

# MicroRNA-Target Binding Structures Mimic MicroRNA Duplex Structures in Humans

Xi Chen<sup>1</sup>, Lu Shen<sup>2</sup>, Hui-Hsien Chou<sup>1,3\*</sup>

**1** Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa, United States of America, **2** Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Department of Computer Science, Iowa State University, Ames, Iowa, United States of America

## Abstract

Traditionally, researchers match a microRNA guide strand to mRNA sequences using sequence comparisons to predict its potential target genes. However, many of the predictions can be false positives due to limitations in sequence comparison alone. In this work, we consider the association of two related RNA structures that share a common guide strand: the microRNA duplex and the microRNA-target binding structure. We have analyzed thousands of such structure pairs and found many of them share high structural similarity. Therefore, we conclude that when predicting microRNA target genes, considering just the microRNA guide strand matches to gene sequences may not be sufficient — The microRNA duplex structure formed by the guide strand and its companion passenger strand must also be considered. We have developed software to translate RNA binding structure into encoded representations, and we have also created novel automatic comparison methods utilizing such encoded representations to determine RNA structure similarity. Our software and methods can be utilized in the other RNA secondary structure comparisons as well.

**Citation:** Chen X, Shen L, Chou H-H (2014) MicroRNA-Target Binding Structures Mimic MicroRNA Duplex Structures in Humans. PLoS ONE 9(2): e88806. doi:10.1371/journal.pone.0088806

**Editor:** Junming Yue, The University of Tennessee Health Science Center, United States of America

**Received:** November 12, 2013; **Accepted:** January 15, 2014; **Published:** February 13, 2014

**Copyright:** © 2014 Chen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is supported by National Science Foundation Grant DBI-0850195, ([http://www.nsf.gov/awardsearch/showAward?AWD\\_ID=0850195](http://www.nsf.gov/awardsearch/showAward?AWD_ID=0850195)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: hhchou@iastate.edu

## Introduction

MicroRNAs (miRNAs) are small noncoding RNAs about 21 nucleotides (nt) in length that regulate gene expressions. miRNAs target messenger RNAs (mRNAs) and trigger either their translational repression or degradation [1–4]. After pre-miRNA processing, mature miRNA duplexes are loaded into the Argonaute (Ago) proteins within the RNA-Induced Silencing Complex (RISC) [5–7]. The assembly activates RISC, which is then directed to target mRNAs [8–10]. Before RISC assembly, small RNA duplexes, which include both miRNAs and small interfering RNAs (siRNAs) in some species, can be sorted according to their structure features [11–14]. This sorting process is species-specific and determines into which Ago protein will each duplex be loaded. In *Drosophila*, if mismatches happen on the central 9–10th locations of a guide strand counted from the 5' end, the small RNA duplex will be loaded into its Ago1 protein [11]. Otherwise, it will be loaded into *Drosophila* Ago2 protein [11]. However, it has not been observed that human has a strict small RNA sorting system. Loading of miRNAs into human Ago proteins seems random but predictable — among the 4 human Ago proteins, Ago2 is the most abundant and it interacts with the majority of miRNAs (60%) [15,16]. A recent study showed that the human miRNA hsa-miR-451 only associates with Ago2 rather than the other Ago proteins [17], which suggested that there may be some human miRNA sorting rules yet to be discovered.

After loading, the unwinding of small RNA duplexes in RISC goes through two different pathways that are either slicer-dependent or slicer-independent [18–20]. Slicer-dependent unwinding, which cleaves the passenger strand, requires extensive

base-pairing. Human Ago2, the only human Ago protein known to have slicer activity, can cut a perfectly base-paired passenger strand between the 10–11th locations [15]. The nicked passenger strand is then degraded. Most of the human miRNA duplexes have mismatches in this location, i.e., their guide and passenger strands are not perfectly complementary [1,21]. Therefore, most human miRNAs will instead go through the slicer-independent unwinding [22]. Comparing to the slicer-dependent unwinding, the slicer-independent unwinding is much slower at producing activated RISCs [18].

In the past few years, there were extensive studies on the human miRNA target recognition mechanism. Most of these studies were based on sequence comparisons [23–25]. However, considering just miRNA matches to mRNA sequences is not sufficient to correctly predict miRNA targeting sites, resulting in many false-positive candidates [26–28]. The miRNA-RISC assembly process has been shown to determine the inhibition efficiency that a miRNA can exert on its target genes [18]. It is not clear how the assembly process affects the inhibition efficiency. It could be that the assembly efficiency determined the amount of activated RISC complexes that can inhibit the target genes, or it could be that the assembly process determined the guide strand structure whose efficiency in binding to target genes is governed by thermodynamics. We hypothesize that the miRNA-RISC assembly process affects the miRNA-target recognition efficiency at the guide strand structural level, i.e., miRNA target prediction precision may be improved by considering sequence matches as well as RNA structure comparisons.

In this work we have tested our hypothesis by studying the correlation of known miRNA duplex structures with known miRNA-target binding structures. All selected miRNA duplexes have experimentally validated guide strand and passenger strand sequences, allowing their duplex structures to be inferred. All selected miRNA-target binding structures have also been experimentally validated. The selected miRNA duplexes were then precisely paired with the validated miRNA-target binding structures — the same miRNA guide strand must be shared between the two structures in any pair. The similarity between the paired structures was then measured by a common RNA structure comparison method, and we have found that 69% of the pairs exhibit high similarity. An alternative statistical measurement was also applied and it revealed that most paired structures share common structural patterns that could not have been generated randomly. Further analysis indicated that this close resemblance of miRNA-target binding and miRNA duplex structures is unrelated to the evolutionary age of the miRNAs. Therefore, our study concludes that the miRNA-target binding structure closely mimics the miRNA duplex structure, and this relationship may need to be considered to improve miRNA target prediction.

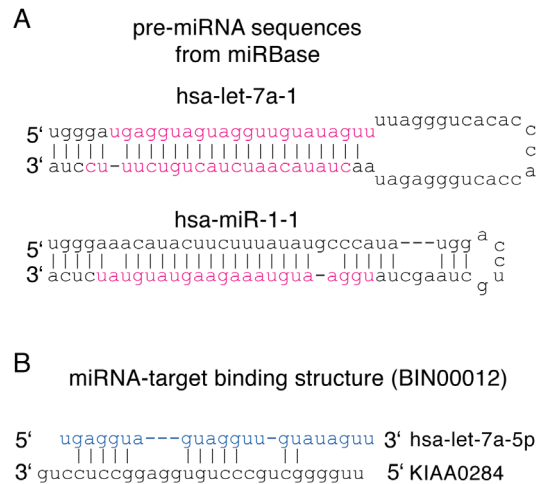
## Results

### Selection of paired miRNA duplexes and miRNA-target binding structures

Since our purpose is to compare the structure similarity between miRNA duplexes and miRNA-target binding structures, it is required to collect these structures and pair them together for this study. Human miRNA duplex sequences were selected from release version 18 of the public database miRBase, which is the most comprehensive miRNA database and includes 1527 published human miRNA duplexes (<http://www.mirbase.org>) [29]. Some mature miRNA duplexes in this database have experimentally verified sequences of both strands but others have only a single verified strand (Figure 1A). To infer the duplex structure of a mature miRNA, we need both strand sequences. As a result, 547 mature miRNA duplexes having both strand sequences were selected from miRBase (Table S1). Human miRNA-target binding structures were selected from the experimental results of Helwak *et al.* [30]. They provided a high confident data set of 18502 miRNA-target binding structures from 413 miRNAs (Figure 1B and Table S2). Because each miRNA normally have multiple target genes, the set of miRNA-target binding structures is significantly larger than the number of participating miRNAs.

Intersecting the set of 547 miRNA duplexes having both strand sequences and the set of 413 miRNAs forming the 18502 validated miRNA-target binding structures produced 321 *unique* miRNA duplexes (Figure 2 and Table S3). The *uniqueness* here refers to the duplex structure, not the name of a duplex. For example, the longer pre-miRNA sequences of hsa-let-7a-1 and hsa-let-7a-3 are different, thus their different names, but their shorter mature miRNA duplexes are the same; both have the same guide strand *hsa-let-7a-5p* and the same passenger strand *hsa-let-7a-3p*. To avoid redundancy, only one of the duplicated mature miRNA duplexes was retained in the 321 unique miRNA duplexes. The guide strand within a miRNA duplex was determined according to the miRNA-target binding structure data — the strand that binds to target mRNAs was identified as the guide strand.

After the 321 unique miRNA duplexes were identified, the miRNA-target binding structures associated with them were also selected. Several miRNA duplexes have the same guide strand but different passenger strands (they are considered different among



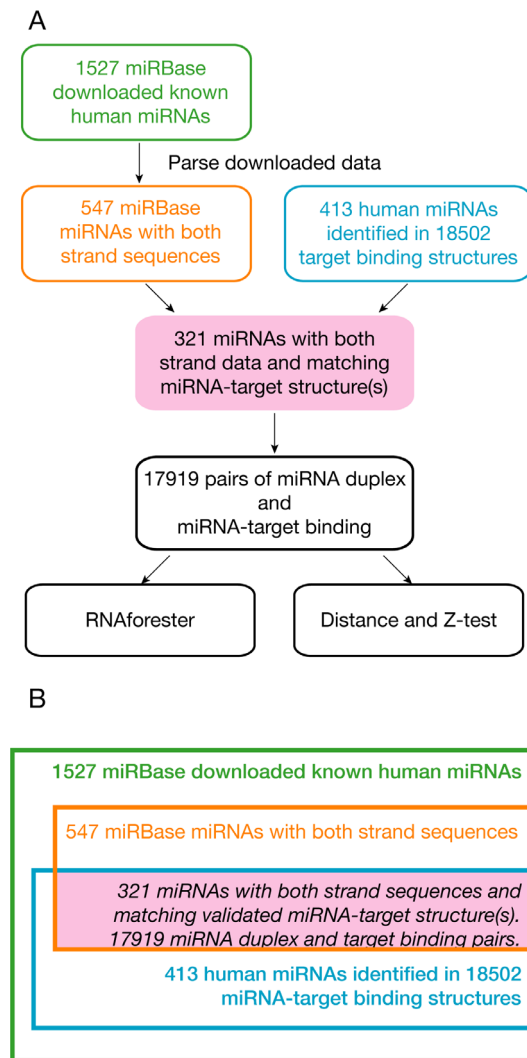
**Figure 1. Examples of miRNA structure selections.** (A) miRBase includes mature miRNAs with either two known strands like hsa-let-7a-1 (2 red sequences) or only one known strand like hsa-miR-1-1 (1 red sequence). To infer the miRNA duplex structure, we have selected only mature miRNAs with both strand sequences. (B) miRNA-target interaction data were obtained from the paper of Helwak *et al.* In this example, *hsa-let-7a-5p* is the guide strand of hsa-let-7a-1 (colored blue), and its validated binding with one of its targets, KIAA0284, has the structure as shown.  
doi:10.1371/journal.pone.0088806.g001

the 321 duplexes). The miRNA-target binding structures associated with the same guide strand were thus paired to all these miRNA duplexes. For example, the guide strand *hsa-let-7a-5p* has 310 different target bindings (Table S2). *hsa-let-7a-5p* is shared by two different miRNA duplexes hsa-let-7a-1 and hsa-let-7a-2. Therefore, 620 structure pairs, the first 310 associated with hsa-let-7a-1 and the other 310 associated with hsa-let-7a-2, were produced. In total, 17919 miRNA duplex and miRNA-target binding structure pairs were produced (Figure 2 and Table S4).

### Global features of miRNA duplexes and miRNA-target binding structures

As mentioned in the Introduction, the way a small RNA duplex will unwind within RISC depends on its mismatch locations. If mismatches happen on the central 8–12th locations of its guide strand, the small RNA duplex will follow the slicer-independent unwinding pathway. Alternatively, if its passenger strand is perfectly complementary to its guide strand, the small RNA duplex will take the slicer-dependent unwinding pathway [18–20]. Therefore, mismatch patterns on the guide strand generally have been used in RNA duplex studies [18,31–33]. However, considering only mismatches on the guide strand ignores subtle signals on the passenger strand that may also be important. For example, the pre-miRNA hsa-miR-1-1 on Figure 1A has 2 mismatched adenines on its guide strand shown in red color. When counting only mismatches on its guide strand, both adenines will be counted as 1 mismatch despite their different mismatch patterns on the passenger strand. Therefore, to better compare the global features of miRNA duplexes and miRNA-target binding structures, we have decided to consider the *base-paired* locations on *both* strands.

Figure 3A illustrates the base-pairing distribution of the 321 miRNA duplexes we have selected. The distribution can be visualized in 3 different ways: a 3-D surface graph, a 2-D contour map and a 1-D projection of the base-pairing counts onto the guide strand axis. Note that the 1-D projection adds up all base-



**Figure 2. Selection of miRNA duplexes and miRNA-target binding structures for pairwise comparisons.** (A) The data selection workflow of miRNA duplexes and miRNA-target binding structures from miRBase and the paper of Helwak *et al.* (B) The selection criteria and the resulted numbers of miRNA duplexes and miRNA-target binding structures. At the end, 321 miRNAs with both strand sequences and validated miRNA-target binding structures were selected and 17919 miRNA duplex and miRNA-target binding pairs were selected for pairwise comparisons.  
doi:10.1371/journal.pone.0088806.g002

pairing counts along the passenger strand axis that are projected onto the same guide strand location, thus the height of the 1-D projection is higher than the 3-D graph. The similar set of graphs obtained from the 18502 miRNA-target binding structures is provided in Figure 3B. We can see that on miRNA duplexes there is a noticeable dip of base-pairing counts roughly in the central 10–13th locations of the guide strand. The guide strand seed region (the 2–7<sup>th</sup> nucleotides) has lower base-pairing counts among miRNA-target binding structures than among miRNA duplexes; most base-pairing counts in this region is below 80% for miRNA-target binding structures. This supports the notion that seed region interactions with targets are not canonical for many miRNAs [30]. The other locations on the miRNA-target binding structures also have lower base-pairing counts than the duplexes, but their counts remain high at over 60% throughout most of the structures.

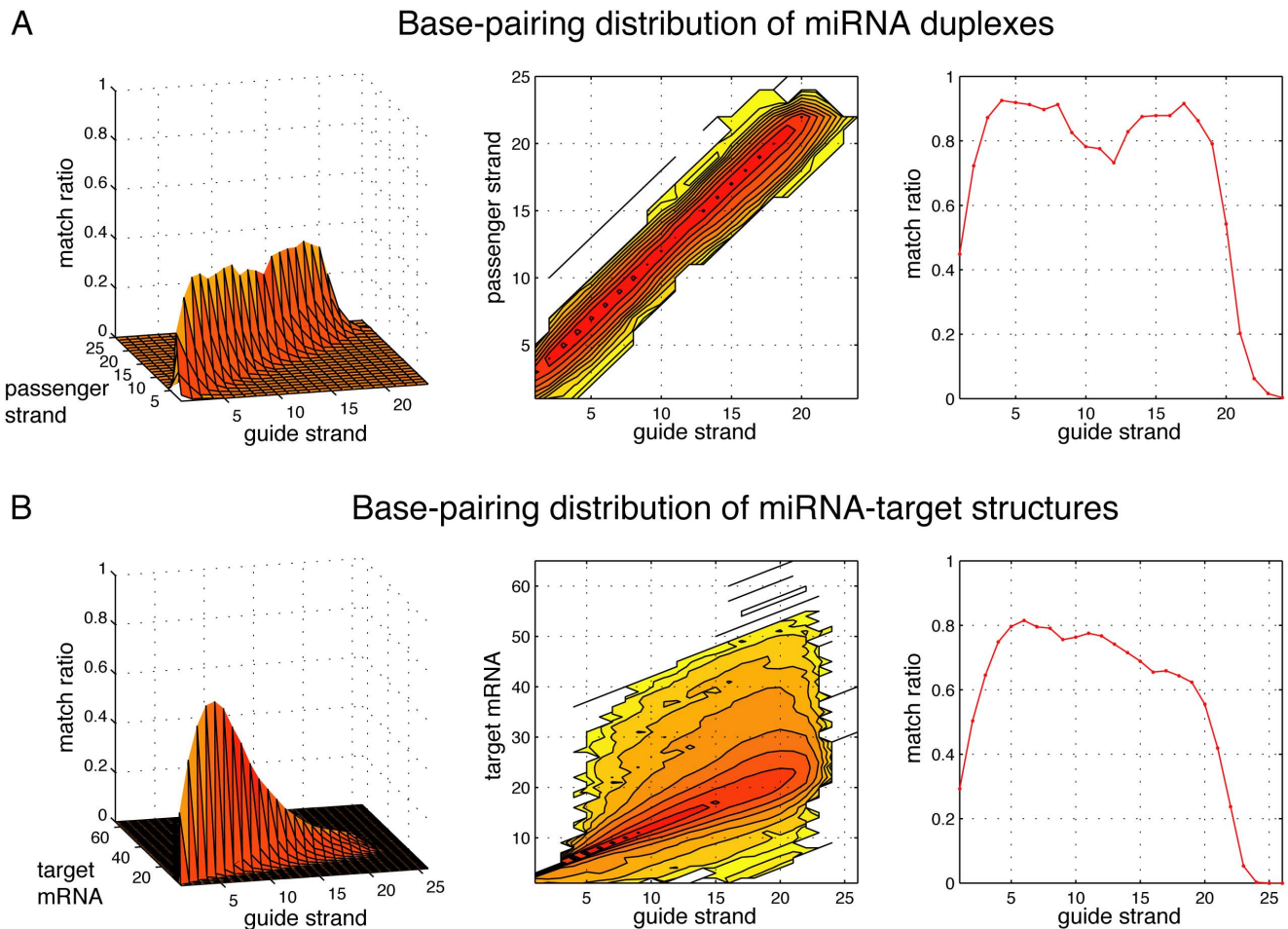
The analysis above supports the validity of our method to count the *base-paired* locations on *both* strands instead of the mismatched locations only on the guide strand, because it reveals many important miRNA features as follows. Among miRNA-target binding structures, the seed region of the guide strand seems to have mixed canonical and non-canonical match patterns, which is consistent with existing knowledge [34,35]. In addition, miRNA duplexes tend to have lower base-pairing counts in its central region, which is also consistent with previous studies [36,37]. Noticeably, non-seed region base-pairing counts are significantly lower on miRNA-target binding structures than on miRNA duplexes. A possible explanation is that a miRNA must be packaged well in double-stranded format within pre-miRNA before it can be successfully processed and incorporated into RISC. Significant mismatches anywhere along the duplex can disturb the stability of its structure and ruin its function. However, once incorporated into RISC, the miRNA guide strand does not require perfect base-pairing beyond the seed region to target genes. The global features of miRNA duplexes and miRNA-target binding structures seem to suggest that they are not very different in the seed region but are noticeably different in the non-seed region. What we need to find out next is if such distinction holds at the individual structure pair level.

### Pairwise structural comparisons between miRNA duplexes and miRNA-target bindings

To study if the distinctive base-pairing distributions of miRNA duplexes and miRNA-target binding structures hold at individual miRNA level, we systematically compared the 17919 pairs of miRNA duplex and miRNA-target binding structure that share a common guide strand. The RNAforester software was applied on the 17919 pairs to obtain the relative similarity score between the two members in each pair. Remarkably, nearly 69% of the pairs (12006/17919) have high similarity scores above 0.7 in a range of 0–1 (Figure 4A). The score cutoff value 0.7 was noted by RNAforester authors to indicate significant structure similarity [38]. In each of the 12006 pairs with high scores, the miRNA duplex structure is similar to its miRNA-target binding structure (Figure 4D, left). A closer way to look at this phenomenon is to individually consider the top 200 guide strands from the selected 321 miRNA duplexes that have the most number of targets, thus they have received the most number of structure similarity scores. Figure 4C reveals that most of their similarity scores are above 0.7 (green color). Among all 249 guide strands of the selected 321 miRNA duplexes that have received more than one similarity scores, 78% have received more scores above 0.7 than below, and only 14% have received more scores below 0.7 than above (Figure 4D, right). This indicates that miRNA-target binding structures generally mimic the associated miRNA duplex structure sharing the same guide strand.

We have developed an alternative method to compare the structure similarity within a pair of miRNA duplex and miRNA target binding structure that share the same guide strand. In this novel method, the guide strand match pattern in each structure was converted into a series of numbers: matched bases were set to 1 and unmatched bases were set to 0 (Figure 5). As a result, two series of binary numbers were generated from each pair of structures. Subsequently, we calculated the length-scaled Hamming distance between these two series of numbers: distance values closer to 0 indicate the two series are very similar, distance values closer to 1 indicate the two series are almost exactly opposite, and distance values near 0.5 indicate the two series are likely unrelated. The mean of the 17919 distance values is 0.354 (Figure 4B, red line), and the 1<sup>st</sup> quartile and 3<sup>rd</sup> quartile are at 0.272 and 0.435





**Figure 3. The base-pairing distribution of miRNA duplexes and miRNA-target binding structures.** (A) The distribution corresponding to the 321 miRNA duplexes that is drawn in a 3-D surface graph, a 2-D contour map and a 1-D projection onto the guide strand locations. (B) The distribution corresponding to the 18502 miRNA-target binding structures obtained from the Helwak *et al.* paper that are drawn in similar graphs. doi:10.1371/journal.pone.0088806.g003

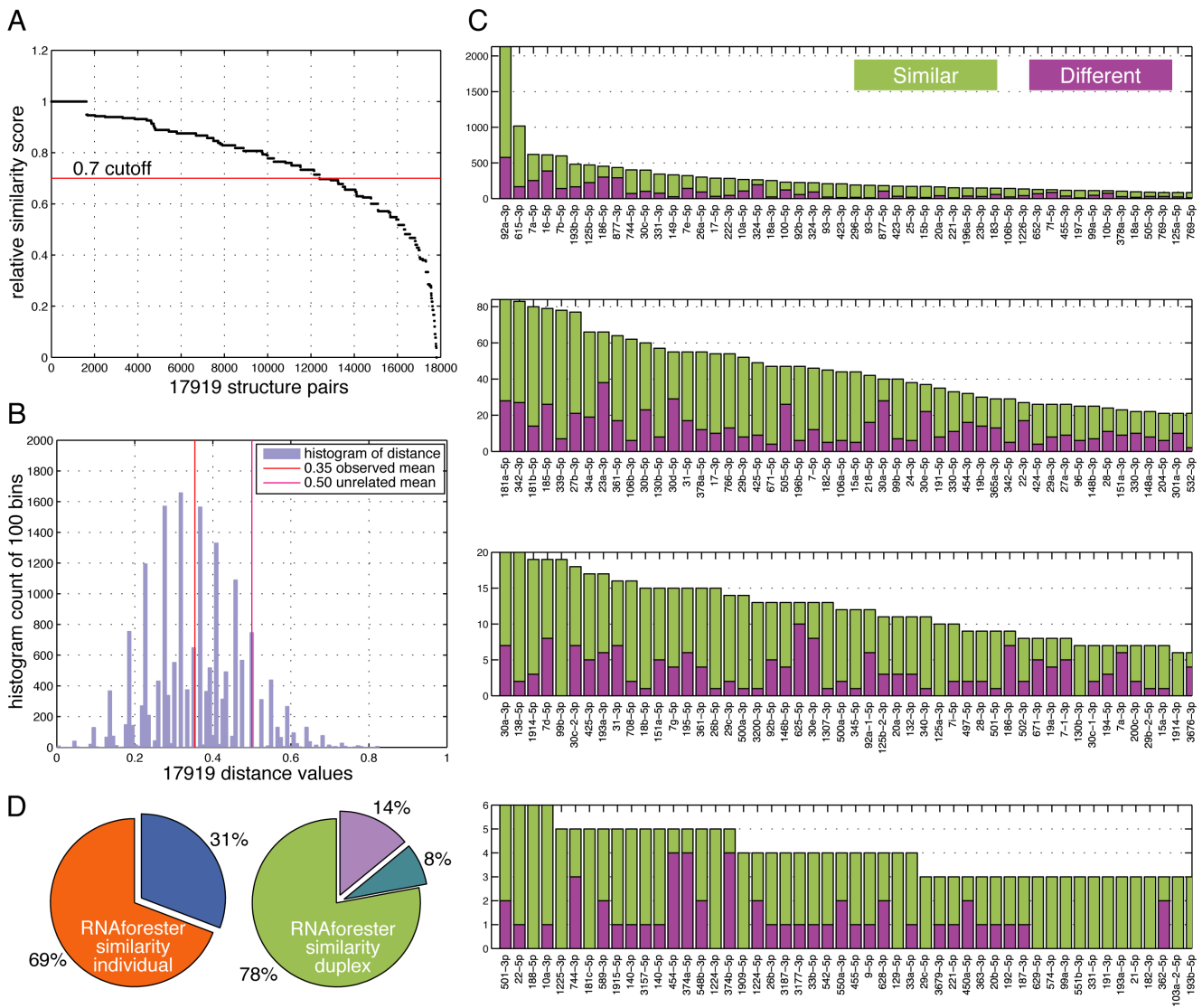
respectively. Z-test was applied to test whether 0.354 is significantly different from the expected mean value of 0.50 if the two number series in all 17919 pairs were completely random and unrelated to each other (Figure 4B, magenta line). The Z-test revealed a p-value less than 0.0001, thus we can conclude that most paired structures share a similar pattern and this similarity is not random. Taken together, our data suggest that the two types of structures, a miRNA duplex and its associated target binding structures, tend to be very similar when compared individually.

### The high pairwise structural similarity is unrelated to miRNA evolution

miRNAs are ancient. In the animal kingdom, miRNAs were present at the dawn of metazoan [39–42]. Two miRNA expansions were observed at the basin of bilaterian lineage and vertebrate lineage [41]. With the advent of next-generation sequencing, more miRNAs were discovered from various species. These studies suggested that the number of miRNAs in each species is closely correlated with its morphological complexity [36,37]. So far, more than one thousand miRNAs have been found in humans according to the data from miRBase [43]. Some miRNAs are highly conserved whereas others are younger — they emerged just recently [44]. A total of 304 unique guide strands can

be identified from the 321 miRNAs involved in the 17919 pairs selected for structure comparisons (see Figure 2 and Materials and Methods). We have selected for analysis 15 representative metazoan species that are from shorter to longer evolutionary distances from the humans. The number of known homologs among the 304 human miRNA guide strands in each of the 15 species is summarized in Figure 6A. None of the 304 guide strands has homologs in *Amphimedon queenslandica*, and only one, hsa-miR-100, has homologs in both bilaterian species and *Nematostella vectensis* — this is consistent with previous studies [45]. Most human miRNA guide strands have homologs in the other four vertebrates: *Mus musculus*, *Gallus gallus*, *Xenopus tropicalis* and *Danio rerio*. The number of miRNA guide strands that have homologs in protostomia decreases, which is also consistent with the notation that when some miRNAs became important in a particular lineage, they are rarely lost in the descendant lineages [39,46].

When the 304 human miRNA guide strands are individually considered, about 85% of them have homologs in less than 6 other species (Table S5). The logical question to ask is whether the conclusion we have drawn in the previous section, that a miRNA duplex and its associated target binding structures tend to be similar, is only applicable to younger miRNAs? To answer this



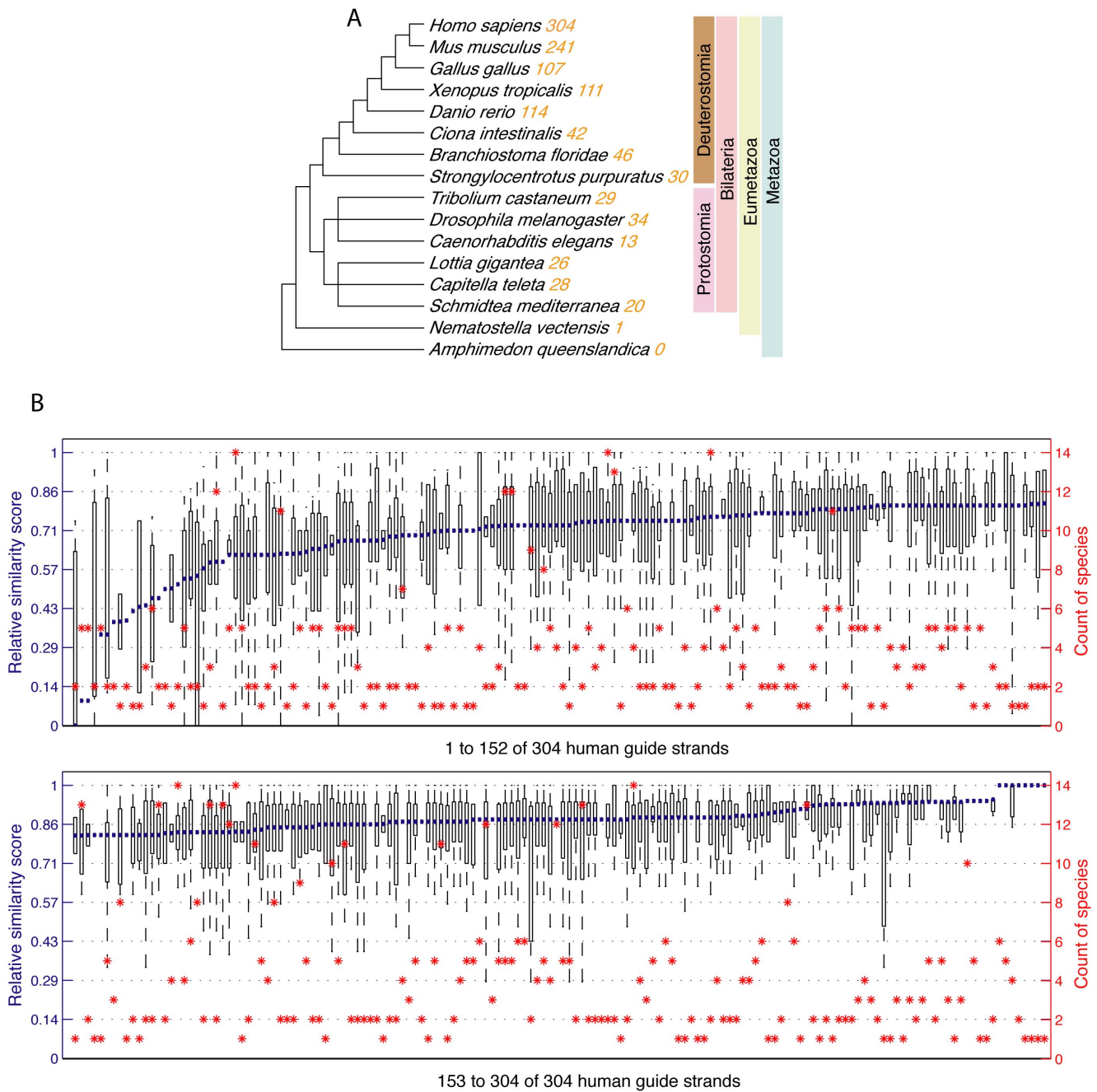
**Figure 4. Pairwise comparison of miRNA duplexes and miRNA-target binding structures.** (A) The similarity scores of the 17919 structure pairs were calculated by RNAforester and ordered from high to low; scores above 0.7 are considered similar. (B) The histogram of the length-scaled Hamming distance values of the 17919 structure pairs divided into 100 bins (bar chart) with the mean value at 0.354 (red line); the expected mean value is at 0.50 for random structure pairs (magenta line). (C) The similarity score ratio for each miRNA duplex. Many miRNA duplexes were paired with multiple target bindings and thus received multiple similarity scores; the top 200 miRNA duplexes are shown here and are ordered by the count of scores received. Green: similarity score  $\geq 0.7$  and considered similar; Purple: similarity score  $< 0.7$  and considered different. Note the Y-axes on the four subgraphs were not drawn to the same scale. (D) Overall similarity ratios. Left: among the 17919 structure pairs considered, 69% are similar with RNAforester scores  $\geq 0.7$ ; Right: among the 249 miRNA duplexes that received more than 1 similarity scores, 78% have more scores  $\geq 0.7$ , 14% has more scores  $< 0.7$  and 8% have equal counts. doi:10.1371/journal.pone.0088806.g004

question, the similarity scores of the 17919 pairs of miRNA duplex and miRNA-target binding structures were mapped onto the 304 unique guide strands, which are then ordered and summarized in Figure 6B. For guide stands that received multiple mapped similarity scores, their score distributions are represented by boxplots and their median values are used in the ordering which goes from the least similar to the most similar (blue lines). Superimposed on Figure 6B is the count of the other species that have homologs to each of the ordered 304 human miRNA guide strands (red dots). From this figure it can be concluded that the similarity between miRNA duplex and miRNA-target binding structure is not correlated to miRNA evolution age.

### Discussion

The traditional understanding about miRNA is that the passenger strand is excluded during duplex unwinding and only the guide strand is needed to guide the activated RISC to target mRNAs; the passenger strand does not seem to play any role in the targeting process. However, recent studies in plants showed that the passenger strand may affect miRNA-triggered transitivity, a production of secondary siRNAs through antisense transcription [47,48]. Further study revealed that the asymmetry in miRNA duplex structure is actually triggering the transitivity and the asymmetry is determined by both the passenger and guide strands [49]. These novel observations contradict the traditional guide-strand-centric





**Figure 6. Correspondence of structure similarity to miRNA evolution.** (A) The 321 selected miRNA duplexes contain 304 distinct guide strands; their homolog counts in each of the other 15 species is listed. (B) The similarity scores received by each miRNA guide strand are drawn in a boxplot with the median value identified by a blue line; the guide strands are then ordered according to their medians. The number of the other species that contain homologs to each human guide strand is also identified by a red dot. There is no discernible correspondence between the similarity score distribution in blue lines and the evolutionary ages of miRNAs in red dots.  
doi:10.1371/journal.pone.0088806.g006

data identified the guide strand of each miRNA. Oriented by the identified guide strand, the secondary structure of each miRNA duplexes was encoded into the dot-bracket and number notations (Figure 5A). The dot-bracket notation uses parentheses and dots to represent base-paired and mismatched bases; this notation was used by the RNAForester software to compare RNA structure similarities [53]. The number notation uses 1 and 0 to represent base-paired and mismatched bases; this notation was used in the Hamming distance calculation followed by the Z-test to also

compare RNA structure similarities. There is a difference between the two notations: the dot-bracket notation encodes both strands of a miRNA duplex but the number notation encodes only the guide strand. The miRNA-target binding structures from the paper of Helwak *et al.* [30] were also determined and encoded into the dot-bracket and number notations in the same way (Figures 5B-D).

The Perl program *mirna\_structure\_notations\_v2.pl* was created to automatically encode the structures of miRNA duplexes and miRNA-target bindings into the dot-bracket and number notations.



Usage instructions can be obtained by running this program without any command-line argument. This program can also produce a global summary matrix counting the base-paired locations on the guide strand and passenger strand from all input structures. For example, if the 3rd base on the guide strand is base-paired with the 4th base on the passenger strand in a particular input structure, then the count at matrix location (3,4) will be increased by 1. Such matrices were used to create Figure 3.

### Pairwise comparison of miRNA duplex and miRNA-target binding structures

RNAforester is one of the structure comparison methods we have used [53]. This software compares two RNA secondary structures in the dot-bracket notation and determines their similarity score [53]. The Perl program mentioned above can convert the structures of miRNA duplexes and miRNA-target bindings to this notation. However, RNAforester only compares single-strand RNA foldings, not RNA duplex structures. Therefore, structures such as miRNA duplexes or miRNA-target bindings need to be converted into single-strand RNA folding representations. This is customarily accomplished by connecting together the dot-bracket notations obtained from the two strands of a duplex structure with a minimum of 3 dots to mimic a loop between them (see Figure 5) [54]. In our tests, we have found that the RNAforester similarity score is not sensitivity to the length of the loop (data not shown). To execute RNAforester solely for RNA secondary structure comparisons without also considering the primary sequence similarities, its parameter **bm** (base match score), **br** (base mismatch score) and **bd** (base indel score) were all set to 0, whereas its parameters **pm** (pairing match score) and **pd** (pairing indel score) were both set at the default values 10 and -5, respectively.

The miRNA duplex and miRNA-target binding structure in each selected pair share a common guide strand. The shared guide strand can serve as an *anchor* to determine the similarity between the two RNA secondary structures. We have created a novel method that can also determine RNA structure similarity. The Perl program mentioned above can convert the secondary structures of miRNA duplex and miRNA-target binding into a series of binary numbers. On the guide strand, a base-paired location is converted to 1 and an unmatched location is converted to 0. Subsequently, we calculated the length-scaled Hamming distance between the two series of numbers. If the two numbers on the same location is the same, 0 is added to the sum; otherwise  $1/n$  is added to the sum, where  $n$  is the length of the guide strand. The sum of all values represents the distance between the two series of binary numbers. If the two series of numbers are the same, their distance should be 0. However, if they are complete reversals of each other, their distance should be 1. Any two random series should have an expected distance close to 0.50 because at each location there is equal probability to add or not to add  $1/n$  to the sum, so the sum should approach  $1/n \times n/2 = 1/2$ . We have obtained 17919 distance values from the 17919 structure pairs. The mean of our distance values is 0.354, and the 1<sup>st</sup> quartile and 3<sup>rd</sup> quartile are at 0.272 and 0.435 respectively. This distribution is determined to be significantly different from the expected mean value of 0.50 for completely unrelated structure series after performing a Z-test ( $p$ -value < 0.0001).

### The relationship between structure similarity and miRNA evolution

To determine if the similarities of miRNA duplexes and miRNA-target binding structures are related to miRNA evolutionary age,

homologs to human miRNAs from the other 15 species were also selected from miRBase [55]. The miRNA names reflect their association to specific miRNA families. For example, the four miRNAs dre-miR-21 (*Danio rerio*), gga-miR-21 (*Gallus gallus*), mmu-miR-21 (*Mus musculus*) and hsa-miR-21 (*Homo sapiens*) all belong to the miR-21 family. We took a human miRNA-centric approach for the evolution study and considered only miRNAs from the other species that can be associated to the 321 human miRNAs selected for our study. Because only the guide strand is shared between a miRNA duplex and its miRNA-target bindings, we considered only guide strands in the evolution study. Some consolidation of guide strand data was performed. If two human miRNAs share the same guide strand, only one guide strand was retained for the study. Conversely, if both strands of a miRNA duplex are functional guide strands, both were retained for the study. For example, the two strands *hsa-let-7a-5p* and *hsa-let-7a-3p* of the human miRNA *hsa-let-7a-1* are both guide strands, but they have different homologs in the other 15 species — 13 other species contain *hsa-let-7a-5p* homologs but only 5 other species contain *hsa-let-7a-3p* homologs.

After the consolidation, 304 distinct human miRNA guide strands were retained for the evolution analysis. The non-human miRNAs were initially assigned to the 304 human guide strands based on their name. If a non-human miRNA can be assigned to different human guide strands, they will be assigned in duplicate to simplify the analysis. Because some miRNA strand names are incomplete or incorrect, all non-human miRNAs assigned to a human guide strand were subsequently sequence-aligned to the guide strand to validate the assignment. If a non-human miRNA aligned poorly to the assigned guide strand, it was manually inspected and its assignment was removed when necessary. The counts of human miRNA homologs that can be found in the other species are listed in Table S5.

### Supporting Information

**Table S1** 547 mature miRNA duplexes selected from miRBase that have both strand sequences. (XLSX)

**Table S2** 18502 validated miRNA-target binding structures that were collected from the paper of Helwak *et al.* (XLSX)

**Table S3** 321 unique miRNA duplexes from the intersection of the set of 547 miRNA duplexes having both strand sequences and the miRNAs involved in the set of 18502 validated miRNA-target binding structures. (XLSX)

**Table S4** Dot-bracket and number notations for 17919 pairs of miRNA duplex and associated miRNA-target binding structure. Note that the two structures in each pair share a common guide strand. Also listed in the last two columns of the table are the RNAforester similarity score and the Hamming distance value for each pair of structures. (XLSX)

**Table S5** The main part of the matrix listed the number of homologs of the 304 distinct human miRNA guide strands that can be found in each of the other 15 species; the 15 species were ordered according to their evolutionary distance from humans. The total number of human miRNA guide strands that have homologs in each species is summarized on the right side of the table under the column “Count of human siRNAs having homologs in each species”. The number of species covered by each of the 304 distinct human miRNA guide strand is also



summarized at the bottom row of the table labeled “The number of species covered”.  
(XLSX)

**Program S1 Perl Program (downloadable Perl code).**  
The Perl program `mirna_structure_notations_v2.pl` used to generate the dot-bracket and number notations on Table S4.  
(PL)

## Acknowledgments

We thank Dr. Xun Gu for the discussion about miRNA evolution, and our colleagues Hyejin Cho and Jennifer Chang for their assistance. We also

## References

- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215–233. doi:10.1016/j.cell.2009.01.002.
- MacFarlane L-A, Murphy PR (2010) MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics* 11: 537–561. doi:10.2174/138920210793175895.
- Djuranovic S, Nahvi A, Green R (2011) A parsimonious model for gene regulation by miRNAs. *Science* 331: 550–553. doi:10.1126/science.1191138.
- Pasquinelli AE (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* 13: 271–282. doi:10.1038/nrg3162.
- Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101: 25–33. doi:10.1016/S0092-8674(00)80620-0.
- Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell* 123: 631–640. doi:10.1016/j.cell.2005.10.022.
- Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 9: 22–32. doi:10.1038/nrm2321.
- Lewis BP, Shih I -hun., Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of Mammalian MicroRNA Targets. *Cell* 115: 787–798. doi:10.1016/S0092-8674(03)01018-3.
- Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of MicroRNA-Target Recognition. *PLoS Biol* 3: e85. doi:10.1371/journal.pbio.0030085.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. *Nat Genet* 39: 1278–1284. doi:10.1038/ng2135.
- Tomari Y, Du T, Zamore PD (2007) Sorting of Drosophila small silencing RNAs. *Cell* 130: 299–308. doi:10.1016/j.cell.2007.05.057.
- Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* 12: 19–31. doi:10.1038/nrg2916.
- Kawamata T, Tomari Y (2010) Making RISC. *Trends Biochem Sci* 35: 368–376. doi:10.1016/j.tibs.2010.03.009.
- Kaya E, Doudna JA (2012) Guided Tour to the Heart of RISC. *Science* 336: 985–986. doi:10.1126/science.1223549.
- Höck J, Meister G (2008) The Argonaute protein family. *Genome Biol* 9: 210. doi:10.1186/gb-2008-9-2-210.
- Wang D, Zhang Z, O’Loughlin E, Lee T, Houel S, et al. (2012) Quantitative functions of Argonaute proteins in mammalian development. *Genes Dev* 26: 693–704. doi:10.1101/gad.182758.111.
- Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G (2012) microRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res.* Available: <http://www.ncbi.nlm.nih.gov/pubmed/22844086>. Accessed 6 August 2012.
- Kawamata T, Seitz H, Tomari Y (2009) Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* 16: 953–960. doi:10.1038/nsmb.1630.
- Seitz H, Tushir JS, Zamore PD (2011) A 5’-uridine amplifies miRNA/miRNA\* asymmetry in Drosophila by promoting RNA-induced silencing complex formation. *Silence* 2: 4. doi:10.1186/1758-907X-2-4.
- Betancur JG, Tomari Y (2012) miRNA-like duplexes as RNAi triggers with improved specificity. *Front Non-Coding RNA* 3: 127. doi:10.3389/fgene.2012.00127.
- Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115: 209–216.
- Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123: 607–620. doi:10.1016/j.cell.2005.08.044.
- Lewis BP, Burge CB, Bartel DP (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* 120: 15–20. doi:10.1016/j.cell.2004.12.035.
- Kruger J, Rehmsmeier M (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 34: W451–W454. doi:10.1093/nar/gkl243.
- Maragkakis M, Reczek M, Simossis VA, Alexiou P, Papadopoulos GL, et al. (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res* 37: W273–W276. doi:10.1093/nar/gkp292.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, et al. (2004) Human MicroRNA Targets. *PLoS Biol* 2: e363. doi:10.1371/journal.pbio.0020363.
- Barbato C, Arisi I, Frizzo ME, Brandi R, Da Sacco L, et al. (2009) Computational Challenges in miRNA Target Predictions: To Be or Not to Be a True Target? *BioMed Res Int* 2009. Available: <http://www.hindawi.com/journals/bmri/2009/803069/abs/>. Accessed 15 April 2013.
- Witkos TM, Koscianska E, Krzyzosiak WJ (2011) Practical Aspects of microRNA Target Prediction. *Curr Mol Med* 11: 93–109.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2007) miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36: D154–D158. doi:10.1093/nar/gkm952.
- Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153: 654–665. doi:10.1016/j.cell.2013.03.043.
- Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, et al. (2008) Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci* 105: 7964–7969. doi:10.1073/pnas.0800334105.
- Ahmed F, Ansari HR, Raghava GP (2009) Prediction of guide strand of microRNAs from its sequence and secondary structure. *BMC Bioinformatics* 10: 105. doi:10.1186/1471-2105-10-105.
- Li N, You X, Chen T, Mackowiak SD, Friedländer MR, et al. (2013) Global profiling of miRNAs and the hairpin precursors: insights into miRNA processing and novel miRNA discovery. *Nucleic Acids Res.* Available: <http://nar.oxfordjournals.org/content/early/2013/02/08/nar.gkt072>. Accessed 15 April 2013.
- Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, et al. (2009) miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to “seedless” 3’UTR microRNA recognition elements. *Mol Cell* 35: 610–625. doi:10.1016/j.molcel.2009.08.020.
- Chi SW, Hannon GJ, Darnell RB (2012) An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* 19: 321–327. doi:10.1038/nsmb.2230.
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901–906. doi:10.1038/35002607.
- Zeng Y, Yi R, Cullen BR (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci* 100: 9779–9784. doi:10.1073/pnas.1630797100.
- Höschmann M (2005) The tree alignment model: algorithms, implementations and applications for the analysis of RNA secondary structures. Available: <http://pub.uni-bielefeld.de/pub?func=drec&id=2305233>. Accessed 15 April 2013.
- Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, et al. (2009) The deep evolution of metazoan microRNAs. *Evol Dev* 11: 50–68. doi:10.1111/j.1525-142X.2008.00302.x.
- Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and Functional Diversification of MIRNA Genes. *Plant Cell Online* 23: 431–442. doi:10.1105/tpc.110.082784.
- Berezikov E (2011) Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* 12: 846–860. doi:10.1038/nrg3079.
- Guerra-Assunção JA, Enright AJ (2012) Large-scale analysis of microRNA evolution. *BMC Genomics* 13: 218. doi:10.1186/1471-2164-13-218.
- Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 39: D152–157. doi:10.1093/nar/gkq1027.
- Lu J, Shen Y, Wu Q, Kumar S, He B, et al. (2008) The birth and death of microRNA genes in Drosophila. *Nat Genet* 40: 351–355. doi:10.1038/ng.73.
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, et al. (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455: 1193–1197. doi:10.1038/nature07415.
- Zhu Y, Skogerboe G, Ning Q, Wang Z, Li B, et al. (2012) Evolutionary relationships between miRNA genes and their activity. *BMC Genomics* 13: 718. doi:10.1186/1471-2164-13-718.
- Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, et al. (2010) 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci U S A* 107: 15269–15274. doi:10.1073/pnas.1001738107.

thank a prior anonymous reviewer for the suggestion to use the Helwak *et al.* data set, which dramatically improves both the quality and quantity of the miRNA-target binding data that are analyzed in our work.

## Author Contributions

Conceived and designed the experiments: XC HC. Performed the experiments: XC HC. Analyzed the data: XC LS HC. Contributed reagents/materials/analysis tools: HC. Wrote the paper: XC HC.

48. Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, et al. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol* 17: 997–1003. doi:10.1038/nsmb.1866.
49. Manavella PA, Koenig D, Weigel D (2012) Plant secondary siRNA production determined by microRNA-duplex structure. *Proc Natl Acad Sci* 109: 2461–2466. doi:10.1073/pnas.1200169109.
50. Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP, et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27: 91–105. doi:10.1016/j.molcel.2007.06.017.
51. Didiano D, Hobert O (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* 13: 849–851. doi:10.1038/nsmb1138.
52. Dimitrov RA, Zuker M (2004) Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids. *Biophys J* 87: 215–226. doi:10.1529/biophysj.103.020743.
53. Hochsmann M, Toller T, Giegerich R, Kurtz S (2003) Local similarity in RNA secondary structures. *Proceedings of the 2003 IEEE Bioinformatics Conference, 2003. CSB 2003*. pp. 159 – 168. doi:10.1109/CSB.2003.1227315.
54. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, et al. (2004) Structural Features of MicroRNA (miRNA) Precursors and Their Relevance to miRNA Biogenesis and Small Interfering RNA/Short Hairpin RNA Design. *J Biol Chem* 279: 42230–42239. doi:10.1074/jbc.M404931200.
55. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, et al. (2003) A uniform system for microRNA annotation. *RNA* 9: 277–279. doi:10.1261/rna.2183803.