Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 365 (2006) 2–10
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Computational analysis of microRNA targets in *Caenorhabditis elegans*

Yuka Watanabe [a,b], Nozomu Yachie [a,c], Koji Numata [a,c], Rintaro Saito [a], Akio Kanai [a,*], Masaru Tomita [a,b]

[a] *Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, 997-0017, Japan*
[b] *Department of Environmental Information, Keio University, Fujisawa, 252-8520, Japan*
[c] *Bioinformatics Program, Graduate School of Media and Governance, Keio University, Fujisawa, 252-8520, Japan*

Received 28 March 2005; received in revised form 4 August 2005; accepted 27 September 2005
Available online 13 December 2005

Abstract

MicroRNAs (miRNAs) are endogenous ~22-nucleotide (nt) non-coding RNAs that post-transcriptionally regulate the expression of target genes via hybridization to target mRNA. Using known pairs of miRNA and target mRNA in *Caenorhabditis elegans*, we first performed computational analysis for specific hybridization patterns between these two RNAs. We counted the numbers of perfectly complementary dinucleotide sequences and calculated the free energy within complementary base pairs of each dinucleotide, observed by sliding a 2-nt window along all nucleotides of the miRNA-mRNA duplex. We confirmed not only strong base pairing within the 5′ region of miRNAs (nts 1–8) in *C. elegans*, but also the required mismatch within the central region (nt 9 or nt 10), and we found weak binding within the 3′ region (nts 13–14). We also predicted 687 possible miRNA target transcripts, many of which are thought to be involved in *C. elegans* development, by combining the abovementioned hybridization tendency with the following analyses: (1) prediction of the miRNA-mRNA duplex with free-energy minimization; (2) identification of the complementary pattern within the miRNA–mRNA duplex; (3) conservation of target sites between *C. elegans* and *C. briggsae*, a related soil nematode; and (4) extraction of mRNA candidates with multiple target sites. Rigorous tests using shuffled miRNA controls supported these predictions. Our results suggest that miRNAs recognize their target mRNAs by their hybridization pattern and that many target mRNAs may be regulated through a combination of several specific miRNA target sites in *C. elegans*.

Keywords: Bioinformatics; Nematode; *C. briggsae*; MicroRNA targets; MiRNA–mRNA duplex

1. Introduction

Noncoding RNA (ncRNA) genes produce various kinds of functional RNA molecules. The short sequences known as microRNAs (miRNAs), a family of ncRNAs, imperfectly hybridize to target mRNAs and regulate their messages (Bartel, 2004; He and Hannon, 2004). Although the first miRNA was identified not long ago, experimental and computational approaches have now found more than 1000 different miRNAs. Many endogenously encoded miRNAs have been cloned from a number of species (Lau et al., 2001; Lee and Ambros, 2001; Reinhart et al., 2002; Ambros et al., 2003; Lim et al., 2003), and a combination of computational approaches has yielded even more (Ambros et al., 2003; Lai et al., 2003; Lim et al., 2003). MiRNA sequences are archived and accessible at the Rfam miRNA registry website (Griffiths-Jones, 2004). The database contains a total of 1420 miRNAs from 13 species, including *Homo sapiens*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and even the Epstein-Barr virus. However, experimental studies have touched on only a handful of functions (Lee et al., 1993; Wightman et al., 1993; Moss et al., 1997; Reinhart et al., 2000; Brennecke et al., 2003; Johnston and Hobert, 2003) of miRNAs such as *lin-4* (Lee et al., 1993) and *let-7* (Reinhart et al., 2000), well known as the founding members of the miRNA family and having central roles as key regulators of developmental timing through cell fate decisions. It was found recently that some miRNAs regulate a large number of messages (Lim et al., 2005), supporting the proposition that miRNAs work as key participants in gene regulatory networks (Ke et al., 2003; John et al., 2004).

Abbreviations: miRNA, microRNA; *C. elegans*, *Caenorhabditis elegans*; *C. briggsae*, *Caenorhabditis briggsae*; UTR, untranslated region(s).

* Corresponding author. Tel.: +81 235 29 0524; fax: +81 235 29 0525.
E-mail address: akio@sfc.keio.ac.jp (A. Kanai).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.09.035

The functional characterization of miRNAs relies heavily on the identification of their target mRNAs. To reveal the function of miRNAs in a relatively straightforward and powerful way, a number of bioinformatic or combined bioinformatic and experimental approaches to the prediction of target mRNAs have been performed. The binding pattern of miRNA–mRNA duplexes has been studied computationally (Enright et al., 2003; Lewis et al., 2003; John et al., 2004) and experimentally (Stark et al., 2003; Kiriakidou et al., 2004). Because miRNAs are very short and miRNA–mRNA duplexes are not entirely complementary, nucleotides within the 5′ region of miRNA that participate in perfect or nearly perfect matches have been the targets for identifying potential miRNA–mRNA duplexes computationally. At the same time, some confirmed miRNA–mRNA duplexes are known to have mismatches in the 5′ region. Several computer programs have been developed for genome-wide searches of miRNA targets. Target Scan searches for target mRNAs with complementary sites within the 5′ region of miRNA and predicts potential binding sites (Lewis et al., 2003), and miRNA predicts miRNA hybridization sites with a position-weighted local alignment algorithm (Enright et al., 2003). These approaches have successfully predicted a number of miRNA–mRNA duplexes for *D. melanogaster*, *H. sapiens*, *Mus musculus*, and *A. thaliana*. Although *C. elegans* is one of the organisms for which the functions of a number of miRNAs are known, no genome-wide prediction of miRNA targets in *C. elegans* has been reported.

We sought to elucidate the regulatory roles of miRNA by applying a bioinformatic approach to the *C. elegans* genome. We computationally analyzed the base-pairing pattern of known *C. elegans* miRNA–mRNA duplexes, in which strong base pairing occurred at positions 1 through 8, and confirmed the fact that the 5′ region of miRNA plays an important role within miRNA target recognition. We also confirmed weak binding at the center of the miRNA–mRNA duplex, and observed that the required mismatch occurs especially at position 9 or 10. We found a weak binding region at positions 13 and 14. By analyzing the strength of each binding position according to its free energy instead of a perfect match pattern, we could evaluate strong binding within the 5′ regions of miRNAs without missing many known miRNAs that show less-than-perfect binding patterns in this region. By applying the analyzed hybridization tendency to *C. elegans* mRNAs, we identified 687 target mRNA candidates and estimated the number of possible miRNA-binding sites in their 3′ UTRs. Our results suggest that many mRNAs may be regulated by formation of specific binding patterns with miRNAs and a combination of multiple target sites.

2. Materials and methods

2.1. Datasets

Mature *C. elegans* and *C. briggsae* miRNA sequences were obtained from the Rfam miRNA database (Griffiths-Jones, 2004) in a FASTA-formatted sequence file. The file contained a total of 117 *C. elegans* miRNA sequences and 78 *C. briggsae* miRNA sequences. For *C. briggsae* miRNA datasets, we used an additional 39 *C. elegans* miRNAs to cover the sequences in *C. briggsae* where data were incomplete.

The Ensembl database (Birney et al., 2004) served as the source of genomic data. EnsMart version 26 (Kasprzyk et al., 2004), an application user interface available at Ensembl Archives, was used to generate the 3′ UTR sequence file. In total, 9779 transcript sequences (including alternatively spliced isoforms) from 8894 genomic loci were available for the *C. elegans* 3′ UTR dataset. For *C. briggsae*, 698 transcripts were retrieved from the database. To compensate for the lack of *C. briggsae* 3′ UTR sequence data, predictions for the missing 3′ UTR sequences were generated by taking 1000 bp of genomic sequence downstream of the last exon conserved in *C. elegans*. If the predicted sequences overlapped with the next coding sequences on either strand, the 3′ UTR extension was halted at that point. In total, 8436 transcripts (including alternatively spliced isoforms) sequenced from 7122 genomic loci were retrieved as the *C. briggsae* 3′ UTR dataset.

2.2. Analysis of known miRNA–mRNA duplexes

We selected *let-7* and *lin-4* targeting the mRNAs *lin-14* (GenBank/EMBL/DDBJ accession no. NM_077516), *lin-28* (NM_059880), *lin-41* (NM_060087), and *hlb-1* (NM_076575), which yielded a total of 25 miRNA–mRNA duplexes that have been experimentally defined (Banerjee and Slack, 2002; Abrahante et al., 2003; Lin et al., 2003; Stark et al., 2003; He and Hannon, 2004). We then counted the number of dinucleotides showing perfect matches in the known miRNA–mRNA duplexes. Perfect matches were counted and compared in two different ways: with and without considering GU wobble as mismatch. The free energy within each dinucleotide was calculated by using the thermodynamic parameters described by Freier et al. (1986). For dinucleotides causing a bulge or internal loop, +0.0 (kcal/mol) was assigned as the free-energy score.

The conservation of 8-nt sequences within *C. elegans* miRNAs was calculated by sliding an 8-nt window along each miRNA nucleotide and counting the number of miRNAs with exactly the same 8-nt sequence at the same position among the 117 known *C. elegans* miRNAs (see Section 2.1). The number of conserved octamers was taken as one if no matching octamer sequence occurred within the other miRNAs. Also, the number of gaps was calculated within every 8-nt window of the miRNA–mRNA duplexes.

2.3. MiRNA target prediction

Complementary miRNA and mRNA target sites were predicted by using RNAhybrid software (Rehmsmeier et al., 2004) and a free-energy threshold of less than −20.0 kcal/mol. The RNAhybrid software predicts potential hybridization with short and long RNAs by using minimum free-energy calculation.

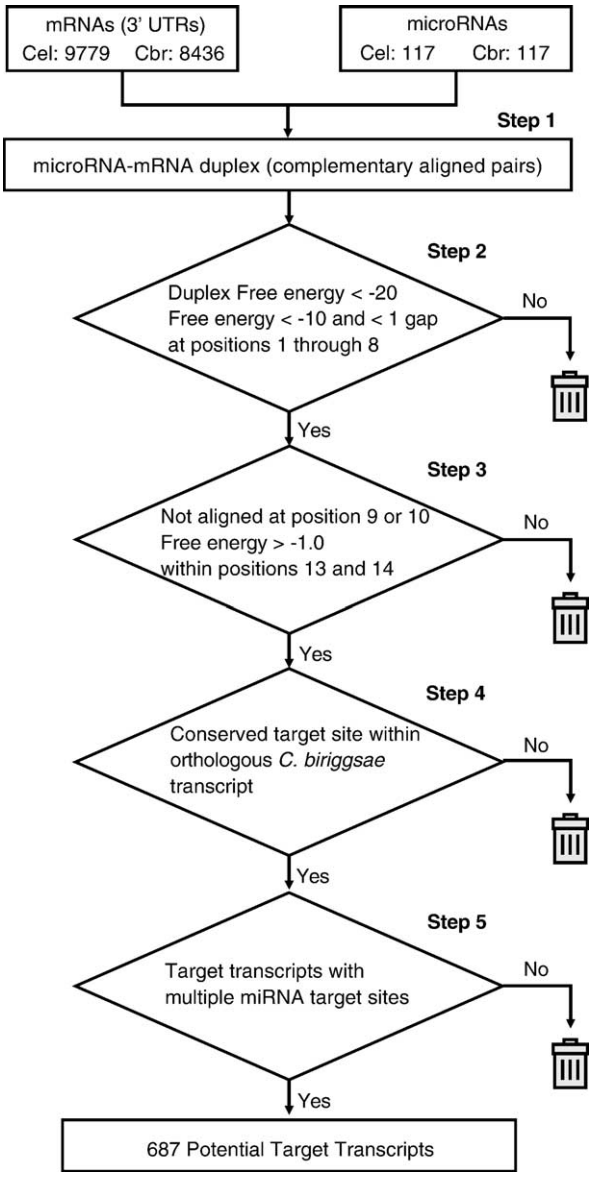

Fig. 1. Computational extraction of mRNA candidates containing miRNA target sites. Candidate mRNAs containing miRNA target sites are extracted through five strategies computationally: hybridization selection (step 1), hybridization pattern selection with perfect match and free-energy thresholds (steps 2 and 3), conservation analysis between the two related nematodes *Caenorhabditis elegans* (Cel) and *Caenorhabditis briggsae* (Cbr) (Step 4), and selection of candidates with multiple target sites (step 5).

Of those miRNAs and mRNAs that formed duplexes with free-energy values less than −20.0 kcal/mol, we then determined which met the described thresholds for specific regions of the duplexes. The requirements were as follows: the sum of the free energy calculated for positions 1 through 8 must be lower than −10.0 (kcal/mol); there must be one or no gap at positions 1 through 8; mismatch must occur at position 9 or 10; and the free energy of positions 13–14 must be higher than −1.0 kcal/mol. These thresholds were determined in light of the characteristics of known miRNA–mRNA duplexes (see Section 3.1).

The 3′ UTR sequences of *C. elegans* and *C. briggsae* were used to identify potential miRNA targets. For this purpose, we asked whether predicted target sites for the same miRNA were found in the conserved target mRNAs of both of these related species. Conserved target mRNAs were required to be orthologous on the basis of the Ensembl classification. Finally, we counted the number of miRNA binding sites for each target transcript, to check whether there was more than one (Fig. 1).

We listed the candidates that had the Gene Ontology (Ashburner et al., 2000) classification of "development (GO:0007275)" (Table 1) and had three or more target sites in their 3′ UTRs as examples of target candidate transcripts that might be involved in development. To observe genes lower in the Gene Ontology hierarchy, we extracted candidates according to the slimed Gene Ontology hierarchy (Martin et al., 2004). WormBase (Harris et al., 2004) was used for identification of these genes.

Table 1
Target mRNAs that may be involved in development in *Caenorhabditis elegans*

Gene ID	Gene name	Target sites	miRNA
C18A3.5f	*C18A3.5*	3	*miR-256 miR-271*
C34E7.4	*C34E7.4*	3	*miR-38 miR-42 miR-56*
C34F11.3b	*C34F11.3*	6	*miR-35 miR-36 miR-37 miR-39 miR-40 miR-271*
C34F11.9a	*dsh-1*	10	*miR-35 miR-36 miR-38 miR-39 miR-78 miR-249 miR-271*
D2021.1	*D2021.1*	3	*miR-34 miR-265*
E01A2.6	*E01A2.6*	3	*miR-2 miR-264*
F11C7.5	*F11C7.5*	5	*miR-56* miR-81 miR-250 miR-268*
F23H12.5	*F23H12.5*	3	*miR-55 miR-56 miR-56**
F32A7.5a	*F32A7.5*	3	*miR-38 miR-264 miR-271*
F32A7.5b	*F32A7.5*	3	*miR-38 miR-264 miR-271*
F32A7.5c	*F32A7.5*	3	*miR-38 miR-264 miR-271*
F32B5.6c	*F32B5.6*	3	*miR-49 miR-56**
F33A8.5	*F33A8.5*	4	*miR-56* miR-78 miR-264*
K08D12.3b	*K08D12.3*	4	*miR-56* miR-78 miR-264*
K08F8.1b	*K08F8.1*	3	*miR-39 miR-55 miR-273*
M01F1.3	*M01F1.3*	6	*miR-34 miR-55 miR-56* miR-264 miR-273*
R07E5.12	*R07E5.12*	4	*miR-42 miR-264*
T01B11.2b	*T01B11.2*	3	*miR-55 miR-271*
T05C12.10	*qua-1*	3	*miR-35 miR-56* miR-271*
T08B2.7c	*T08B2.7*	3	*miR-39 miR-40 miR-124*
T14D7.2	*T14D7.2*	6	*miR-35 miR-38 miR-39 miR-40 miR-41 miR-271*
W09B6.1a	*W09B6.1*	4	*miR-34 miR-35 miR-249 miR-264*
Y105E8B.2b	*Y105E8B.2*	3	*miR-233 miR-264*
Y113G7B.17	*Y113G7B.17*	11	*miR-34 miR-42 miR-271*
Y39G10AR.12a	*Y39G10AR.12*	4	*miR-34 miR-56* miR-264*
Y47G6A.18	*Y47G6A.18*	3	*miR-56 miR-235 miR-273*
Y54E5B.1a	*smp-1*	3	*miR-124 miR-246 miR-268*
Y57E12AL.1a	*Y57E12AL.1*	3	*miR-35 miR-268 miR-271*
Y57E12AL.1b	*Y57E12AL.1*	3	*miR-35 miR-268 miR-271*
Y64G10A.6	*Y64G10A.6*	5	*miR-35 miR-36 miR-39 miR-41 miR-271*
Y66H1B.2	*Y66H1B.2*	3	*miR-34 miR-78*

Listed are miRNA target transcripts with a Gene Ontology (Ashburner et al., 2000) classification of "development (GO:0007275)" and with three or more target sites.

2.4. Characterization of predicted target sites

First, we counted the number of potential target transcripts with one (counted before excluding candidates with single target sites) or more miRNA binding sites within their 3′ UTRs. We also counted the number of target site(s) in each target transcript that originated from the same miRNA. We calculated the observed/expected (*O*/*E*) value in the following way. The count derived from actual miRNA sequences was defined as the observed value, whereas that originating from randomized trials was defined as the expected value. The observed value (*O*) was divided by the expected value (*E*) to indicate the significance of the characteristics within an actual miRNA. Each randomized trial was achieved by shuffling all miRNAs while preserving their dinucleotide compositions, since the free energy of the RNA structure is known to be strongly influenced by their dinucleotide composition (Freier et al., 1986). The targets of these shuffled miRNAs were predicted in the same way as described for miRNAs, in order to generate randomized controls. A total of 100 randomized experiments were performed and averaged across all experiments. To calculate *O*/*E* values in cases where expected value (*E*) happened to be 0, we added 0.5 to all of the observed (*O*) and expected (*E*) values as background (Fleiss, 1979).

3. Results

3.1. Binding patterns of known miRNA–mRNA duplexes

To retrieve information on the rules for miRNA–mRNA recognition in *C. elegans*, we initially analyzed the pattern of hybridization by using reported pairs of miRNAs and mRNAs. We counted the numbers of perfect matches among units of two nucleotides (Fig. 2a,b), which is the minimum length for counting continuous sequences. We gained insight into the significance of perfect matching by determining the percentage of perfect match within known miRNA–mRNA duplexes. Perfect match binding occurred frequently, especially for the octamer of nts 1 through 8 in the 5′ region of miRNAs. In comparison, the ratio of perfect matching for positions 9 and 10 was especially low, at only 4% (1 in 25). In other words, most of the miRNA–mRNA duplexes had mismatches at position 9 or 10.

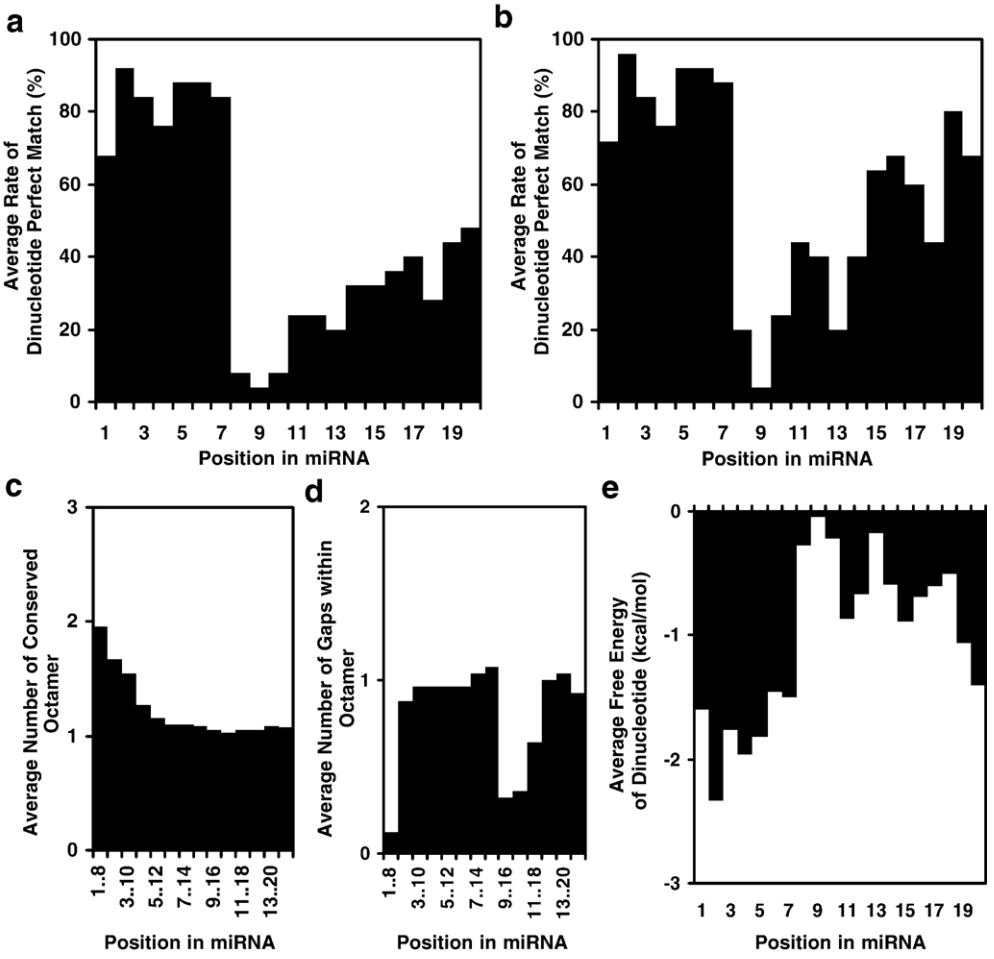

Fig. 2. Analysis of nucleotide recognitions in known miRNA–mRNA hybrids. The average rate of dinucleotide perfect match was calculated for each nucleotide of the miRNA–mRNA hybrids considering GU wobble as either (a) a mismatch or (b) a match. (c) 5 regions of miRNAs were relatively well conserved compared to the other regions in *Caenorhabditis elegans.* Numbers of the same octamer sequences in the same position on each miRNA were counted and the averages are shown. (d) The 5 regions of miRNAs had fewer gaps compared to the other regions. Average numbers of gaps within each octamer are shown. (e) Free-energy analysis of each dinucleotide in miRNA–mRNA hybrids. Horizontal axis indicates the nucleotide position on the miRNA.

Another low perfect match binding pattern occurred at positions 13 and 14 and was especially low when GU wobble was considered a match. Compared with the relatively high rate of perfect match binding within the 3′ region of miRNA–mRNA duplexes (especially when GU wobble was considered a match), the percentage of perfect match binding was very low at positions 13 and 14. These results indicate a relatively low usage of GU wobble within the regions comprising nts 1 through 8, 9 and 10, and 13 and 14, although GU wobbles tended to be common within the 3′ region of miRNAs. We then counted the number of conserved octamers among the 117 *C. elegans* miRNAs (Fig. 2c). Our analysis confirmed that the 5′ regions of miRNAs were well conserved among *C. elegans* miRNAs. In comparison, hardly any miRNAs had conserved octamers within the 3′ region. We also counted the number of gaps within 8-nt nucleotides of miRNA sequences (Fig. 2d), and observed that there were very few gaps within positions 1 through 8.

Calculation of free energy yielded similar end results but more detailed information on the binding strengths within miRNA–mRNA duplexes (Fig. 2e). Using Freier's experimentally determined parameters for free energy (Freier et al., 1986), we found that the free energy value was fairly low for nts 1 through 8, rose at positions 8 and 9, and nearly reached +0.0

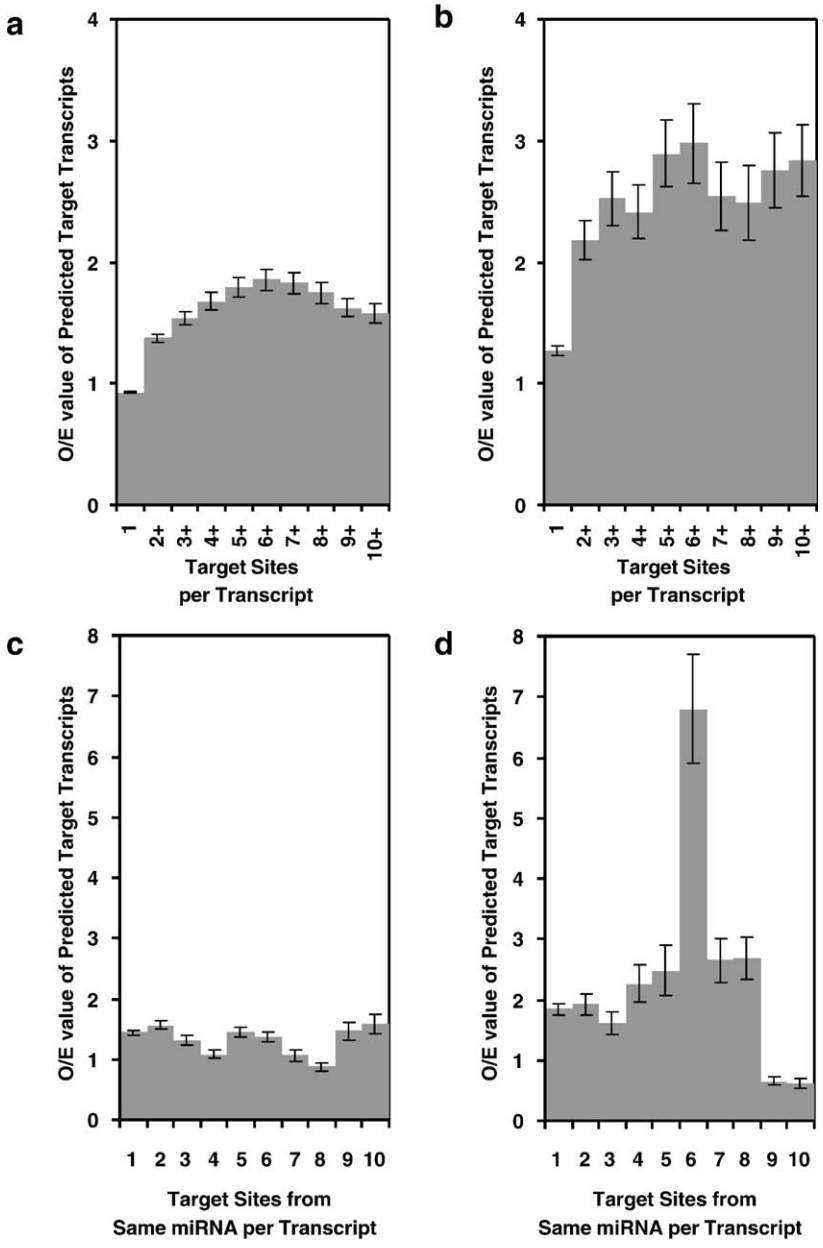

Fig. 3. Multiple target sites within target mRNAs. Each bar reflects the number of predicted transcripts with a given number of target sites. The analysis was performed by observing the number of target sites for either all miRNAs (a: *Caenorhabditis elegans*; b: conserved in *C. elegans* and *Caenorhabditis briggsae*) or each miRNA species (c: *C. elegans* and d: conserved within *C. elegans* and *C. briggsae*). The observed value (*O*) was the count derived from actual miRNA sequences and the expected value (*E*) was the average value of 100 randomized trials.

kcal/mol at positions 9 and 10. The free energy at positions 13 and 14 was also high compared with that of neighboring dinucleotides. These results indicate that the dinucleotides at positions 9 and 10 and those at positions 13 and 14 not only had a very low number of perfect matches but also had sequences that bound weakly to their target mRNAs.

3.2. Prediction and characterization of candidate miRNA targets

To identify miRNA–target mRNA pairs in *C. elegans*, we computationally screened 9779 3′ UTR regions in mRNAs. The scheme is summarized in Fig. 1. Briefly, we used RNAhybrid software to predict *C. elegans* miRNA–mRNA duplexes with free energy less than −20.0 (kcal/mol). In addition, predicted miRNA–mRNA duplexes had to have a total free energy value less than −10.0 kcal/mol and only one or no gap for nts 1 through 8, mismatch (not including GU wobble) at position 9 or 10, and free energy exceeding −1.0 kcal/mol at positions 13 and 14. Thresholds of miRNA–mRNA hybridization pattern were determined from analyses of known miRNA–mRNA duplexes and covered 84% of them. MiRNA–mRNA duplexes that appeared in both *C. elegans* and *C. briggsae* were extracted as conserved candidates. After comparing the *O*/*E* values of miRNA target transcripts with single potential target sites with those of transcripts having multiple target sites (Fig. 3a,b), we suggest that many miRNAs require multiple target sites to regulate a message. Therefore, we extracted mRNAs with multiple target sites as candidates. By applying the described miRNA–mRNA duplex criteria to the *C. elegans* genome, we extracted 687 transcripts (including alternatively spliced isoforms) from 626 different genes (detailed information about the target mRNAs is provided in the supplementary table). The use of evolutionally conserved candidates led us to infer the significance of conserved target sites because of their increased *O*/*E* values (Fig. 3). Many of the known *C. elegans* miRNAs that have been characterized so far seem to regulate various aspects of development (He and Hannon, 2004).

Therefore, we listed candidates with the Gene Ontology (Ashburner et al., 2000) classification of "development (GO:0007275)" (Table 1) to give some examples of target candidate transcripts that might be involved in development.

Next, we analyzed the number of target site(s) in each target transcript that originated from the same miRNA, in order to observe the trends in *O*/*E* values of mRNAs with more than two target sites. For this purpose, we calculated the *O*/*E* value for each number of the target sites originating from the same miRNA within every predicted target mRNA (Fig. 3c) and every predicted target mRNA that was conserved between the two related nematodes *C. elegans* and *C. briggsae* (Fig. 3d). Among the conserved candidates, the *O*/*E* value was high; it was over 1.5 for duplexes with one through eight target sites and especially high when the number of target sites rose to six. Therefore, we extracted candidates with one through eight target

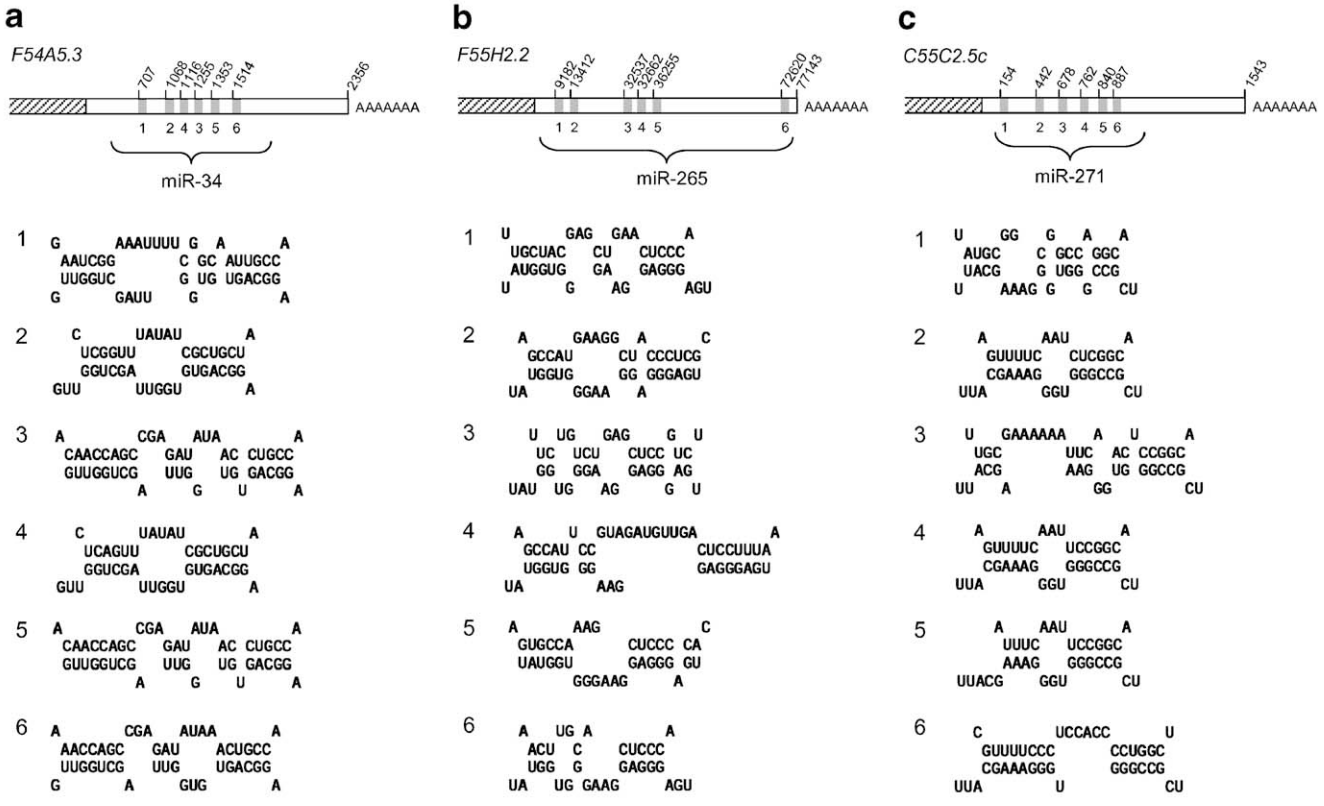

Fig. 4. Three examples of possible miRNA–mRNA duplexes with multiple target sites. Each miRNA (miR-34 [a], miR-265 [b], and miR-271 [c]) potentially hybridizes at multiple locations in a specific target mRNA. The positions of miRNA target sites in the 3 UTR of each target mRNA are shown (gray bars). Poly(A) sequence is shown in the 3 terminal end. Sequences and possible binding patterns of the miRNA(bottom)–mRNA(top) hybrids are shown.

sites for a single miRNA without lowering the significance within each set of target sites, and checked for the possibility that miRNAs might most commonly regulate their targets through six sites. Contrary to the *O*/*E* values for the conserved candidates (Fig. 3b,d), the *O*/*E* value was relatively low within each number of target sites for the *C. elegans* candidates (Fig. 3a,c). Three hypothetical examples of conserved miRNA–mRNA duplexes with six target sites per transcript are shown in Fig. 4. These mRNAs have relatively long 3′ UTRs, which may increase the likelihood of there being a large number of target sites. Meanwhile, the specificity of miRNA targeting is shown by the fact that one miRNA can have as many as six target sites in a single 3′ UTR. Because real miRNA target mRNAs also have relatively long 3′ UTRs, we hypothesize that the 3′ UTRs need to be long to accommodate a large number of target sites. Further analysis is necessary before we can substantiate this view.

4. Discussion

4.1. MiRNA target recognition in C. elegans

We analyzed miRNA–target mRNA recognition in *C. elegans* and observed that some patterns of hybridization between miRNA and mRNA may be crucial for regulation of the target mRNA. The ratio of perfect match binding between the target mRNA and nts 1 through 8 of the partner miRNA was high and the calculated free energy of this region was low, indicating strong binding between miRNAs and their target mRNAs. In mammals, perfect match binding of nts 2 through 8 has been identified computationally by comparing the number of predicted target sites within actual and shuffled miRNAs (Lewis et al., 2003). These investigators searched for targets at the genome level by locating perfect-match complementary sites for nts 2 to 8 of miRNAs and calculating the free energy within predicted miRNA–mRNA duplexes. Our result suggesting the importance of nts 1 through 8 of miRNAs is very similar and confirms their results. A similar requirement for strong binding at nts 1 to 8 is observed experimentally within *D. melanogaster* (Brennecke et al., 2005). Brennecke et al. suggested that 8-mer binding at nts 1 through 8 confers stronger regulation than does 7-mer binding at nts 2 through 8. By comparing 117 miRNAs of *C. elegans*, we confirmed that the sequence at nts 1 to 8 is very well conserved. For example, the 8-nt sequence UCACCGGG occurs within 7 different kinds of miRNAs. Nucleotides 1 through 8 of *C. elegans* miR-2 and miR-43 are perfectly complementary to the motif called the K box and are conserved among other species, including *H. sapiens* and *D. melanogaster* (Lai, 2002). Another example is the prediction of human miRNA target sites using strong binding at the 5′ region (John et al., 2004). Within the report of John et al., GU wobbles were treated as perfect matches, just like AU and GC binding. According to the results of experimental analysis (Vella et al., 2004), GU wobble in the 5′ region of the miRNA–mRNA duplex tends to weaken regulation by miRNA. In addition, the results of our analyses indicate that the 5′ region of miRNA tended not to permit GU wobble in the duplex with the target mRNA.

Therefore, during our prediction of miRNA–mRNA duplexes, we assigned an increased free-energy score to GU wobble, differentiating it from AU and GC matches. Our analysis of *C. elegans* miRNAs confirmed these nucleotide tendencies and supported the view that 8-mer binding at nts 1 through 8 at 5′ region plays an important role in miRNA targeting. In contrast, the binding is weaker and there is hardly any conservation among the 3′ regions of miRNAs. Therefore, as has been said, the 3′ region of miRNAs is likely to be required for the selectivity of each target site (Lim et al., 2003).

We also confirmed the fact that most of the known miRNA–mRNA duplexes bind weakly at their centers, and from our analysis of the perfect-match sequences, these mismatches tend to occur at position 9 or 10. Only one known miRNA–mRNA duplex that lacked mismatch at position 9 or 10 was identified– for the gene *hbl-1*, which is not conserved within *C. briggsae*– suggesting that the target site may not be very important for regulation of the message. The 2-nt segment containing mismatch within the central portion of miRNAs (nts 9 and 10) seems to be a requirement in miRNA–mRNA binding of *C. elegans*. Other experimental results point to the importance of a 2- to 5-nt bulge within the target mRNA sequence, a 6-to 9-nt bulge within the miRNA sequence, and a 2- to 3-nt internal loop in the central region of the duplex (Kiriakidou et al., 2004). We suggest that some mismatch in duplexes, especially at nt 9 or nt 10, is important in miRNA–mRNA target recognition. For small interfering RNAs (siRNAs), experimental results suggest that perfect-match binding at nt 9 is essential for target cleavage. Moreover, we used this structural constraint in the prediction of miRNA targets, resulting in enhancement of the *O*/*E* value (data not shown). In addition, mutations at nts 12 through 15 are said to prevent cleavage of the target mRNA (Martinez and Tuschl, 2004). Our analysis of known miRNA–mRNA duplexes revealed trends of mismatch at position 9 or 10 and relatively high free energy at nts 13 and 14. From those findings, we suggest that the miRNAs avoid forming perfect matches at positions that might lead to cleavage of target mRNAs, in order not to mark target mRNAs for degradation but instead to regulate their messages.

4.2. Multiple miRNA-binding sites are located on each target mRNA

We detected that, in nematodes, as many as 40% of all of the target transcripts contained multiple binding sites for miRNAs (data not shown). From this result, we suggest that many target mRNAs are regulated by miRNAs with multiple binding sites. This feature also occurs in documented miRNA–mRNA duplexes, in which all of the target mRNAs have more than two binding sites for either the same or different miRNA(s). For example, *lin-14* has a total of 10 target sites, seven for *lin-4* (Wightman et al., 1993; Ha et al., 1996) and three for *let-7* (Reinhart et al., 2000). We assume that specificity of miRNAs in regard to the developmental stage (as for *lin-4* and *let-7*) or the tissue is required for appropriate regulation of target mRNAs.

We also noticed that some candidate mRNAs contain multiple target sites for the same miRNA, a pattern repeated among

known miRNA–mRNA duplexes. For example, two target sites in the *let-7* miRNA-*lin-41* mRNA duplex are essential for target gene regulation (Vella et al., 2004), suggesting that some miRNAs require multiple target sites in order to regulate a message. Our analysis of the number of target sites for each miRNA revealed that 86% (1562 out of 1814 target sites) of the miRNAs seemed to have one target site per mRNA, suggesting that miRNAs with single target sites work in combinations with other miRNAs to regulate the target mRNAs.

MiRNAs are known to fall into number of families, and miRNAs in the same family are very similar in their 5′ parts from nucleotides 1 through 8 (Griffiths-Jones, 2004). To check the effect that this family similarity might have on *O*/*E* values, we chose one miRNA from each family and repeated the analysis. No difference was detected within the pattern of *O*/*E* values (data not shown).

We extracted candidates with the Gene Ontology classification of "development (GO:0007275)" and with three or more target sites and listed them (Table 1). There were some examples of alternatively spliced isoforms, such as those of *F32A7.5*, which were regulated by exactly the same miRNAs, whereas there were other examples, such as *C34F11.3b*, where only one of the spliced isoforms was regulated by miRNAs. These results suggest that alternative splicing may contribute to miRNA-directed target mRNA regulation, by splicing-out miRNA target sites for some of the spliced isoforms. A full list of extracted candidates is available as supplementary material.

In conclusion, our miRNA-based pattern searches detected a region with strong binding at nts 1 through 8 of the miRNA 5′ region and mismatch at the center of miRNA–mRNA duplexes, suggesting essential mismatch especially at nt 9 or nt 10 and weak binding at nts 13 and 14. These results confirmed our understanding of how miRNAs bind to their target mRNAs within *C. elegans*, and enabled us to extract a list of candidate miRNA target mRNAs within *C. elegans*. Moreover, from our bioinformatic analysis of predicted miRNA targets, we suggest that many mRNAs may be regulated by multiple miRNAs.

Acknowledgments

We thank Dr. Yasuhiro Naito (Keio University) for his helpful discussions. This research was supported in part by grants from the Project for Development of a Technological Infrastructure for Industrial Bioprocesses on R&D of New Industrial Science and Technology Frontiers (Ministry of Economy, Trade, and Industry of Japan) and from the New Energy and Industrial Technology Development Organization of Japan; a Grant-in-Aid for Scientific Research on Priority Areas; a Grant-in-Aid for 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; and a grant from Keio University.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2005.09.035.

References

Abrahante, J.E., et al., 2003. The *Caenorhabditis elegans* hunchback-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. Dev. Cell 4, 625–637.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., Jewell, D., 2003. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr. Biol. 13, 807–818.

Ashburner, M., et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.

Banerjee, D., Slack, F., 2002. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. BioEssays 24, 119–129.

Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 116, 281–297.

Birney, E., et al., 2004. Ensembl 2004. Nucleic Acids Res. 32, D468–D470.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., Cohen, S.M., 2003. *Bantam encodes* a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. Cell 113, 25–36.

Brennecke, J., Stark, A., Russell, R.B., Cohen, S.M., 2005. Principles of microRNA-target recognition. PLoS Biol. 3, e85.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S., 2003. MicroRNA targets in *Drosophila*. Genome Biol. 5, R1.

Fleiss, J.L., 1979. Confidence intervals for the odds ratio in case-control studies: the state of the art. J. Chronic Dis. 32, 69–77.

Freier, S.M., et al., 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. U. S. A. 83, 9373–9377.

Griffiths-Jones, S., 2004. The microRNA registry. Nucleic Acids Res. 32, D109–D111.

Ha, I., Wightman, B., Ruvkun, G., 1996. A bulged *lin-4/lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. Genes Dev. 10, 3041–3050.

Harris, T.W., et al., 2004. WormBase: a multi-species resource for nematode biology and genomics. Nucleic Acids Res. 32, D411–D417.

He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev., Genet. 5, 522–531.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human microRNA targets. PLoS Biol. 2, e363.

Johnston, R.J., Hobert, O., 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature 426, 845–849.

Kasprzyk, A., et al., 2004. EnsMart: a generic system for fast and flexible access to biological data. Genome Res. 14, 160–169.

Ke, X.S., Liu, C.M., Liu, D.P., Liang, C.C., 2003. MicroRNAs: key participants in gene regulatory networks. Curr. Opin. Chem. Biol. 7, 516–523.

Kiriakidou, M., et al., 2004. A combined computational–experimental approach predicts human microRNA targets. Genes Dev. 18, 1165–1178.

Lai, E.C., 2002. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. 30, 363–364.

Lai, E.C., Tomancak, P., Williams, R.W., Rubin, G.M., 2003. Computational identification of *Drosophila* microRNA genes. Genome Biol. 4, R42.

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294, 858–862.

Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862–864.

Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell 75, 843–854.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B., 2003. Prediction of mammalian microRNA targets. Cell 115, 787–798.

Lim, L.P., et al., 2003. The microRNAs of *Caenorhabditis elegans*. Genes Dev. 17, 991–1008.

Lim, L.P., et al., 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769–773.

Lin, S.Y., et al., 2003. The *C. elegans* hunchback homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target. Dev. Cell 4, 639–650.

Martin, D., Brun, C., Remy, E., Mouren, P., Thieffry, D., Jacq, B., 2004. GOToolBox: fucntional analysis of gene datasets on Gene Ontology. Genome Biol. 5, R101.

Martinez, J., Tuschl, T., 2004. RISC is a 5′ phosphomonoester-producing RNA endonuclease. Genes Dev. 18, 975–980.

Moss, E.G., Lee, R.C., Ambros, V., 1997. The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the *lin-4* RNA. Cell 88, 637–646.

Rehmsmeier, M., Steffen, P., Hochsmann, M., Giegerich, R., 2004. Fast and effective prediction of microRNA/target duplexes. RNA 10, 1507–1517.

Reinhart, B.J., et al., 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403, 901–906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P., 2002. MicroRNAs in plants. Genes Dev. 16, 1616–1626.

Stark, A., Brennecke, J., Russell, R.B., Cohen, S.M., 2003. Identification of *Drosophila* MicroRNA targets. PLoS Biol. 1, E60.

Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., Slack, F.J., 2004. The *C. elegans* microRNA let-7 binds to imperfect *let-7* complementary sites from the *lin-41* 3′UTR. Genes Dev. 18, 132–137.

Wightman, B., Ha, I., Ruvkun, G., 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. Cell 75, 855–862.