The sufficient minimal set of miRNA seed types

Daniel C. Ellwanger¹, Florian A. Büttner¹,*; Hans-Werner Mewes¹,²
and Volker Stümpflen¹,*

¹Institute of Bioinformatics and Systems Biology (MIPS), Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuhberg and ²Chair of Genome-oriented Bioinformatics, Technische Universität München, Center of Life and Food Science, D-85350 Freising-Weihenstephan, Germany

Supplementary Information: Supplementary data are available at Bioinformatics online.

Received on October 18, 2010; revised on March 15, 2011; accepted on March 17, 2011

ABSTRACT

Motivation: Pairing between the target sequence and the 6–8 nt long seed sequence of the miRNA presents the most important feature for miRNA target site prediction. Novel high-throughput technologies such as Argonaute HITS-CLIP afford meanwhile a detailed study of miRNA:mRNA duplicates. These interaction maps enable a first discrimination between functional and non-functional target sites in a bulky fashion. Prediction algorithms apply different seed paradigms to identify miRNA target sites. Therefore, a quantitative assessment of miRNA target site prediction is of major interest.

Results: We identified a set of canonical seed types based on a transcriptome wide analysis of experimentally verified functional target sites. We confirmed the specificity of long seeds but we found that the majority of functional target sites are formed by less specific seeds of only 6 nt indicating a crucial role of this type. A substantial fraction of genuine target sites are non-conserved. Moreover, the majority of functional sites remain uncovered by common prediction methods.

Contact: florian.buettner@helmholtz-muenchen.de; v.stuemplfen@helmholtz-muenchen.de

© The Author(s) 2011. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Ago HITS-CLIP sites were located almost at the center of an average Ago-target site. Thus, we got three supersets composed of seed matches of different seed match types with their superset. Due to the hierarchical structure of seed matches, the longest possible type was graded. To reduce unnecessary complexity, we determined all sites complementary to a minimum of six nucleotides beginning at either position one, two or three relative to the 5′-end of the miRNA. Seed matches were classified functional or non-functional by means of their distance to Ago HITS-CLIP sites. To account for all seed start positions, each seed match located within a distance of two nucleotides to an Ago HITS-CLIP site was tagged functional. Since the Ago HITS-CLIP sites were located almost at the center of an average Ago-mRNA footprint, matches located within a distance of 3–31 nt could also be functional. Since the locations of the footprints were not available, an unambiguous classification was not feasible. To avoid false positives, these sites remained unclassified. All seed matches located beyond the footprint (distance > 31) were classified as non-functional.

Further, two miRNAs whose target sites were not significantly enriched in the footprints were removed. Ago HITS-CLIP included 20 miRNAs, whereas 18 of which are broadly conserved (according to Friedman et al., 2009). We proceeded with conserved miRNAs. All miRNA and miRNA data were downloaded from UCSC (Karolchik et al., 2004) and miRBase (Griffiths-Jones, 2010) on October 2010. Based on the set of conserved miRNA sequences and mRNA 3′ UTR sequences, we determined all sites complementary to a minimum of six contiguous nucleotides beginning at either position one, two or three relative to the 5′-end of the miRNA. Seed matches were classified functional or non-functional by means of their distance to Ago HITS-CLIP sites. To account for all seed start positions, each seed match located within a distance of two nucleotides to an Ago HITS-CLIP site was tagged functional. Since the Ago HITS-CLIP sites were located almost at the center of an average Ago-mRNA footprint, matches located within a distance of 3–31 nt could also be functional. Since the locations of the footprints were not available, an unambiguous classification was not feasible. To avoid false positives, these sites remained unclassified. All seed matches located beyond the footprint (distance > 31) were classified as non-functional. Further, two miRNAs whose target sites were not significantly enriched in the footprints were removed from the dataset (Supplementary Table S2). Finally, we got 7342 functional, 64689 non-functional and 1755 unclassified seed sites. Verifying a required minimum target site length of 6 nt, we determined all 5mer matches. The frequency of seed matches within a footprint (distance ≤ 31) and beyond of it was calculated for each seed match length. Additionally, to support our results we prepared the data of the PAR-CLIP experiment in a quite similar fashion (Harfen et al., 2010) (Supplementary Material).

We defined the background set f of the functional and non-functional sites. A seed match s ∈ f was distinguished by its start position relative to the miRNA 5′ UTR (1 = a, 2 = β, 3 = γ) and its length. The outcome of this were 20 match types S for a length k and a start position type p. The distributions of all seed match types were disjoint that is each seed match was graded by the longest possible type. To reduce unnecessary complexity of the seed match type set, we merged iteratively non-significant seed match types with their superset. Due to the hierarchical structure of the set of seed matches, we were able to apply a separate-and-conquer algorithm (Supplementary Algorithm S1). First we divided the target sites by their seed match start position. Thus, we got three supersets composed of seed matches of a minimum length of 6 nt containing all seed types: S, S, S. These sets were separated into 5mers having a mismatch at their subsequent position (S, S, S) and seed matches having a minimum length of 7 nt: S, S. We tested the null hypothesis stating that the distribution of functional and non-functional target sites is independent of a mismatch at the 5′ most subsequent position of a seed match. Thus, if the proportions of functional to non-functional target sites between the S, S, S and S, S seed types were not significantly varying (P > 0.05), the separation terminated otherwise the procedure was continued for the next seed type length. A P-value was calculated by means of a two-tailed Fisher’s exact test. The seed site separation terminated after three steps, the seed match types contained two significant subsets and γ yielded no significant subsets. We termed the found significant seed types based on their start position and their length: S, S. For standardization, we renamed the endmost subsets:

To estimate the specificity of our seed type set, we compared the distribution of the functional sites with a randomized pool of functional seed matches. We drew without replacement a subset of 7803 instances of the multinomial distribution f from functional and non-functional seed matches. A P-value was calculated by means of a χ² test of independence.

The impact of the seed types on miRNA target site prediction was evaluated in terms of recall and specificity. The recall estimates how many of the functional target sites OP are covered by a certain seed type S and the specificity computes the fraction of correctly excluded non-functional target sites ON.

Recall(S,p) = \frac{|\{s \in S_p : s \in OP\}|}{|\{s \in S_p\}|}

Specificity(S,p) = \frac{|\{s \in S_p : s \not\in ON\}|}{|\{s \in S_p\}|}

The recall and specificity of each miRNA target prediction algorithm was determined in terms of pure seed finding. Their seed type selection was assigned as described in the related literature. Due to ambiguous seed type assignments based on the first position of the target sequence, the specificity and recall values for TargetScan were computed by executing predictions on our miRNA set.

To estimate the miRNA seed type usage, we calculated the relative frequencies of a seed type for a certain miRNA. These values were normalized by the mean μ and the SD σ:

Z-score = \frac{x - \mu}{\sigma}

The conservation of each seed site was determined as described in (Betel et al., 2008). We used the software package PHAST (Siegel et al., 2005). The algorithm PhastCons is based on a phylogenetic hidden Markov model, which is fitted to the input sequence by maximum likelihood. Each nucleotide gets a score, which measures the evolutionary conservation across 17 vertebrates. For each seed match, the absolute chromosomal coordinates were determined and a conservation score was calculated. Only if the score of each nucleotide within a functional seed match exceeded the threshold of 0.57 (Betel et al., 2008), the site was tagged conserved in mammals. The background conservation of a seed type was computed by calculating the fraction of conserved nucleotides of a non-redundant set of UTRs holding a specific seed type. For all statistical computations, the R programming language was applied (R Development Core Team, 2010).

3 RESULTS

3.1 Canonical seed types of miRNA target recognition

In this work, we defined a set of canonical seed types by analyzing the seed matches of experimentally verified functional target sites in the 3’UTR. The Ago HITS-CLIP miRNA-mRNA interaction map (mRNA assay of 2006) (Chi et al., 2009) lists 15,665 chromosomal positions of perfectly matching seed sites of length 6–8 nt belonging to 20 miRNAs frequently bound in Ago complexes. We mapped these sites to annotated protein-coding mRNA transcripts and retained sites located within the 3′ UTR. For each miRNA, we scanned the 3′ UTRs of the transcript set for all sites complementary to a miRNA subsequence beginning at either position one (α-position), two (β-position) or three (γ-position) relative to the miRNA 5′-end. We required a minimum length of 6 nt. Seed matches of length five [as reported by (Brembeck et al., 2005)] were not significantly enriched in average Ago footprints (Supplementary Table S1). We classified these sites by means of their distance to an Ago HITS-CLIP site and retained miRNAs significantly enriched in footprints. This resulted in 2369 mouse genes containing 7070 Ago HITS-CLIP sites of 16 broadly conserved miRNAs.

Each contiguous seed match was defined by its start position type and its length. The dataset was composed of eight α-, seven β- and five γ- seed match types (Supplementary Table S3). Following the law of Occam’s razor, the simplest seed type setting for target prediction should usually be the correct one. To reduce unnecessary complexity of the seed type set, we identified unique seed types differing significantly from their superset in terms of functional and non-functional site distribution. For the murine and the human dataset, we achieved six different, disjunct types of seeds: three 6mers either beginning at the first nucleotide (6merα), the second nucleotide (6merβ) or the third nucleotide (6merγ), two 7mers
Table 1. Determined canonical seed types

| Seed type | Functional | Non-functional | LORe | P-value |
|-----------|------------|----------------|-------|---------|
| 6merα     | 1793       | 24 20746       | 32    | −0.12   | 1.20E−02 |
| 6merβ     | 1382       | 19 13500       | 21    | −0.04   | 2.57E−04 |
| 6merγ     | 1755       | 24 17954       | 28    | −0.06   | 2.26E−09 |
| 7merα     | 760        | 10 5036        | 8     | 0.12    | 2.03E−13 |
| 7merβ     | 959        | 13 5250        | 8     | 0.21    | 1.34E−15 |
| 8merα     | 693        | 9 2203         | 3     | 0.44    | 7.60E−12 |

Log odds ratio based on sampling.

Fig. 1. Definition of seed types. The seed types were termed by the start position relative to the 5′-end of the miRNA and the length of the consecutive seed match. The defined set of canonical seed types can be surjectively projected to the seed type set of (Bartel, 2009). Equivalent definitions could be found for 6merα, 6merβ and 6merγ. In the case of miRNAs having a seed sequence beginning with an uracile, 7merα complies with 7mer-A1 and 8merα is equal to 8mer. Otherwise 6merβ equates 7mer-A1 and 6merβ complies with 8mer. If the first position within the target sequence is not an adenine, 8merα equates 7mer-m8 and 7merα is equal to 6mer. Additionally, our set considered 6mer matches that are complementary to the first position of a miRNA seed (6merα). Common target site prediction tools focus on seeds of length seven and eight to increase precision.

3.2 Majority of functional sites are based on 6mer seeds

We examined recall and specificity affected by the individual seed type to functional sites, 6mer seeds make up the highest fraction of true target sites (recall: 0.67). On the other hand, 6mer types either starting at position one (7merα) or position two (7merβ) and one 8mer beginning at the first nucleotide (8merα) (Supplementary Fig. S1A). These canonical seed types terminated within the first 8 nt of the miRNA in 97% of cases. This underscores the importance of the octamer at the 5′-end of the miRNA and the length of the consecutive seed match. The significance of this seed type to functional sites, 6mer seeds make up the highest fraction of true target sites (recall: 0.67). On the other hand, 6mer types do not reflect the ranking of predictions based on the algorithms’ scoring schemes.

in silico target site classification, the usage of a short seed type causes an inverse prediction (Matthews correlation coefficient (MCC) < 0). In terms of in vitro target site classification, the usage of a short seed type causes an inverse prediction (Matthews correlation coefficient (MCC) < 0). In terms of in vitro target site classification, the usage of a short seed type causes an inverse prediction (Matthews correlation coefficient (MCC) < 0). In terms of in vitro target site classification, the usage of a short seed type causes an inverse prediction (Matthews correlation coefficient (MCC) < 0). In terms of in vitro target site classification, the usage of a short seed type causes an inverse prediction (Matthews correlation coefficient (MCC) < 0). In terms of in vitro target site classification, the usage of a short seed type causes an inverse prediction (Matthews correlation coefficient (MCC) < 0).
Table S4 lists the numbers of third of the seed matches are conserved, whereas 7 number of non-conserved sites. In terms of 7mer seeds, about two-thirds of conserved sites in this case is more than three times as high as the avoidance of miRNA regulation. This goes well with our observation that 37% of the 3′UTRs (Supplementary Table S5). The 6mer sites reveal an almost UTs regulated by less effective 6mer matches. The impact of the (default) seed type selection of prediction The majority of functional target sites is conserved (60%). All seed types have a higher fraction of conserved sites than one would expect by chance, given the conservation of their 3′UTRs (Supplementary Table S5). The 6mer sites reveal an almost equal partitioning in conserved and non-conserved sites. A clear discrepancy between the numbers of conserved and non-conserved sites emerges for 7- and 8mer seeds. Particularly, 8mer seed matches exhibit a significant tendency to be conserved. The number of conserved sites in this case is more than three times as high as the number of non-conserved sites. In terms of 7mer seeds, about two-thirds of the seed matches are conserved, whereas 7mer exceeds 7merβ.

In summary, the mean probability to be conserved is about 55% for a 6mer seed. In contrast, 7mer and 8mer seeds have a probability up to 77% to be conserved. Further, a total of 75% of the functional non-conserved sites. Particularly, 8mer seed matches exhibit a significant tendency to be conserved. Therefore, non-conserved or species-specific targeting relies to a large extent on target sites containing short seeds.

Fig. 3. Heatmap showing the seed type distribution for each miRNA. The colors affected by the row z-score indicate the bias of miRNAs to prefer targets holding a specific seed type. A red/blue coloration implies a higher/lower usage of a seed type compared to other miRNAs.

Moreover, we wondered if the seed type distributions differ between functional sites. The relative frequencies of the seed types were computed for each miRNA. A z-score indicates miRNAs holding a frequency over or below the mean frequency given a specific seed type (Fig. 3). Interestingly, 6mer seed types and long seed types are grouped to clusters, respectively, demonstrating that a miRNA either binds to long sites or to short sites but not to both. Further, two main miRNA cluster appeared. The larger group contains miRNAs binding primarily to 6mer-based functional sites. Seven of the 16 miRNAs carry out stronger repression by pairing to rather long seed matches.

The importance of short seed types gains further support by the observation that 37% of the 3′UTRs contain exclusively seed matches of length six in their Ago footprints (Supplementary Table S4 lists the numbers of 3′UTRs containing seed matches of exclusively one type.) Interestingly, the sequences of this subset of 3′UTRs are significantly shorter than that of the superset (t-test, \(P = 4.53E^{-06}\), Stark et al. (2005) studied the impact of miRNA regulation on 3′UTR evolution and found that short 3′UTRs indicate avoidance of miRNA regulation. This goes well with our observation of short 3′UTRs regulated by less effective 6mer matches.

### 3.3 Non-conserved targeting relies on short seeds

We used a strategy established by (Betel et al., 2008) to identify seed sites conserved across mammals (Fig. 4 and Supplementary Fig. S1B). The majority of functional target sites is conserved (60%). All seed types have a higher fraction of conserved sites than one would expect by chance, given the conservation of their 3′UTRs (Supplementary Table S5). The 6mer sites reveal an almost equal partitioning in conserved and non-conserved sites. A clear discrepancy between the numbers of conserved and non-conserved sites emerges for 7- and 8mer seeds. Particularly, 8mer seed matches exhibit a significant tendency to be conserved. The number of conserved sites in this case is more than three times as high as the number of non-conserved sites. In terms of 7mer seeds, about two-thirds of the seed matches are conserved, whereas 7mer exceeds 7merβ.

In summary, the mean probability to be conserved is about 55% for a 6mer seed. In contrast, 7mer and 8mer seeds have a probability up to 77% to be conserved. Further, a total of 75% of the functional non-conserved sites. Particularly, 8mer seed matches exhibit a significant tendency to be conserved. Therefore, non-conserved or species-specific targeting relies to a large extent on target sites containing short seeds.

Fig. 4. Observed and expected fraction of conserved seed matches for each seed type illustrated for functional target sites.

Keeping only the conserved sites from the set of seed matches lifts specificity of all seed types (Fig. 2, Supplementary Table S9). In particular, the 6mer seeds show a significant increase of specificity leading to a classification better than an average random prediction (MCC > 0, Supplementary Table S6).

### 3.4 Target prediction focuses on 7- and 8mer seed matches

We reviewed frequently used approaches for target prediction in mammals with regard to the implemented seed types (Table 2). The TargetScan algorithm (Grimson et al., 2007) seeks mainly for seeds of length seven and eight via seed types 7mer-A1, 7mer-m8 and 8mer. The 7mer-A1 sites may be of type 6merf in the event the miRNA sequence starts with a nucleotide different to uracile. However, the majority of mammalian miRNAs begins with an U (Lewis et al., 2005). Both PicTar (Krek et al., 2005) and EIMMO (Gaidatzis et al., 2007) require stringent seed pairing between 7 nt starting at either the α or the β-position. A novel approach called TargetSpy seed (Sturm et al., 2010) restricts the set of seed matches to predictions containing a perfect 7mer.

Some algorithms allow for custom-defined seed searching: PITA (Kertesz et al., 2007) seeks by default for sites of length six, seven and eight that start at position two of the miRNA. The standard setting of PACMIT (Marn and Vaníček, 2011) is even more restrictive by considering merely sites matching to miRNA positions starting at either the α or the β-position. The latter is a general approach to predict RNA:RNA interactions. Both do not suggest default seed search parameters.

The impact of the (default) seed type selection of prediction algorithms on recall and specificity was evaluated (Fig. 2, Supplementary Tables S8 and S9).

Prediction methods implement scoring schemes to value target site characteristics beside the seed. In contrast to common evaluations of miRNA target site prediction algorithms, this is not an assessment of a subset of top scored instances but of all predictions. Therefore, the denoted specificity values represent the minima while the recall values show the maxima for the (default) seed choice, respectively. Subsets composed of top scored predictions would achieve significantly higher coverage by 6mer seed values.

Obviously, all prediction models exhibit a considerable constraint regarding their ability of finding potential target sites. PITA holds...
the highest recall of 52% (specificity: 60%) owing to the exhaustive search for 6merβ seed matches, whereas PACMIT has the lowest recall of 23% (specificity: 88%) restricted to find less than a quarter of all functional seed sites. Additional filtering by removing conserved sites increases the specificity but consequently lowers the recall. Here, PACMIT could only find 16% of all functional sites (specificity: 75%). A higher recall but a lower specificity can be observed for the prediction of miRNA:mRNA interactions (Supplementary Fig. S2). Concluding, due to the significant gain of precision, tool developers recommend to use long seeds. Our study quantified the loss of recall accompanied by this proceeding.

4 CONCLUSION

In this study, we present an analysis of the most important feature for miRNA target recognition, the so-called miRNA seed, using a large-scale dataset of functional target sites. Based on the Ago HITS-CLIP and PAR-CLIP miRNA:mRNA interaction maps, we analyzed seeds properties and their influences on miRNA target site prediction methods. Due to the definite specification of Ago binding sites, we were able to classify miRNA recognition elements contained in the miRNA 3′UTR as either functional or non-functional. We defined a minimal set of seed types that is sufficient for accurate miRNA target site predictions. The final data pool allows for enhanced analysis of miRNA target prediction algorithms compared to earlier studies that were restricted by experimental constraints (Alexiou et al., 2009; Selbach et al., 2008). We found that most conserved miRNAs interact predominantly with target sites endowed with short seed matches; 67% of functional sites are based on 6mer seeds. In contrast, common prediction algorithms focus mainly on seeds of length seven or eight. At present, prediction algorithms have to accept severe deficiencies of recall to ensure high specificity that is naturally considered to be more important. Moreover, the preferential search for long seeds lifts the proportion of conserved sites. But we found that a substantial fraction (40%) of all functional target sites is not conserved. Target sites including 6mer seeds are enriched among these.

Concluding, the problem of recall can be easily translated to the problem of precision. However, this strongly intensifies the need for features beyond seed pairing that realistically describe miRNA targeting, in particular non-conserved target sites. It may also raise the basic question for the potential of seed-based approaches in discriminating between functional and non-functional sites.

Funding: Helmholtz Association within the Helmholtz Alliance on Systems Biology (project CoReNe); Federal Ministry of Education and Research (BMBF) within project SysMBio (Förderkennzeichen: 031549/A).

Conflict of Interest: none declared.

REFERENCES

Alexiou,P. et al. (2009) Lost in translation: an assessment and prospective for computational microRNA target identification. Bioinformatics, 25, 3049–3055.
Baek,D. et al. (2008) The impact of microRNAs on protein output. Nature, 455, 64–71.
Bagga,S. et al. (2008) Regulation by let-7 and lin-4 mirrors results in target miRNA degradation. Cell, 122, 553–563.
Baes,D.P. (2009) MicroRNA: target recognition and regulatory functions. Cell, 136, 215–233.
Betz,D. et al. (2008) The microRNA resource: targets and expression. Nucleic Acids Res., 36, D100–D102.
Bienzle,J. et al. (2005) Principles of miRNA-target recognition. PLoS Biol., 3, e85.
Busch,A. et al. (2008) Instantaneous efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics, 24, 2849–2856.
Chu,S.W. et al. (2009) Argonaut hits-clip decodes microRNA-mRNA interaction maps. Nature, 460, 479–486.
Das,K.K.-H. et al. (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. Science, 310, 1817–1821.
Friedma,R.C. et al. (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res., 19, 92–105.
GuidaK.D. et al. (2007) Inference of mRNA targets using evolutionary conservation and pathway analysis. BMC Bioinformatics, 8, 69.
Griffiths-Jones,S. (2010) microRNA sequences and annotation. Curr. Protoc. Bioinformatic, Chapter 12, Unit 12.9.1–12.9.10.
Grimson,A. et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell, 27, 91–105.
Guo,H. et al. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature, 466, 835–840.
Hafner,M. et al. (2010) Transcriptome-wide identification of ssa-binding protein and microRNA target sites by par-clip. Cell, 141, 129–141.
Hausser,J. et al. (2009) Relative contribution of sequence and structure features to the miRNA binding of argonaute2L-mRna complexes and the degradation of mRNA targets. Genome Res., 19, 2009–2020.
Karsch-Mizrachi,D. et al. (2004) The UCSC table browser data retrieval tool. Nucleic Acids Res., 32, D493–D496.
Kent,W.J. et al. (2007) The role of site accessibility in microRNA target recognition. Nat. Genet., 39, 1278–1284.
Krek,A. et al. (2005) Combinatorial microRNA target predictions. Nat. Geen., 37, 495–500.
Krüger,J. and Rehmsmeier,M. (2006) Rnahybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res., 34, W451–W454.
Lewis,B.P. et al. (2003) Prediction of mammalian microRNA targets. Cell, 115, 787–798.
Lewis,B.P. et al. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120, 15–20.
Li,P. et al. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature, 433, 769–773.
Mlambo,R.M. and Vaníček,J. (2011) Efficient use of accessibility in microRNA target prediction. Nucleic Acids Res., 39, 19–29.
Nielsen,C.B. et al. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA, 13, 1894–1910.
Pruitt,K.D. et al. (2009) NCBI reference sequences: current status, policy and new initiatives. Nucleic Acids Res., 37, D132–D136.
Rehmsmeier,M. et al. (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Selbach,M. et al. (2008) Widespread changes in protein synthesis induced by microRNAs. Science, 321, 900–902.
Siegel,A. et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res., 15, 1054–1060.
Stark,A. et al. (2005) Animal microRNAs confer robustness to gene expression and have a significant impact on mRna evolution. Cell, 123, 1133–1146.
Stum,M. et al. (2010) TargetSpy, a supervised machine learning approach for microRNA target prediction. BMC Bioinformatics, 11, 292.

Table 2. Default miRNA seed type selection of prediction algorithms

| Algorithm | 6merα | 6merβ | 6merγ | 7merα | 7merβ | 7merγ |
|-----------|-------|-------|-------|-------|-------|-------|
| PITAg     | ✓     | ✓     | ✓     | ✓     |       |       |
| TargetScan| ✓     | ✓     | ✓     | ✓     | ✓     |       |
| PsTar     | ✓     | ✓     | ✓     | ✓     | ✓     |       |
| EIMMO     | ✓     | ✓     | ✓     | ✓     | ✓     |       |
| TargetSpy S.| ✓     | ✓     | ✓     | ✓     | ✓     |       |
| PACMITα   | ✓     | ✓     | ✓     | ✓     | ✓     |       |
| PACMITβ   | ✓     | ✓     | ✓     | ✓     | ✓     |       |
| PACMITγ   | ✓     | ✓     | ✓     | ✓     | ✓     |       |

αConfigurable seed length, default seed types ensure high precision.
βIf miRNA seed sequence starts with an adenine, guanine, cytosine.
γIf miRNA seed sequence starts with a uracile.