FEBS Letters 581 (2007) 4603–4610

Computational analysis of microRNA-mediated antiviral defense in humans

Yuka Watanabe[a,b,1], Atsuko Kishi[a,1], Nozomu Yachie[a,b], Akio Kanai[a,b,c,*], Masaru Tomita[a,b,c]

[a] *Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan*
[b] *Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa 252-8520, Japan*
[c] *Faculty of Environment and Information Studies, Keio University, Fujisawa 252-8520, Japan*

Received 27 June 2007; revised 18 August 2007; accepted 22 August 2007

Available online 31 August 2007

Edited by Takashi Gojobori

Abstract **Recent studies have proposed the interesting perspective that viral gene expression is downregulated by host microRNAs (miRNAs), small non-coding RNAs well known as post-transcriptional gene regulators. We computationally predicted human miRNA target sites within 228 human-infecting and 348 invertebrate-infecting virus genomes, and we observed that human-infecting viruses were more likely than invertebrate-infecting ones to be targeted by human miRNAs. We listed 62 possible human miRNA-targeted viruses from 6 families, most of which consisted of single-stranded RNA viruses. These results suggest that miRNAs extensively mediate antiviral defenses in humans.**

Keywords: MicroRNA; ssRNA-positive strand virus; TargetScan; Bioinformatics; *Homo sapiens*

1. Introduction

MicroRNAs (miRNAs) are short non-coding RNAs, approximately 22 nucleotides long, which are processed as stem–loop precursor transcripts and hybridize incompletely to specific target sites in mRNAs to downregulate mRNA messages [1,2]. Recent studies have reported a large number of computationally screened and experimentally identified miRNAs in various animal and plant species, and a growing number of identified miRNAs have been classified and registered in the Rfam database [3,4]. As the next step, numerous approaches are now being taken to reveal the regulatory networks used by miRNAs. Regulation by miRNAs is complicated because multiple and/or cooperative regulation can occur, and the target-recognition mechanism is complex (e.g., there is incomplete base complementation between miRNAs and their corresponding target sites) [5]. As a result of efforts to elucidate such miRNA mechanisms and roles, many interactions between miRNAs and their target mRNAs have been identified [6]. For the efficient prediction of miRNA targets, a number of software packages, including TargetScan [7,8], miRanda [9,10], RNAhybrid [11], PicTar [12], and DIANA-microT [13], have been developed and applied. In this context, our previous analysis of miRNA target sites predicted in *Caenorhabditis elegans* by using RNAhybrid software suggested that many developmental genes are regulated by miRNAs [14]. Furthermore, recent miRNA target predictions in humans have demonstrated that numerous genes may be regulated by miRNAs [7,10]. The specific downregulation of mRNAs by miRNAs is supported experimentally to some extent [15], suggesting that miRNAs play important roles in gene regulatory networks.

Moreover, some miRNAs encoded by viral species, such as the Epstein-Barr virus (EBV) and human cytomegalovirus, have been identified by experimental studies [16,17], and one computational analysis has suggested the existence of even more viral miRNAs [18]. This suggests that an miRNA pathway exists in viral species as well as in human developmental genes, indicating that miRNA-mediated gene regulation is universal among species. Analyses of host miRNA target sites within virus genomes have recently been launched, and miRNAs have been reported to interact in several ways with viral genes. For example, human miR-122a induces hepatitis C virus (HCV) replication by targeting the 5′ noncoding region of the viral genome [19], and miR-32 restricts accumulation of the retrovirus primate foamy virus type 1 (PFV-1) in human cells [20]. Within PFV-1, the *Tas* protein, which has the ability to suppress miRNA-directed functions in mammalian cells, has been detected, along with miRNA-mediated viral gene regulation [20]. Moreover, a computational analysis to find miRNA target genes in human immunodeficiency virus-1 (HIV-1) has suggested that several genes in HIV-1—*nef*, *vpr*, *env*, and *vif*—are regulated by certain types of human miRNAs [21]. A recent report provided experimental evidence that the miRNA-based RNA-regulating machinery plays a key role in the control of HIV-1 replication [22]. In order to integrate such documented and putative viral target sites of human, mice, rat, and chicken miRNAs, a comprehensive database named ViTa has provided a list of host miRNA targets in viral species [23]. These findings offer a glimpse into the offence and defense of viruses and their host organisms, and they implicate miRNAs in antiviral defense in host cells. However, there is still much to be revealed about the relationship between host miRNAs and infecting viruses.

*Corresponding author. Address: Institute for Advanced Biosciences, Keio University, 403-1 Nipponkoku, Daihoji, Tsuruoka, Yamagata 997-0017, Japan. Fax: +81 235 29 0525.
E-mail address: akio@sfc.keio.ac.jp (A. Kanai).

[1]These authors contributed equally to this work.

Abbreviations: miRNA, microRNA; *H. sapiens*, *Homo sapiens*; ssRNA-positive virus, single-stranded RNA positive-strand virus; dsDNA virus, double-stranded DNA virus

0014-5793/$32.00
doi:10.1016/j.febslet.2007.08.049

In response to our predictions of human miRNA target sites within the genome sequences of various viral species, we show here the patterns interactions between human miRNAs and human-infecting viruses. We constructed a bioinformatics workflow with TargetScan software [8], and initially screened possible human miRNA target sites within the genomes of 228 viruses that infect humans and 348 viruses that infect invertebrate species. Comparison of these results suggested that human miRNAs hybridize to the genomes of human-infecting viruses rather than to the genomes of viral species that infect invertebrates. We supported this suggestion by performing the same analyses with miRNAs encoded in species other than humans. Finally, our prediction revealed 62 viral genomic sequences of 6 families that are the most likely to be regulated by specific human miRNAs. The list included a large number of viruses belonging to the Picornaviridae family, such as echovirus and coxsackievirus. In accordance with these results, we propose that human miRNA regulatory pathways may defend against infections by a wide range of viruses.

2. Materials and methods

2.1. Data set preparation

miRNA sequences were obtained from the miRBase website (http://microrna.sanger.ac.uk/, accessed 2006/11/06) [3,4], a comprehensive registry of published miRNA sequences and their annotations. We obtained 470 miRNA sequences of *Homo sapiens*, 375 of *Mus musculus*, 78 of *Drosophila melanogaster*, 115 of *C. elegans*, and 119 of *Arabidopsis thaliana* in FASTA format. The viral genomic sequences were downloaded in EMBL format from the EMBL nucleotide sequence database (http://www.ebi.ac.uk/embl, accessed 2006/11/07) [25]. We extracted two sets of viral genomes; a ''human-infecting group'' consisting of 228 viruses that infect humans, and an ''invertebrate-infecting group'', consisting of 348 viruses that infect not humans but invertebrate species. The set of genomic sequences within the ''human-infecting group'' was collected according to the EMBL ''natural host'' annotations and/or according to viral species name and inclusion of the term ''human''. The set of sequences of the 'invertebrate-infecting group' was composed according to the host species annotations available in the Universal Virus Database of the International Committee on Taxonomy of Viruses (ICTVdB) (http://www.ncbi.nlm.nih.gov/ICTVdb/,2006/11/07). Of the 576 viruses with extracted hosts, 228 were categorized into the human-infecting group and 348 into the invertebrate-infecting group.

2.2. Use of two measures to evaluate viral species regulated by miRNAs

Two measures—''normalized miRNA targeting scores'' (see below) and observed-to-expected (O/E) values—were used to screen viral genome sequences for viruses that may be regulated by human miRNAs. For each viral genome sequence, the target sites of human-encoded miRNAs were initially predicted with TargetScan software version 1.0 [8], and predicted pairs of miRNAs and their corresponding target sites with binding free energies under the threshold of −20.0 kcal/mol, a threshold value that has been used in several previous miRNA target prediction analyses [8–10,14,21], were further screened for candidates. We first defined the normalized miRNA targeting score of every viral genome within the human-infecting group and the invertebrate-infecting group. To do this, we calculated the sum of the free energies calculated from every pair of miRNAs and predicted target sites within the queried viral genome; we then divided the value by the genome size and multiplied by −100 to give the targeting score. We then screened for those viral genomes that had higher normalized miRNA targeting scores (higher than the cutoff score of 80) and were therefore more likely to be regulated by human miRNAs with stable interactions.

To estimate the importance of each normalized miRNA targeting score compared with the scores obtained by randomized trials (see below), the O/E values of the targeting scores were calculated for every viral genome by the following steps. Each normalized miRNA targeting score derived from a set of actual human miRNA sequences was defined as the observed value (O), whereas each normalized miRNA targeting score derived from randomized miRNA sequences was defined as the expected value (E); the distribution of observed values was divided by that of the expected values to indicate the relative importance of the normalized miRNA targeting score calculated by using actual miRNAs. Randomized trials were achieved by using human miRNAs shuffled in a way that preserved their nucleotide compositions. To generate randomized controls, the normalized miRNA targeting score of these shuffled miRNAs was calculated in the same way as described for the actual miRNAs. We repeated the randomized experiments 100 times to calculate the average O/E values and 95% confidence intervals for candidate extraction. Also, the distributions of O/E values were averaged for 100 experiments to compare the virus-targeting potentials of miRNAs encoded by humans and those encoded by other species. The 95% confidence intervals for these O/E values were also calculated and are shown in the figures as error bars.

2.3. Determination of the best thresholds for the two measures

To compare the performance of separation of the viruses within the human-infecting group and invertebrate-infecting group by using the normalized miRNA targeting score and the average O/E value of the normalized miRNA targeting score, we calculated the overall accuracies (Q) [26] of the two score lines, as below:

$$Q = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

where TP (true positives) is the number of viruses within the human-infecting group that scored above each threshold (80 for normalized miRNA targeting score, and 1.4 for the average O/E value of the normalized miRNA targeting score); TN (true negatives) is the number of viruses in the invertebrate-infecting group that scored below the threshold; FP (false positives) is the number of viruses in the invertebrate-infecting group that scored above the threshold; and FN (false negatives) is the number of viruses in the human-infecting group that scored below the threshold. To efficiently screen for human-infecting viruses that may have been downregulated by their host miRNAs, the best threshold of each score line was determined as the score at which the accuracy was highest.

2.4. Prediction of human-infecting viruses downregulated by their host miRNAs

We used four steps to extract human-encoded miRNAs targeting viruses (Fig. 1). Using TargetScan software version 1.0 [8], we predicted the target sites of human-encoded miRNAs within the viral genomic sequences of the human-infecting group (Step 1). We then screened those miRNA-target pairs that had stable duplex free energy, i.e. below the default threshold of −20.0 kcal/mol (Step 2). Then, human-infecting virus genomes of which the normalized miRNA targeting scores were above a threshold score of 80 (Step 3) and the average O/E values were above a threshold score of 1.4 (Step 4) were further filtered as final candidate human-infecting viruses that may have been regulated by human-encoding miRNAs. For details of determination of the thresholds of the normalized miRNA targeting score and the average O/E value, see Sections 2.3 and 3.1.

2.5. Specificity analyses of human miRNAs targeting human-infecting viruses

To validate our hypothesis that human-infecting viruses would be markedly more likely to be targeted by human miRNAs than viruses that do not infect humans, we conducted the following two analyses. First, normalized miRNA targeting scores and their O/E value distributions were calculated for viral genomic sequences within the human-infecting group and compared with those within the invertebrate-infecting group. Second, to analyze whether the results of the analysis were driven specifically by human miRNAs as opposed to invertebrate or other mammalian miRNAs, we calculated the normalized miRNA binding score and O/E values for the viruses in both human-infecting and invertebrate-infecting group, by using miRNAs encoded in 4 species other than humans, i.e. *M. musculus*, *D. melanogaster*, *C. elegans*, and *A. thaliana*. These miRNAs and their target sites have been well studied. The distributions of the scores obtained by using these miRNAs were compared with that obtained with the human miRNAs in both virus groups.

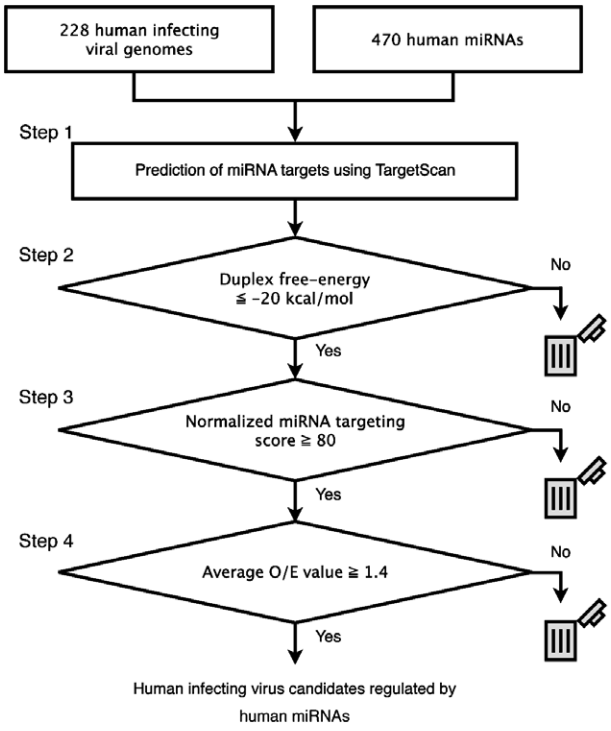

Fig. 1. Computational scheme for the extraction of viruses putatively targeted by human miRNAs. Candidate human miRNA target viruses were extracted computationally by the following four steps: miRNA targets within each human-infecting virus were predicted by using TargetScan software (Step 1); candidates were filtered by using a hybridization free-energy threshold of −20 (kcal/mol) (Step 2); candidates with normalized miRNA targeting scores above a threshold of 80 (Step 3) and average O/E values above a threshold of 1.4 were extracted (Step 4). For details, refer to the text.

3. Results and discussion

3.1. Prediction of human-infecting viruses downregulated by human miRNAs

We predicted energetically favorable human miRNA target sites within the virus genomes included in the human-infecting group and invertebrate-infecting group, according to the calculated hybridization free energy of respective pairs of human miRNAs and their target sites by using TargetScan software [8]. TargetScan software was the first miRNA target-predicting software to be developed, and it has been used efficiently to predict miRNA target genes in mammals [8]. This software searches for miRNA target sites with complete hybridization in the miRNA seed region (2nd to 8th nucleotides in the 5′ region of the miRNA), and calculates the free energy for each of the predicted miRNA-target duplexes. To discard miRNA-target pairs that, because of weak hybridization, were unlikely to interact with each other, we adopted a free energy threshold of −20.0 kcal/mol. miRNA-target pairs that met this criterion were used for further target extraction.

If some human miRNAs do function specifically in defense against human-infecting viruses, human miRNAs will have more target sites within the genomic sequences of human-infecting viruses than in those of invertebrate-infecting viruses. To test this hypothesis, and to incorporate this characteristic into our miRNA target extraction, we calculated the normalized miRNA targeting scores and average O/E values for the viral sequence of each member of both the human-infecting group and the invertebrate-infecting group. The normalized miRNA targeting scores were calculated by dividing the total miRNA binding free energy by the length of the corresponding genomic sequence and then multiplying by -100 to eliminate the effect of sequence length, because longer genomic sequences have stochastically better chances of generating favorable miRNA binding. Thus, a high normalized miRNA targeting score indicated that many miRNAs potentially targeted multiple sites on the genomic sequence. The distribution of the normalized miRNA targeting scores of the human-infecting group ranged over markedly higher values did than that of the invertebrate-infecting group (Fig. 2A). A *t*-test comparing these two groups demonstrated a statistically significant difference between these two distributions ($P < 0.01$). This result suggests that human miRNAs are able to bind to more target sites on those viral genomes that infect humans than on those that infect invertebrate species. To select candidate human miRNA target viruses, we screened viral genomic sequences with normalized miRNA targeting scores higher than 80. This threshold was taken as the point of highest overall accuracy ($Q = 65.27$) in a comparison of the scores of the human-infecting group as a positive set with those of the invertebrate-infecting group as a negative set (Fig. 2B; for details, see Section 2.3).

In parallel, the average O/E values of normalized miRNA targeting scores were also calculated for viral genomes in both the human-infecting group and the invertebrate-infecting group to produce a reliability score evaluating whether the human-infecting viral genomes had significantly more target sites of actual human miRNA sequences than of randomized miRNA sequences. We generated 100 sets of shuffled miRNAs and used the same approaches to calculate the normalized miRNA targeting scores of viral genomes within each of the randomized trials, which were used for the calculation of average O/E value and 95% confidence intervals. An O/E value of over 1.0 indicate the significance of the scores originating from actual miRNAs against randomized controls. In our previous work, we proposed and used the same method to evaluate the prediction accuracy of miRNA target sites of *C. elegans* [14]. As was the case in our evaluation of normalized miRNA targeting scores, the distribution of average O/E values of the human-infecting group ranged over markedly higher values than did that of the invertebrate-infecting group, indicating that the miRNAs tended to target viruses that infect humans (Fig. 2A). The majority of members of the invertebrate-infecting group had average O/E closer to 1.0, suggesting that their result did not differ greatly when either real or randomized miRNA was used. In contrast, a large number of viruses in the human-infecting group had average O/E values around 1.5, showing that human miRNA sequences were significantly more likely to be associated with human-infecting viruses than were randomized miRNA sequences ($P < 0.01$). The threshold of the average O/E value for miRNA target virus detection was determined to be 1.4; this was the point at which the accuracy (Q) took the highest value of 71.18 when the average O/E values of the human-infecting group and the invertebrate-infecting group were compared (Fig. 2C). The distributions of the normalized miRNA targeting scores and average O/E values for the human-infecting group and invertebrate-infecting group also suggested differences in the human miRNA recognition of these 2 groups (Fig. 3A and Suppl. Fig. S1A). By

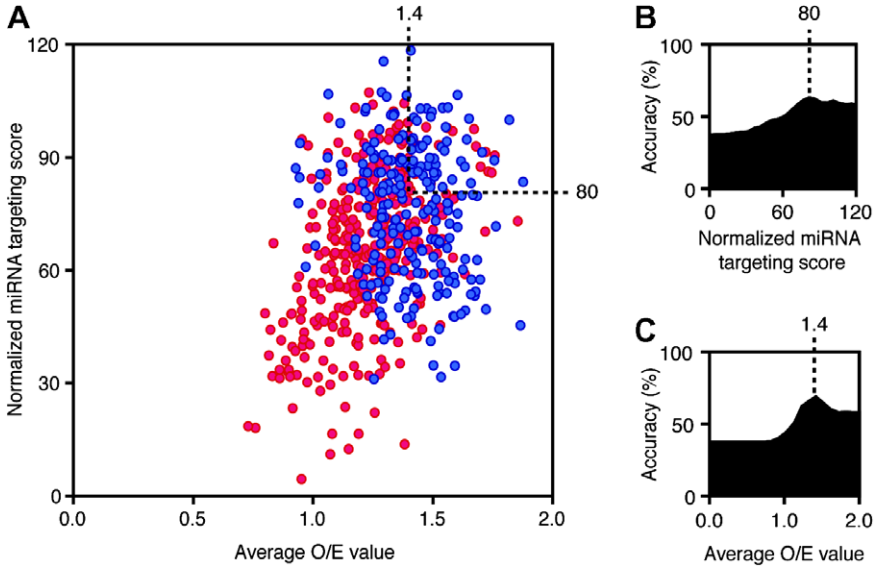

Fig. 2. Classification of human-infecting viruses and invertebrate-infecting viruses by using two scores. (A) Average O/E value (horizontal axis) and 'normalized miRNA targeting score' (vertical axis) of human-infecting viruses are plotted in blue, and those of invertebrate-infecting viruses are plotted in red. The average 95% confidence intervals for average O/E values, calculated from 100 randomized trials, were 0.023 for the human-infecting group (maximum 0.074 and minimum 0.006) and 0.025 for the invertebrate-infecting group (maximum 0.106 and minimum 0.005). Threshold values (1.4 for average O/E value and 80 for normalized miRNA targeting score) are indicated by dotted lines. Accuracies (Q) calculated for each normalized miRNA targeting score (B) and average O/E value (C) are shown. The threshold values are indicated by dotted lines.

using the normalized miRNA targeting scores and average O/E values, and their determined thresholds, from among those viruses infecting humans we extracted 62 as possible human miRNA targets (Table 1). Also, we listed the miRNAs that were predicted to target human-infecting viruses, along with their "tissue specific enrichment" (Heart, Connective tissue, Embryonic tissue and cell lines, Endocrine glands and tissues, Hematopoetic system, Kidney, Liver, Nervous system, Reproductive System, and Respiratory system) analyzed within the previous work by Landgraf et al. [24] (Suppl. Table 1). Although the tissue specificity was observed for some miRNAs, most of the virus targeting miRNAs express ubiquitously as in the case of majority of human encoded miRNAs [24]. This result suggests that miRNAs may defense against viruses by expressing in various tissues.

3.2. Significance of targeting of human-infecting viruses by human miRNAs rather than by miRNAs from other species

To determine the significance of the human miRNAs regulating human-infecting viruses, normalized miRNA targeting scores and average O/E values calculated by using human-infecting viruses were compared with those calculated by using invertebrate-infecting viruses (see Section 3.1). Moreover, we analyzed of O/E values by using miRNAs encoded in 4 species other than human, *M. musculus*, *D. melanogaster*, *C. elegans*, and *A. thaliana*. These results were compared with those obtained with human miRNAs to determine whether the distributions calculated by using human miRNAs were dependent on miRNAs encoded in humans. The distributions of the O/E values for the human-infecting group and invertebrate-infecting group calculated by using the miRNAs encoded in the five different species are shown in Fig. 3. The majority of human-infecting viruses gave O/E values higher than 1.0 when miRNAs from human or mouse were used for the analysis. In contrast, the O/E values clustered around 1.0 when miRNAs from *D. melanogaster*, *C. elegans*, and *A. thaliana* were used for the same analysis, suggesting that trials with real and randomized miRNAs were not markedly different. In other words, human or mouse miRNAs target human-infecting viruses significantly more frequently than randomized controls do. The direct comparison of O/E values within five different species and possible interaction between miRNAs and human-infecting viruses are summarized (Suppl. Fig. S2).

The close similarity in the distributions of O/E values calculated with human and mouse miRNAs may have been the result of a high level of conservation of miRNA sequences between these two species. Of the human miRNAs, 77.9% are of the same miRNA family [4] as those of *M. musculus*, whereas only 24.7% of human miRNAs share a family with those of *D. melanogaster*, 11.7% with those of *C. elegans*, and 1.7% with those of *A. thaliana* (data not shown). This trend of conservation among the majority of known miRNAs in humans and mice suggests that a high level of conservation may also stand true for the unknown miRNAs of these species. Another reason for the similar distributions of the O/E values calculated with human and mouse miRNAs may be the relatively close evolutionary distance between humans and mice, suggesting that their virus infection mechanisms are to some extent shared.

Similar trends were observed for both the normalized miRNA targeting scores and the O/E values calculated for viruses in the human-infecting group and invertebrate-infecting group using miRNAs of *M. musculus*, *D. melanogaster*, *C. elegans*, and *A. thaliana* (Fig. 3 and Suppl. Fig. S1). These results suggest that the predicted miRNA–virus pairings depend on both the human-infecting viruses and the human miRNA sequences, and they support the reliability of our miRNA target virus prediction (Fig. 3 and Suppl. Fig. S1).

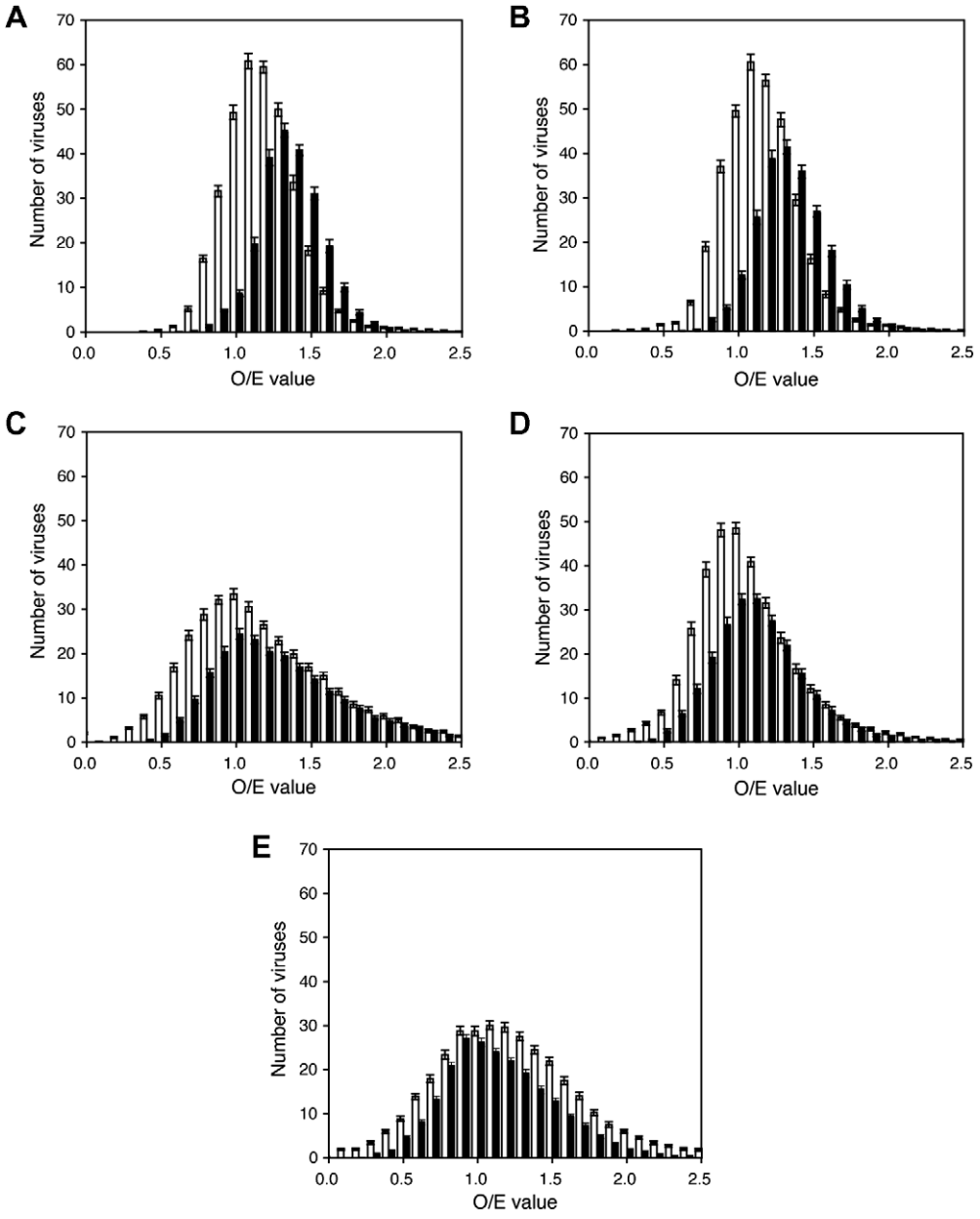

Fig. 3. O/E values of human-infecting viruses and invertebrate-infecting viruses, calculated by using miRNAs from five species. Average O/E values of human-infecting group and invertebrate-infecting group were calculated by using miRNAs encoded by five organisms: (A) *Homo sapiens*, (B) *Mus musculus*, (C) *Drosophila melanogaster*, (D) *Caenorhabditis elegans*, and (E) *Arabidopsis thaliana*. Solid bars: human-infecting group; open bars: invertebrate-infecting group. Average O/E values were calculated by taking the average value of each distribution calculated by using 100 different sets of shuffled miRNAs. Error bars indicate 95% confidence intervals.

3.3. Candidates for viruses regulated by human miRNAs

The list of extracted candidate viruses that may be regulated by human miRNAs includes 62 viral genomic sequences. The higher-scoring members targeted by various human encoded miRNAs were echovirus and papillomavirus in means of both normalized miRNA targeting score and O/E value. RNA virus families accounted for 66.7% (4 out of 6) of the predicted number of miRNA-targeted virus families, and all of the RNA virus members were ssRNA positive-strand viruses (Table 1). Moreover, ssRNA- positive virus genomes represented 56.5% (35 of 62) of the total predicted number of miRNA target virus genomes. In view of the fact that ssRNA-positive virus genomes accounted for 43.4% (99 out of 228) of the primary data set (i.e. before candidate extraction), this trend is unlikely to have been the result of a pre-existing trend in the initial data set. Therefore, ssRNA-positive viruses were extracted specifically by the miRNA target-prediction scheme. These results suggest that ssRNA-positive viruses are more strongly regulated by miRNAs than are other types of viruses. Extracted virus genomes other than ssRNA-positive ones were mainly double-stranded DNA (dsDNA) viruses (37.1%; 23 out of 62). Interestingly, all of the extracted dsDNA viruses were human papillomaviruses, indicating that these viruses may be exceptionally targeted by miRNAs among dsDNA viruses, although further analysis is necessary before we can confidently subscribe to this view.

Table 1
Human-infecting viruses that may be regulated by human miRNAs

Virus type	Virus family name	Virus name	Sequence length (bp)	Normalized miRNA targeting score	O/E value
(+) ssRNA virus	*Picornaviridae*	Human echovirus 11 isolate HUN-1108	7433	107.0	1.70
		Human echovirus 6	7417	103.4	1.63
		Human echovirus 15 strain CH 96-51	7437	101.1	1.56
		Human echovirus 31 strain Caldwell	7432	94.9	1.51
		Human echovirus 9 isolate DM	7453	93.1	1.43
		Human echovirus 27 strain Bacon	7412	92.9	1.45
		Human echovirus 16 strain Harrington	7437	91.2	1.41
		Human echovirus 3 strain Morrisey	7428	90.2	1.44
		Human echovirus 33 strain Toluca-3	7394	89.7	1.40
		Human echovirus 17 strain CHHE-29	7416	89.1	1.41
		Human echovirus 30	7440	88.8	1.41
		Human echovirus 30 strain Bastianni	7445	88.3	1.41
		Human coxsackievirus B3	7399	100.2	1.56
		Human coxsackievirus A16	7413	95.1	1.49
		Human coxsackievirus B5	7402	94.7	1.55
		Human coxsackievirus A24	7461	94.3	1.60
		Human coxsackievirus B6	7393	93.5	1.50
		Human coxsackievirus B2	7403	93.1	1.44
		Human coxsackievirus A18	7460	91.2	1.58
		Human coxsackievirus B4	7395	90.4	1.42
		Human coxsackievirus B4 strain E2	7397	89.4	1.43
		Human coxsackievirus A21 (strain Coe)	7401	83.1	1.45
		Human poliovirus 1	7441	95.5	1.59
		Human poliovirus 1 Mahoney	7440	92.2	1.56
		Human poliovirus 2	7440	88.0	1.42
		Human poliovirus 3	7432	85.1	1.40
		Enterovirus 95 strain CIV03-10361	7442	89.7	1.42
		Human enterovirus 90 isolate F950027	7438	86.3	1.47
		Human rhinovirus 14	7212	80.2	1.81
	Caliciviridae	Human calicivirus NLV/GII/Langen1061/2002/DE	7556	102.0	1.43
		Snow Mountain virus	7537	101.6	1.44
	Retroviridae	Simian-Human immunodeficiency virus	10006	83.5	1.59
	Astroviridae	Human astrovirus	6828	83.9	1.61
		Human astrovirus 4	6723	81.8	1.52
		Human astrovirus 1	6771	81.3	1.47
dsDNA virus	*Papillomaviridae*	Human papillomavirus type 2a	7860	106.6	1.65
		Human papillomavirus type 57b	7868	103.7	1.50
		Human papillomavirus type 29	7916	103.5	1.76
		Human papillomavirus type 106	8035	101.6	1.63
		Human papillomavirus type 27	7823	101.2	1.53
		Human papillomavirus type 27b	7831	99.5	1.51
		Human papillomavirus type 57	7861	98.5	1.43
		Human papillomavirus type 94	7881	97.8	1.62
		Human papillomavirus type 3	7820	92.5	1.61
		Human papillomavirus type 28	7959	92.2	1.55
		Human papillomavirus – cand87	7998	91.8	1.60
		Human papillomavirus type 61	7989	90.3	1.50
		Human papillomavirus – 81	8070	89.7	1.60
		Human papillomavirus type 10	7919	89.6	1.50
		Human papillomavirus – cand62	8092	88.7	1.48
		Human papillomavirus type 84	7948	88.5	1.47
		Human papillomavirus type 101	7259	88.0	1.69
		Human papillomavirus – cand86	7983	87.5	1.47
		Human papillomavirus type 55	7822	86.3	1.86
		Human papillomavirus type 102	8078	84.6	1.44
		Human papillomavirus – 83	8104	83.6	1.42
		Human papillomavirus type 8	7654	82.6	1.63
		Human papillomavirus type 40	7909	81.2	1.52
ssDNA virus	*Parvovirinae*	Human parvovirus B19	5594	86.7	1.65
		Human parvovirus 4	5268	85.9	1.52
		Human bocavirus WLL-1	5297	82.8	1.67
		Human bocavirus isolate st2	5299	82.5	1.67

Candidate human miRNA target viruses are listed along with their virus types, virus family names, lengths of genomic sequences, normalized miRNA targeting scores, and average O/E values. The normalized miRNA targeting score was calculated by dividing the total free energy by the length of the viral genomic sequence and multiplying by −100.

Different types of echoviruses were predicted as miRNA target viruses. Echoviruses are ssRNA positive-strand viruses with an RNA length of approximately 7.5 kb; they contain an RNA replicase that is responsible for the formation of

structural proteins and other proteins necessary for cellular replication [27]. This virus is one of the leading causes of acute febrile illness and aseptic meningitis, especially in young children [27]. Predicted miRNA target sites were distributed within both coding and non-coding regions, although the distribution of target sites was relatively intense within the regions encoding the capsid proteins 1B (VP2) and 1C (VP3) (data not shown). These proteins, together with VP1 and VP4, form a closed capsid enclosing the viral positive-strand RNA genome.

Although the majority of candidates were RNA viruses, three types of DNA viruses were also included on the list. Most of the candidate viruses from this group were human papillomaviruses (23 out of 27). These viruses have circular double-stranded DNA genomes close to 8 kb in size, and they are known to cause cervical cancer and epithelial tumors [28]. The miRNA target sites on these viruses were distributed equally within both coding and non-coding regions (data not shown). Human papillomaviruses have persistent life strategies, i.e. they have biological characteristics that give them persistence in individual hosts and genetic stability [29]. The predicted targeting miRNAs for the viruses in each family are listed along with the number of target sites in Supplementary Table 1. Two known miRNA target viruses, HCV [19] and PFV-1 [20], were not extracted as candidate miRNA target viruses in our search. This was because their genomic sequences were not included within our primary dataset. However, since both HCV and PFV-1 are ssRNA positive-strand viruses, the fact that the majority of our predicted human miRNA target viruses were ssRNA positive-strand viruses supports the likelihood that these viruses will be classed as predicted targets of miRNAs when their sequences become available.

In conclusion, our analyses suggest that human miRNAs regulate various human-infecting viruses via binding and down regulation of their target genes as a form of antiviral defense, as shown by the stronger miRNA binding and higher significance against randomized controls and non-human encoded miRNA controls. Therefore, we showed that there was specificity of binding between human-infecting viruses and human-encoded miRNAs. We extracted 62 viral genomic sequences from 6 different virus families as candidate miRNA targets. The extracted candidates were mainly RNA viruses, suggesting that RNA viruses are more likely to be targets of human miRNAs any other types of viruses. From these results, we believe that human miRNAs function as antiviral defense mechanisms against a wide range of human-infecting viruses. Moreover, these findings would provide the experimentalists with the basis of experimental validation.

Acknowledgements: We thank Kosuke Fujishima and Hikaru Taniguchi (Keio University) for advices on carrying out the computational analysis, and Dr. Satoshi Koike (Tokyo Metropolitan Institute for Neuroscience) for useful comments on the virus tropism. We are also grateful to Dr. Rintaro Saito, Dr. Yuko Osada (Keio University) and Dr. Koji Numata (RIKEN BioResource Center) for their critical discussions throughout this study. This research was supported in part by a Grant-in-Aid from the 21st Century Center of Excellence (COE) Program, entitled "Understanding and Control of Life's Function via Systems Biology (Keio University)"; the Computer Simulation Project, Ministry of Education, Culture, Sport, Science and Technology, Japan; the Japan Society for the Promotion of Science (JSPS); the research funds from the Yamagata government and Tsuruoka City, Japan; and Keio University.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2007.08.049.

References

[1] Ambros, V. (2001) microRNAs: tiny regulators with great potential. Cell 107, 823–826.
[2] Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.
[3] Griffiths-Jones, S. (2004) The microRNA Registry. Nucleic Acids Res. 32, D109–D111.
[4] Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R. and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33, D121–D124.
[5] Rajewsky, N. (2006) microRNA target predictions in animals. Nat Genet. 38 (Suppl.), S8–S13.
[6] Bentwich, I. (2005) Prediction and validation of microRNAs and their targets. FEBS Lett. 579, 5904–5910.
[7] Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.
[8] Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. Cell 115, 787–798.
[9] Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) MicroRNA targets in Drosophila. Genome Biol. 5, R1.
[10] John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. PLoS Biol. 2, e363.
[11] Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10, 1507–1517.
[12] Krek, A. et al. (2005) Combinatorial microRNA target predictions. Nat Genet. 37, 495–500.
[13] Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets. Genes Dev. 18, 1165–1178.
[14] Watanabe, Y., Yachie, N., Numata, K., Saito, R., Kanai, A. and Tomita, M. (2006) Computational analysis of microRNA targets in *Caenorhabditis elegans*. Gene 365, 2–10.
[15] Lim, L.P. et al. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769–773.
[16] Pfeffer, S. et al. (2004) Identification of virus-encoded microRNAs. Science 304, 734–736.
[17] Pfeffer, S. et al. (2005) Identification of microRNAs of the herpesvirus family. Nat Methods 2, 269–276.
[18] Bennasser, Y. and Jeang, K.T. (2006) HIV-1 Tat interaction with Dicer: requirement for RNA. Retrovirology 3, 95.
[19] Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M. and Sarnow, P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science 309, 1577–1581.
[20] Lecellier, C.H., Dunoyer, P., Arar, K., Lehmann-Che, J., Eyquem, S., Himber, C., Saib, A. and Voinnet, O. (2005) A cellular microRNA mediates antiviral defense in human cells. Science 308, 557–560.
[21] Hariharan, M., Scaria, V., Pillai, B. and Brahmachari, S.K. (2005) Targets for human encoded microRNAs in HIV genes. Biochem. Biophys. Res. Commun. 337, 1214–1218.
[22] Triboulet, R. et al. (2007) Suppression of microRNA-silencing pathway by HIV-1 during virus replication. Science 315, 1579–1582.
[23] Hsu, P.W., Lin, L.Z., Hsu, S.D., Hsu, J.B. and Huang, H.D. (2007) ViTa: prediction of host microRNAs targets on viruses. Nucleic Acids Res. 35, D381–D385.
[24] Landgraf, P. et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401–1414.
[25] Kulikova, T. et al. (2007) EMBL nucleotide sequence database in 2006. Nucleic Acids Res. 35, D16–D20.

[26] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A. and Nielsen, H. (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16, 412–424.
[27] Modlin, J.F. (1986) Perinatal echovirus infection: insights from a literature review of 61 cases of serious infection and 16 outbreaks in nurseries. Rev. Infect. Dis. 8, 918–926.
[28] de Villiers, E.M., Fauquet, C., Broker, T.R., Bernard, H.U. and zur Hausen, H. (2004) Classification of papillomaviruses. Virology 324, 17–27.
[29] Villarreal, L.P. (2005) Viruses and the Evolution of Life, American Society for Microbiology Press.