …, *miPred₃/₂₄*, and *miPred₃/₂₅* only include the top ranking 21, 16, 11, 6, 5, 4, 3, 2, and 1 feature(s), respectively. *miPredᵢ* (17 features: 16 dinucleotides frequencies and *%G+C*), *miPredᵢᵢ* (12 features; *MFEI₁*, *MFEI₂*, *dP*, *dG*, *dQ*, *dD*, *dF*, *zP*, *zG*, *zQ*, *zD*, and *zF*), and *miPredᵢᵢᵢ* (9 features; a subset of *miPredᵢᵢ* that excludes *dQ*, *dD*, and *zD*).

Table S7. Putative viral-encoded *pre-miRs* having maximum length (≤ 95-nts), minimum size of terminal loop (≥ 3-nts), minimum free energy of folding (≤ -25 kcal/mol), and *miPred* scores ≥0.815 (except for †*ebv-mir-BHRF1-1* and §*mghv-mir-M1-8*). They are categorized according to *E. Barr virus*, *K. sarcoma-associated herpesvirus*, *M. γ-herpesvirus 68 strain WUMS*, and *H. cytomegalovirus strain AD169*; sorted in descending *miPred* scores.

S	SP	L	*E. Barr virus* (EBV; AJ507799.2); 5' → 3'
+	147303	92	CCAGAGGAGUGUCCCGGGGCCACCUCUUUGGUUCUGUACAUAU**uuu**GUUAUUGUAACAUAACCAUGGAGUUGGCUGUGGUGCACUCCAUCUGG (*ebv-mir-BART10*)
+	7681	94	AUAUAGAUUAGGAUAGCAUAUGCUAUCCAGAUAUUUGGGUAGUA**uaugc**UACCCAGAUAUAAAUUAGGAUAGCAUAUACUACCCUAAUCUCUAU (*ebv-mir-p1*)
+	140016	92	UGACCUUGUUGGUACUUUAAGGUUGGUCCAAUCCAUAGGCUUUUU**uugu**GAAAACCCGGGGAUCGGACUAGCCUUAGAGUAACUCAAGGCCA (*ebv-mir-BART6*)
+	7693	95	AUAGCAUAUGCUAUCCAGAUAUUUGGGUAGUAUAUGCUACCCAGAU**auaa**AUUAGGAUAGCAUAUACUACCCUAAUCUCUAUUAGGAUAGCAUAU (*ebv-mir-p2*)
+	7932	94	GCAUAUGCUACCCAGAUAUAGAUUAGGAUAGCCUAUGCUACCCAG**aua**UAGAUUAGGAUAGCAUAUGCUAUCCAGAUAUUUGGGUAGUAUAUAGC (*ebv-mir-p3*)
+	9007	91	UAGGACCCUUUUACUAAACCCUAAUUCGAUAGCAUAUGCUUCCC**guu**GGGUAACAUAUGCUAUUGAAUUAGGGUUAGUCUGGAUAGUAUAUA (*ebv-mir-p4*)
+	7708	95	CAGAUAUUUGGGUAGUAUAUGCUACCCAGAUAUAAAUUAGGAUAG**cauaua**CUACCCUAAUCUCUAUUAGGAUAGCAUAUGCUACCCGGAUACAG (*ebv-mir-p5*)
+	146422	94	GGAUCCAGUGUCCUGAUCCUGGACCUUGACUAUGAAACAAUUCU**aaa**AAAAUGCAUCAUAGUCCAGUGUCCAGGGACAGUGCACUCGGAAGUCU (*ebv-mir-BART7*)
+	9031	95	UCGAUAGCAUAUGCUUCCCGUUGGGUAACAUAUGCUAUUGAAUUAGG**guuag**UCUGGAUAGUAUAUACUACUACCCGGGAAGCAUAUGCUACCCG (*ebv-mir-p6*)
+	152730	92	CUGGUGGACUUCCAGACUAUUUUCUGCAUUCGCCCUUGCGUGU**ccauu**GUUGCAAGGAGCGAUUUGGAGAAAUAAACUGUGAGUUUCACAG (*ebv-mir-BART2*)
+	146753	95	GGUCGAUGGGUUCACUGAUUACGGUUUCCUAGAUUGUACAGAU**gaacuag**AACUGUCACAAUCUAUGGGGUCGUAGACAGUGUGCUUACCAGACU (*ebv-mir-BART8*)
+	42832	95	AAUGACCCGGCCCCCACUUUUAAAUUCUGUUGCAGCAGAUAGCUGAU**accc**AAUGUUAUCUUUUGCGGCAGAAAUUGAAAGUGCUGGCCAUAUCU (*ebv-mir-BHRF1-2*)
+	139064	93	AGGCAUUGUUAACCUUUGGUGGAACCUAGUGUUAGUGUUGUGCUGU**aaau**AAGUGUCCAGCGCCACCACUAGUCACCAGGUGUCACCGGAGGCU (*ebv-mir-BART3*)
+	139898	95	AACAGGAUGUGGCACCCUAAGAGGACGCAGGCAUACAAGGUU**auuaccc**AGUCCUUGUAUGCCUGGUGUCCCCUUAGUGGGACGCAGGCCUAGGU (*ebv-mir-p7*)
+	12549	93	GGCAGAGGGUCGGCCUAGGCCCGGGGAAGUGGAGGGGGAUC**gcccg**GGUCUCUGUUGGCAGAGUCCGGGCGAUCCUCUGAGACCCUCCGGGCC (*ebv-mir-p8*)
+	139206	95	GGGGCUCUGUAACAUUUGGUGGGACCUGAUGCUGCUGGGUGUGCUGU**aaau**AAGUGCCUAGCACAUCACGUAGGCACCAGGUGUCACCAGGGCUAC (*ebv-mir-BART4*)
+	42950	95	UAUACGCCUGUGGGUGUUCUAACGGGAAGUGUGUAAGCACAC**ACGUAauu**UGCAAGCGGUGCUUCACGCUCUUCGUUAAAAUAACACAAGGACAAG (*ebv-mir-BHRF1-3*)
+	7754	94	AUAUACUACCCUAAUCUCUAUUAGGAUAGCAUAUGCUACCCGGAU**acag**AUUAGGAUAGCAUAUACUACCCAGAUAUAGAUUAGGAUAGCAUAU (*ebv-mir-p9*)
+	156856	94	UUUUGCGCCUGGAAGUUGUACUCCCGGAAGAUGCCCUCCAGGUCAA**agacg**UUGGAGGCACGCUGUUCGUCCCGUGAGUACAGCUCCAGGGAGG (*ebv-mir-p10*)
+	140356	94	CUGGAGACCUGCUAUGUGGCUAGACGUAUGGCCUACCCAAGACGU**ugg**GGGUCUCGGGUAGGCCAUGAUUCUUCCAGGCAUAGGUUACAACCAG (*ebv-mir-p11*)
+	146941	94	UGUGGCAGCUGUUGUUUGUACUGGACCCUGAAUUGGAAACAGUAACU**ugg**AUUCUGUAACACUUCAUGGGUCCCGUAGUGACAACUAUGCUGAA (*ebv-mir-BART9*)
+	139778	95	GCUUUCAGGUGUGGAAUUUAGAUAGAGUGGGUGUGUGCUCUUGUUU**aauu**ACACCAAGAUCACCACCCUCUAUCCAUAUCCCACAAUUGAUAAAC (*ebv-mir-p12*)
+	165115	94	UUCUUGGGUGAGCGAGUCACCCUGACCUCCUACUGGAGGAGG**gugag**CCUCGGUCCAGAGAUUGAGGUCAGCUGGUUUAAACUGGGCCCAGGAG (*ebv-mir-p13*)
+	139333	95	UAACAAACCCGUGGGGGGUCUUAGUGGAAGUGACGUGCUGUGAAU**acag**GUCCAUAGCACCGCUAUCCACUAUGUCUCGCCCGGCUAUAUGUCG (*ebv-mir-BART1*)
+	139658	93	GAUGCUCUGUGGCACCUCAAGGUGAAUAUAGCUGCCCAUCGACGUAUC**gcug**GAAACCGGUGGGCCGCUGUUCACCUAAAGUGACGCAAGGUC (*ebv-mir-BART5*)
+	140396	87	GACGUUGGGGGUCUCGGGUAGGCCAUGAUUCUUCCAGGCAUAGGU**uac**AACCAGUCACUGCUAUCAAGCCUACUCAGUUCCCAACGC (*ebv-mir-p14*)
+	7666	94	GCAUAUGCUACCCAGAUAUAGAUUAGGAUAGCAUAUGCUAUCCAG**auau**UGGGGUAGUAUAUGCUACCCAGAUAUAAAUUAGGAUAGCAUAUAC (*ebv-mir-p15*)
+	165097	95	GGCCAGGCUGUCACCGCUUUCUUUGGGUGAGCGAGUCACCCUAGCCUC**uacu**GGAGGAGGGUGAGCCUCGGUCCAGAGAUUGAGGUCAGCUGGUU (*ebv-mir-p16*)
+	48951	93	GACCGUGGCUCCCGCCUCUUUGGAUGCCAUCAUCCCUGCUUUGG**gac**CCGACCGCAUCUGGCAUGCGGCCGGUAGGUGUCGGGGGGGUGACGGUC (*ebv-mir-p17*)
+	103504	95	UGGCAGAGCUUUCACCGGUGGAACUCGUGACAGAUGUCAUCGCCACC**cua**GGCACGUGGGAGAUCAUCGACGAGCUCUACCGGAGCAGUCGCCUG (*ebv-mir-p18*)
+	68250	94	CUGAGUGUGGGGGCCAUACGAGGCCUUCACUGGCCCUGUGGCC**aaggcuca**GGACGUGGGGGCCGUUGAGGCCCACGUUGUCUGCUCGGUAGCAG (*ebv-mir-p19*)
+	74472	94	UUGUGGCACAAACAAAACAGGCGGAAGCCCUCGUCAGGCCG**cgagagga**UGGCAUCGAGGAUGGCCUCCGCAAUGUCAGUGUUUGAGGCCACAA (*ebv-mir-p20*)
+	153549	95	UGUAGGCUGAGAGCUUGCGGCUGAGCUCCGUUGAAAAGCAGAGCUCCCC**cau**GGGGACCCUGCCUUCACGGAGGUCUGUGUAGGCCUGGUUUAGG (*ebv-mir-p21*)
+	145656	87	GAGUGGGGGAUGCUAGCCAAUUUAGCUUCCCCUCCCCUU**aac**AGGGGGUCUCGCGGGGUGCCAAUUGUCGCCUGCCUUCCCCCGCUU (*ebv-mir-p22*)
+	132971	95	CCGUGCUGGGCAGUCAGGGCCUGGAAGUCUUGGCGGCGUUGGUAUU**uaa**AAACCAGCGAUCCCUGAGAACGCUCCAGGUAGAGUUUCCAGCCCUG (*ebv-mir-p23*)
+	41458	94	AAGGACGGCUCCUUAUUUAACCUGAUCAGCCCCCGGGAGUUGCCGU**uuc**AUCACUAACCCCGGGCCUGAAGAGGUUGACAAGAAGGGUCAAGGUUU (†*ebv-mir-BHRF1-1*)

S	SP	L	*K. sarcoma-associated herpesvirus* (KSHV; U75698.1); 5' → 3'
−	119293	93	UCCAGUAGGUAUACCCAGCUGGGUCUACCCAGCUGCGUAAACCC**cgcu**GCGUAAACACAGCUGGGUAUACGCAGCUGCGUAAACCCGGCUGGG (‡*kshv-mir-K12-9*)
−	119273	94	UGGGUCUACCCAGCUGCGUAAACCCCGCUGCGUAAACACAGCUGG**uaua**CGCAGCUGCGUAAACCCGGCUGGGUAAAUCCAGCUGUAAUUCUA (∆*kshv-mir-K12-9*)
−	120744	94	GCGGGUUUAGAAAGACUUGUCCAGCAGCACCUAAUCCAUCGGC**ggucgg**GCUGAUGGUUUUCGGGCUGUUGAGCGAGUCUUUUUAUCUAGCGC (*kshv-mir-K12-6*)
−	121535	89	GCCUGUGAUGGGCUAUCACAUUCUGAGGACGGCAGCGACGUG**ugutc**UAACGUCAACGUCGCGGUCACAGAAUGUGACACCCUCCAGGU (*kshv-mir-K12-3*)
−	23628	94	GUUUAAUUAUAGAAUUGCAGCUGGGUAUACCCAGCUGGGUUCACCC**accu**GGGUAUACCUACUGGAUUACCCAGCUGGGUAUACCUACUGGAAU (*kshv-mir-p1*)
−	119320	93	GCGCAGCUGAGUCAUCGCAGCCCCUAUUCCAGUAGGUAUACCC**agcu**GGGUCUACCCAGCUGCGUAAACCCCGCUGCGUAAACACAGCUGGGU (*kshv-mir-p2*)
−	121400	95	GAACCGGGCAGUAUAACUAGCUAAACCGCAGUACUCUAGGGC**auucauu**UGUUACAUAGAAUACUGAGGCCUAGCUGAUUAUACUACCUCCGUCC (*kshv-mir-K12-4*)
−	81073	86	UACCCAGUUUGUCAUGACACCGACAGAAGCUGGUGCUGGCCGA**cgucu**UCGCCGCGCCACUCGUCCGGUGGACAGGCUGAUUUGAAA (*kshv-mir-p3*)
−	133793	94	AUUAUGCAGGCUGUUAUGAUAUCCCUGGGGGCAGACCUGCUACCG**cugg**CGGUGCAGGCUUCAACCGGGGACAAUUAUAACGUGGCCAGGUACU (*kshv-mir-p4*)
−	79458	93	AGGACGCUGACGUUGGGGGCCCCCGUAAGGACGUCGGCGAUCGUC**ucg**GCGCUGUCGCCACUCGUACAAAAAAUAACCCUUACUGUCAGCGCCU (*kshv-mir-p5*)
−	9635	95	GCCGUGAUCUCGUUGGCCACAAAGUGGAAGCUGUCCUCGUGGGUAGU**ucgg**AUGGAGCGCGGGAAGGUUUUUCCACAGUGCCAGCGGACACACGGC (*kshv-mir-p6*)
−	77286	95	GUAUAUCUUUGUUUUUUUCAAGUUUGUGGACGAGGUGGUCCAUGCAUA**gac**UGGCAUGUGAUUAUCUGCACAAGCGCUGACGAAAGCUAUGGUUUU (*kshv-mir-p7*)
−	121856	95	GCAGGGUGCGGUGCUGCCCAGGACAGGCCCGGAUGCGGGCGAUUACAG**gaaa**CUGGGUGUAAGCUGUACAUAAUCCCCGGCAGCACCUGUUUCCUGC (*kshv-mir-p8*)
−	108693	94	UGGAGUGUGGAUGGUGAUAUGGUCUCCUGGGCCUGGCCGGCCACC**gu**CUAUGGUGGCAACAGGAGGCCGGCCUGCUCCGCUUCCUGUACCA (*kshv-mir-p9*)
−	120342	95	GCGCAUAUUGGCGUUGAGCGCCACCGGACGGGGAUUUAUGCUGUA**ucu**UACUACCAUGAUCCCAUGUUGCUGGCGCUCACGGCCCGUGUGCCAGC (*kshv-mir-K12-7*)

S	SP	L	*M. γ-herpesvirus 68 strain WUMS* (MGHV68; U97553.2); 5' → 3'
+	1320	59	AACCACCUCCCACAAUUUCAGAGUCU**uagcc**AGAUUAUCUGAAACUGUGUGAGGUGGUU (*mghv-mir-p1*)
+	636	67	ACGAAGUAGCGAACCUCUGCUCACUGCCCGGG**ccc**UCCGGGAGGUGAGCAGGAGUUGCGCUUUUCUU (*mghv-mir-M1-3*)
+	739	93	CACGCUGCCAAUCUCACCCGACAGCUGUCAGGGGUUACAUGA**agaac**UUCAUGUAACCCCUGACAGCUGUCAACCUAAUCCUGACCGUGAG (*mghv-mir-p2*)
+	104268	94	CAGCUAACUGGGUGUUGAGAGUACAUGUUUGCUUUGGAUACACUU**gug**AAGUUUAUUCAAAGUGUAGGGAUGUGUGCUACUAAACAUAACAGCUG (*mghv-mir-p3*)
+	548	91	CCCGAGCCCUGGUUGAGAGGGGGAGUGUGUGGUCUGUAGAGAC**auga**GUUGAUCGGCAGACCCCCUCUCCCCUCUUUCCCUCUUUACG (*mghv-mir-M1-2*)
+	107005	94	UCUUUAGCAGACAGGUUAGAGCACGUUUGUGUGAUGUGAGAGGAGU**aagu**GUGUCUCCACCAUCGCAUAACAGUUGAUAGGUGGGCUUUAAAGA (*mghv-mir-p4*)
+	112453	95	GUUCUAUGGUACCAACAGACUCUUGUGUUUCUUGAAUGGUUCCAG**uuucauu**CUGGACAAGUAAGAAUACUUGAUCUGUUGUCACAUUAAGGAAU (*mghv-mir-p5*)
+	112662	95	GUGAGUAUUUCUUGGAUGGAGCACCUGACCUGUGGCAUCAUGGAC**cggau**GUAACAGGCGUGAGAAAGGUCUUUGGUCCAUUCUUGUAAUACUCU (*mghv-mir-p6*)
+	3794	94	UGUGAGCUCUUCUUUUACCAGCACUCACUGGGGGUUUGGUCAGGAGAUC**aagua**GAUCUGACCAACCCUAAGUGAGUUUUUCUUCUUGCUUAACA ([§]*mghv-mir-M1-8*)

S	SP	L	*H. cytomegalovirus strain AD169* (HCMV; X17403.1); 5' → 3'
−	49486	94	GCAAGGUAAGCCCCACGUCGUUGAAGACACCUGGAAAGAGGACGUUC**gcuc**GGGCACGUUCUUUCCAGGUGUUUUCAACGUGCGUGGAUUUUUU (*hcmv-mir-UL36*)
+	49484	94	AGAAAAAAUCCACGCACGUUGAAAACACCUGGAAAGAACGUGCC**cga**GCGAACGUCCUCUUUCCAGGUGUCUUCAACGACGUGGGGCUUACCUU (*hcmv-mir-p1*)
−	174048	95	AUUUGACGUCAAUGGGGUGGAGUAUUUACGGUAAACUGCCCACUUGG**caguaca**UCAAGUGUAUCAUAUGCCAAGUACGCCCCCUAUUGACGUCAAU (*hcmv-mir-p2*)
+	203097	95	UCUUCGGAAACUGUGGACGCUGUUUCCGAAUACCGGGAGGA**aucgugcuucc**CUCUUCCAAGGAUCGGAAAGUAGCGUCCGUCGUUUCCGCGGA (*hcmv-mir-p3*)
+	93409	94	GCCGCGGAAUGGACGGGACCCGGGGUCCGCGCCCUUCCCCUCC**ccccac**GGGGGGCUGGGUCGCGGACCCCGGUUCCUAGGCUCGUUCCGCGGU (*hcmv-mir-p4*)
−	155177	78	AAAGGACGACCCGUCUCCCCCCCGCACCCGGGUUUUUU**cuuu**GGUCGAACCCGGCUUGCGACGACGGGUUGUUCCUUU (*hcmv-mir-p5*)
+	27628	95	GUUUCUUCCCAUAGCCUGUCUAACUAGCCUUCCCGUGAGAGUUUAU**gaac**AUGUAUCUCACCAGAAUGCUAGUUUGUAGAGGCUAUGCGGGAUGC (*hcmv-mir-UL22A*)
−	35809	94	CAGAAAUAGGGCGACGGUGUUUUUAUAACCGAAAGUAGCGUGUUU**gag**ACACGCGCUUCUGGUCGGUUUUUUUCACCGUCGUCGCUCUAGGUUUG (*hcmv-mir-p6*)
−	147717	94	ACGUGCACGGUGAAAGUGGCGUCGUCGCUCGGCGGGGUGCGCACC**gcc**GGUGCUGCUGCUGACUUCCACGACGUUGUUUUCACCGUCGCCGUCGU (*hcmv-mir-p7*)
+	38054	87	CUCGUCAGCUUCACGGAGCUGUUGUUACCGCCGCCGU**ccgucgcc**GCCGCUGCGGUGGCGGCGACAGCGACGAGCGAGGUGGGCGAG (*hcmv-mir-p8*)
−	65216	95	AGUACCUGCUCGACGACGCGUUCCAUCUGCUUCAGGUCCUCU**accggcaaa**AAGCCGUUAAGGAUGUGAUUGUGCACGCGCGUCAGCAGCUGCGU (*hcmv-mir-p9*)
+	116589	91	AUCACCGGCCUAUUACCGUCGGCGCGACUCUCCGGGCGGUAUGG**augaa**CCACCGUCCGGAUGGGAGCGUUACGACGGUGGUCACCGUGGU (*hcmv-mir-p10*)
−	7091	90	UUCGUCCCGUCUCCCUCGUGGUCGUGGGUGGUGCGAGAGUACAC**gaugg**GUGGCUCUCGUCUCGGGGGACCACAGGGGGAGGGGGGUAA (*hcmv-mir-p11*)
−	25058	95	ACGCCGGUUUCAUCAUAAACCACCGUGAGAACCGGCGCGGGUUU**aacac**GAAACCGCGUCACUCACGGACGUAGGUUAUUUCGAAAACCUACGU (*hcmv-mir-p12*)
+	174048	95	AUUUGACGUCAAUAGGGGGCGUACUUGGCAUAUGAUACACUUGAUG**uac**UGCCAAGUGGGCAGUUUACCGUAAAUACUCCACCCAUUGACGUCAAU (*hcmv-mir-p13*)
−	194927	94	GUACGGUGUCGCCACCGUUGACGUGGGCGGCGAUGAGAACGUCA**gcgg**UGGCGAAACCGCCGUGCGGAAAGUCCCGGUGCCGAAAUCACCGUGU (*hcmv-mir-p14*)
−	49464	94	GAAGACACCUGGAAAGAGGACGUUCGCUCGGGCACGUU**cuuuccaggguguuuc**AACGUGCGUGGAUUUUUUUCUAUUCUCUACCAGGUGCUUAC (*hcmv-mir-p15*)
+	93410	92	CCGCGGAACGAGCCUAGGAACCGGGGUCCGCGACCCAGCCCCCCGU**ggg**GGGAGGGGAAGGGCGCGGACCCCGGGUCCGUCCAUUCCGCGG (*hcmv-mir-p16*)
−	140853	95	GAAGUUUCGCGGCAGCGCAAGCCGUGGUAACCGUCGCCGUCGGC**gcgc**GGUCGCGCGAGACGACGUGGACGGCGACUCAGGCGACGCAGGUUCUC (*hcmv-mir-p17*)
+	196047	90	UCUGAUCCAACACUGAACGCUUUCGUCGUGUUUUUUCAUGCAGCUUU**uac**AGACCUGACAAGCCUGACGAGAGCGUUCAUCGGGGCAUGA (*hcmv-mir-US5-1*)
+	174117	95	UAAAAUACUCCACCCAUUUGACGUCAAUGGAAAGUCCCUAUUGGCGU**uacu**AUGGGAACAUACGUCAUUAUUGACGUCAAUGGGCGGGGUCGUUGG (*hcmv-mir-p18*)
−	27632	90	CAUCCCGCAUAGCCUCUACAAACUAGCAUUCUGGUGAGAUACAU**guuc**AUAAACUCUCACGGGAAGGCUAGUUAGACAGGCUAUGGGAAG (*hcmv-mir-p19*)
−	52491	89	CAUGUGCGCUCACCCGGCGUUCUGGCCACCGGUUACGCCG**ccaaca**UGGCGUAAUUGACGGUGAGAACUCGGAGACCGAGCGGUCCGUG (*hcmv-mir-p20*)
+	35813	92	CCUAGAGCGACGACGGUGAAAAAACCGACCAGAAGCGCGUGU**cuc**AAACACGCUACUUUCGGUUAUAAAAACACCGUCGCCCUAUUUCUGGG (*hcmv-mir-p21*)
−	90766	95	GGCGGUUCUUUGUGAUUUAAAAACACCGUGUUCGUGAAACGUGA**cuuu**UCACGGUUUGUUAGCUGAUGUGAUUUUGGAGGUCACAAACACCGUAC (*hcmv-mir-p22*)
+	25058	95	ACGUAGGUUUUCGAAAUAACCUACGUCCGUGAGUGACGCGGUUUCG**ugu**UGAAACCCGCGCCGGUUCUCACGGUGGUUUAUGAUGAAACCGGCGU (*hcmv-mir-p23*)
−	25024	94	CGCGGGUUUCAACACGAAACCGCGCACUCACGGACGUAGGUUAUU**ucg**AAAACCUACGUUAAUCCUGAACGCGUUUUGUGUCACGCGUCCCCG (*hcmv-mir-p24*)
+	92228	94	GACGUAGCGAGCGUAGCGAGCUACGUCACGUAUGCGUGCGUCAUC**ccggc**GGAAAUCAUCUCUGAUGACGUAGCGAGCGAAGCGAGCUACGUC (*hcmv-mir-p25*)
+	25023	94	GCGGGGACGCGUGACACAAAACGCGUUCAGGAUUAACGUAGGUUU**cga**AAUAACCUACGUCCGUGAGUGACGCGGUUUCGUGUUGAAACCCGC (*hcmv-mir-p26*)
+	139178	89	GUUACCUGUUGUAUCGCAAGGCUCACGUGGAGCUGUCACUCU**ccagca**ACAAGGUGCAACACGUGGAAGCCGUGCUGCGACAGGUGUAC (*hcmv-mir-p27*)
+	146735	94	CGCGCCAGCUAGGGUGCGCUGGCCUCGGCCGUGACUACGGACGCC**gau**GAGCGUCGGCGCGGCCUAGAGCAGCGUAGCGCCGUGUUGGCGCGCG (*hcmv-mir-p28*)
+	37292	95	GCUUCGCUCUGGAUGGGCUCCGGGUCCGUCAACACGCGACUCGC**gcg**GCAAAAGGCACGCGUGUUGACGGCGCGAGAGCCCGUCGUGAUAGUCCAU (*hcmv-mir-p29*)
+	173784	95	GCCCAUUUGCGUCAAUGGGGCGGAGUUGUUACGACAUUUUGGAAAGU**ccc**GUUGAUUUUGGUGCCAAAACAAACUCCCAUUGACGUCAAUGGGGU (*hcmv-mir-p30*)
+	25076	94	ACCUACGUCCGUGAGUGACGCGGUUUCGUGUUGAAACCCGCGCCG**uucu**CACGGUGGUUUAUGAUGAAACCGGCGUUGGGGAUCUACGCGGGU (*hcmv-mir-p31*)
−	134672	93	AUCGUCAGCGAACCGCGCUUCAAACGCCAGAUCCGAAUACAGGUGC**guuuc**CAUAUUCGGAACGCAUCUGUUUCAGAAGCGCGUCCUCGCGCU (*hcmv-mir-p32*)
+	32963	95	GGCCUUGCGGCGGCAGCGGUUGGCGUGGUUGCUCAGCUCGGCGUC**cga**GAGCGCCGAGCUGAACUGCGGCAGCCGCGUGCGAUCCUGCGGCGCGU (*hcmv-mir-p33*)
−	90873	95	GGCGGAGCGCAGCGAAAUCGGUGGUGAUAGCGGCGAUUGAGGUUGC**gaga**CCAGAUUCAUCGCGCCUUGUACCACCGUGGUGCGGUGUUUCUGCU (*hcmv-mir-p34*)
−	162576	95	CGGAGCGAGGGUCGUUGUCCUCCUCGUCCUCGUGGCGGUUGUUU**ucc**GUCGCGAUCUCCGAGAGGAGGAGGACGACGACGAUGCAGCCUGCCG (*hcmv-mir-p35*)
+	30965	94	CCAGAGCCGUUCGGGGCGUGCGGCCGCGCUAGCGCUUCAUUUUCU**cacguc**ACGAAAAGGAGUGACGGACGGCCAGUACGCCACGUCUCUGCGG (*hcmv-mir-p36*)
+	222717	95	UUACUCUCGAGUGCGGUCGGUGUCUCGUCGGUGAGACGAGGCCGC**gcccc**GACAAGUUCGAUCUCAUGUCGCUCUUGGAGCGCGAAGAGAGUUGG (*hcmv-mir-p37*)
−	210170	94	CGCUGCUUUCGCAUGCCCAAGUUCUUUCCGCCGCCCAUGUGCCGCGU**ucc**GUACAACGAAUGCGGCGUCGAAUUACCGGGCGGCGAUAGCAGCG (*hcmv-mir-p38*)
−	203097	95	UCCGCGGAAACGACGGACGCUACUUUCCGAUCCUUGGAAGAGG**gaagcacga**UCUCCUCCCGGUAUUCGGAAACAGCGUCCACAGUUUCCGAAGA (*hcmv-mir-p39*)
−	174118	92	AACGACCCCCGCCCAUUGACGUCAAUAAUGACGUAUGUUCCCAU**agua**ACGCCAAUAGGGACUUUCCAUUGACGUCAAUGGGUGGAGUAUUU (*hcmv-mir-p40*)
+	93225	94	CCCGCUCGACCCCCCCAUCCGACGGCCCGGCCGGGCUGGGACC**ccc**GCACCGGGGUCCCGGUUCCCGUCCGUGGCCCGGGGGGACCCGAGCGGG (*hcmv-mir-p41*)
+	20175	95	ACCCGCUGGGAGGAAAGCAACGUCGUGAGCCAGACGGCCACGC**gagua**CGUACGUGGUUCGUGGGAAAGAACCACGUUUUGGCGUCGCACGUGGGU (*hcmv-mir-p42*)
+	163175	94	UCCCGGCGCUCUGACAGCCUCCGGAUCACAUGGUUACUCAGCGCU**gcc**AGCCUAAGUGACGGUGAGAUCCAGGCUGUCCGUGCACCACGGUGA (*hcmv-mir-UL112*)
−	174625	92	GCCGAUGUGAGUUUCUGUGUAACUGAUAUCGCCAUUUUUCCAAAAG**uga**UUUUUGGGCAUACGCGAUAUCUGGCGAUAGCGCUUAUAUCGUU (*hcmv-mir-p43*)
+	124992	95	CGCGGUUUACGUAGGCUACGCAGGUAUUUGACGUGUAAC**ccagacccau**GUCUACGGUGUUUAAUGUUCUCGCGUGACGUGGUACGUAGUGCUGAUG (*hcmv-mir-p44*)

−	119625	94	AUCCUCGGCGACGGCGUGCACGUCGGGCGUUAUGACACGCGGCC**gccuuaa**GGCCGAGUCCACCGUCGCGCCCGAAGAGGACACCGACGAGGAU (*hcmv-mir-p45*)
−	36838	94	UCCUCUGCCUGGGCACGCGCGUCGGCCGCGUCGCAAACGCUGCUUGG**uac**CCGAGGUCUUUUGCACGCGCGACUUGGCCGACCUGUGCGUGCGA (*hcmv-mir-p46*)
−	197467	90	GUGGGUGCCCACGGACUUGGACCAUCUCACUCUGCAUUUGGUGC**cgu**GCACCAAAUGCAAACCCAUGUGGUGCCAGCCUCGGUACCAUAU (*hcmv-mir-p47*)
+	119625	94	AUCCUCGUCGGUGUCCUCUUCGGGCGCGACGGUGGACUCGGCCUU**aag**GCGGCCGCGUGUCAUAACGCCCGACGUGCACGCCGUCGCCGAGGAU (*hcmv-mir-p48*)
+	147719	93	GACGGCGACGGUGAAAACAACGUCGUGGAAGUCAGCAGCAGCACC**ggc**GGUGCGCACCCGCCGAGCGACGACGCCACUUUCACCGUGCACGUU (*hcmv-mir-p49*)
−	194965	93	UGACGUGACUCUUGACGUUUAUAAACCGCAUGGGAAAGUACGGUG**ucgc**CACCGUUGACGUGGGCGGCGAUGAGAACGUCAGCGGUGGCGAAA (*hcmv-mir-p50*)
+	128612	95	ACUGGGUCGUCUGUUACUGGGACCCGUGGCCGUACCCUGUUUUUG**cga**CGGUGAAGUGGAGGGCCACGGUGAACAUCUGGUACCUACGACGCAGU (*hcmv-mir-p51*)

S (+/− strand), *SP* (start position), and *L* (length of the putative *pre-miRs*). 25 true-positives and 1 false-negative match 25 published *pre-miR* sequences (red regions) and their mature miRNAs (underlined regions) as obtained from miRBase 8.2 (Griffiths-Jones *et al*., 2006); predicted terminal loop $\geq$3-nts (bold lowercase nucleotides). [‡/Δ]*kshv-mir-K12-9* are the accepted and incorrect positives of *kshv-mir-K12-9*. [†]*ebv-mir-BHRF1-1* (0.437 *miPred* score) and [§]*mghv-mir-M1-8* (0.658).