FEBS Letters 585 (2011) 240–248

journal homepage: www.FEBSLetters.org

Systematic analysis of human microRNA divergence based on evolutionary emergence

Qiang-Hu Wang [1], Meng Zhou [1], Jie Sun [1], Shang-Wei Ning, Yan Li, Li Chen, Yan Zheng, Xiang Li, Sa-li Lv, Xia Li *

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

ARTICLE INFO

Article history:
Received 22 October 2010
Revised 23 November 2010
Accepted 29 November 2010
Available online 3 December 2010

Edited by Takashi Gojobori

Keywords:
Divergence
Evolutionary emergence
microRNAs

ABSTRACT

MicroRNAs (miRNAs) play important roles in post-transcriptional gene expression control. To gain new insight into human miRNAs, we performed comprehensive sequence-based homology search for known human miRNAs to study the evolutionary distribution of human miRNAs. Furthermore, we carried out a series of studies to compare various features for different lineage-specific human miRNAs. Our results showed that major expansions of human miRNA genes coincide with the advent of vertebrates, mammals and primates. Further system-level analysis revealed significant differences in human miRNAs that arose from different evolutionary time points for a number of characteristics, implicating genetic and functional diversification for different human miRNAs during evolution. Our finds provide more useful knowledge for further exploring origins and evolution of human miRNA genes.

1. Introduction

MicroRNAs (miRNAs), approximately 22 nucleotides (nt) in length, are a large family of single-stranded endogenous small RNAs that are generated via a multi-stage process involving both the nuclear and cytoplasmic compartments [1,2]. The phylogenetic distribution and evolution of animal and plant miRNAs have been studied in recent reports [2–12]. In animals, the expansion of miRNAs is an ongoing process [5]. Novel miRNA genes are continuously added to the metazoan genome and are maintained within the genome over vast stretches of evolutionary time [5,6,8]. The phylogenetic distribution of miRNAs is related to the broad-scale taxonomic hierarchy of animal relationships [6]. The diversity of lineage-specific miRNAs may contribute to the phenotypic complexity of lineages during evolution. Recently, Zhang et al. [13] examined the evolutionary dynamics of the primate-specific Chr19-linked miRNA family and suggested that this family is of functional importance to primate development. Similarly, the phylogenetic history of the miR-17 family of miRNAs, that are conserved across vertebrates, has been reported to be closely associated with the early evolution of the vertebrate lineage [14].

Inter- and intra-species divergence of miRNA genes has been investigated in recent studies [15,16]. However, little is currently known about the genetic and functional divergence potential of human miRNA genes that have emerged at different evolutionary time points. In this study, we carried out a comprehensive analysis of human miRNA gene evolutionary history in terms of individual genes by a combination of multiple homology search strategies, and further systematically explored their genetic and functional divergence, with the aim of improving our understanding of miRNAs that have emerged at different evolutionary time points in the human genome.

2. Materials and methods

2.1. miRNA genes and target genes data sets

All of the human and animal miRNA genes used in this analysis were obtained from the miRBase Sequence Database, release 15.0 (http://microrna.sanger.ac.uk/) [17]. The predicted miRNA targets were from miRbase [17], miRGen [18] and miRNAMap [19]. Sethupathy et al. who compared the widely used target prediction programs found no one program can be considered as consistently superior to the rest, and suggested that intersection of the prediction results of these programs can yield improved specificity with only a marginal decrease in sensitivity relative to any individual program [18,20]. To reduce false positives, only targets that appeared in at least two databases were used in our analysis.

* Corresponding author. Fax: +86 045186615922.
E-mail address: lixia@hrbmu.edu.cn (X. Li).
[1] These authors contributed equally to this work.

0014-5793/$36.00
doi:10.1016/j.febslet.2010.11.053

2.2. Phylogenetic analysis of human miRNAs

To determine the phylogenetic distribution of human miRNA genes and infer the likely evolutionary origin, we performed comprehensive sequence-based homology search for known human miRNAs to detect both closely-related and distantly-related homologs. First, the human miRNA hairpin sequences were aligned against all animal hairpin sequences using standalone BLAST (blastn, version 2.2.27). The hairpin sequences were considered homologous when they exhibited minimum 85% sequence identity over an alignment length of at least 70% [21]. Second, we performed alignments by both 5′end matching (nt 2–8) and homology analysis across complete mature miRNA sequences, comparing human mature miRNA sequences against all known animal mature miRNA sequences. The mature sequences were accepted as homologs of a human mature miRNA when positions 2–7 of the seed sequence must be identical and the remainder of the alignment may contain no more than three mismatches [22]. Since the data may be biased because a few species are represented in miRBase and independent surveys for miRNAs have been conducted only for a few organisms, so we further carried out exhaustive homology search carefully. All human miRNA hairpin sequences were blasted against the available metazoan genomes available in UCSC Genome Bioinformatics (http://genome.ucsc.edu) [23].

2.3. Genomic enrichment analysis

The genome coordinates of human miRNAs were determined from the miRBase Sequence Database (http://www.mirbase.org) [17] and the UCSC Genome Bioinformatics (http://genome.ucsc.edu) [23]. To examine the genomic distribution of human miRNA genes that emerged at different evolutionary time points and those of their corresponding targets, we performed genomic enrichment analysis based on the cumulative hypergeometric test, using the formula:

$$P(X > m) = 1 - \sum_{i=1}^{m} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the number of all miRNA genes or target genes, n is the number of miRNA genes or target genes located on a specific chromosome, M is the number of lineage-specific miRNA genes or target genes regulated by lineage-specific miRNAs, m is the number of lineage-specific miRNA genes located on a specific chromosome or the number of target genes regulated by lineage-specific miRNAs located on a specific chromosome. P-values were calculated to indicate if miRNA genes had emerged at different evolutionary time points or if their targets are enriched on specific chromosome. A significance level of 0.01 was set in our analysis.

2.4. The human miRNA disease database

The relationship between miRNAs and disease was retrieved from the miR2Disease database (http://www.mir2disease.org/) which is a manually curated database and provided a comprehensive resource of miRNA deregulation in various human diseases [24]. We counted the number of different lineage-specific human miRNAs that associated with at least one disease and the number of diseases that associated with different lineage-specific human miRNAs.

2.5. Network analysis

The human protein–protein interaction data were obtained from the Human Protein Reference Database (HPRD; http://www.hprd.org/) [25]. For each set of lineage-specific miRNA genes, we defined two types of microRNA-regulated PPI network, termed as L0 network and L1 network. L0 network are composed of all of the target proteins regulated by lineage-specific miRNA genes. Then L0 network are extended by adding the interacting proteins with target proteins, noted as L1 network. To characterize the roles that different lineage-specific human miRNAs play in the human PPIN, we performed systematic analysis to compare topological properties of networks regulated by lineage-specific miRNA genes such as degree, clustering coefficient, betweenness centrality, closeness centrality and characteristic path length.

3. Results and discussion

3.1. Evolutionary distribution of human miRNAs

miRNA genes have been found to be widely distributed in animals [11]. Many previous studies have characterized the phylogenetic distribution and evolutionary dynamics of miRNA genes across various metazoan genomes [5,6,8,11,12,22,26,27]. Here, we mainly focused on examining phylogenetic distribution of human miRNA genes and inferring the likely evolutionary origin. To study the evolutionary distribution of human miRNA genes listed in the miRBase database, we first traced the phylogenetic history of these genes and determined evolutionary emergence by a combination of multiple homology search strategies. Then we mapped each human miRNA gene to the ancestral phylogenetic branch. According to the time of evolutionary emergence, all human miRNAs genes and miRNA families were classified into nine distinct age groups: metazoan, bilateria, deuterostoma, chordate, vertebrata, mammalian, primates, hominidae and human-specific. The evolutionary distribution of human miRNA genes are

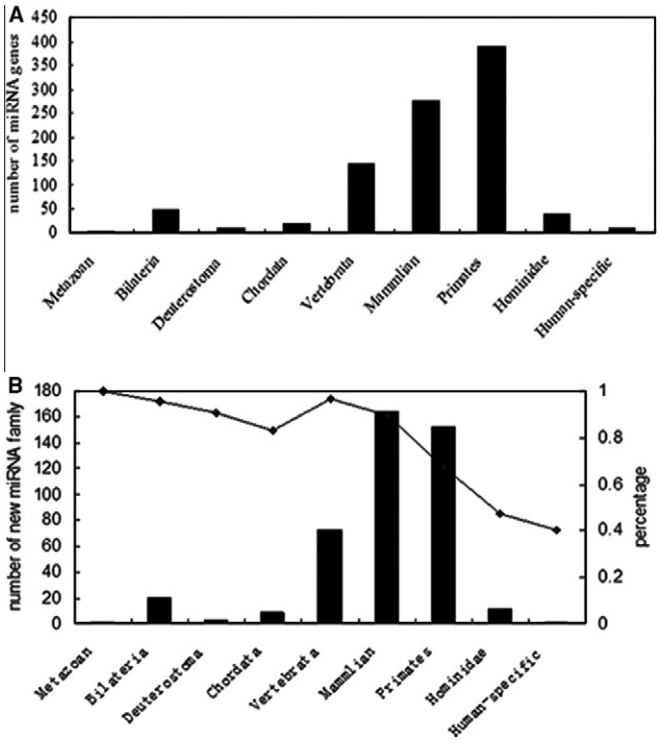

Fig. 1. Evolutionary distribution of human miRNA genes and miRNA families. (A) Evolutionary distribution of human miRNA genes. (B) Evolutionary distribution of miRNA families. Black pane represents number of new miRNAs family and blue pane represents the ratio of the number of miRNAs in new family to total number of miRNAs in each group. The information of animal miRNA family is from the miRBase Sequence Database.

Table 1
The evolutionary distribution of human miRNA families.

	Metazoan	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalian	Primates	Hominidae	Human-specific
miRNAs[a]	2	49	11	18	144	276	390	40	10
miRNA family[b]	1	21	3	9	72	164	153	12	1
New miRNAs[c]	2	47	10	15	139	247	263	19	4
Ratio[d]	100%	95.9%	90.9%	83.3%	96.5%	89.5%	67.4%	47.5%	40%

[a] The number of miRNAs arising from different evolutionary lineages.
[b] The number of miRNA family arising from different evolutionary lineages.
[c] The number of miRNA in new miRNA family.
[d] The ratio of miRNA in new miRNA family to the number of miRNAs arising from different evolutionary lineages.

presented in Fig. 1. We found that miR-99a and miR-100, belonging to miR-99 family, is the most ancient miRNAs and can be traced back to metazoan. 49, 11 and 18 human miRNAs were found to originate from the bilateria, deuterostoma and chordata lineages, respectively. As shown in Fig. 1A, human miRNA genes emerged from the vertebrate lineage showed a marked increase and reached a peak in the primate lineage. Hertel et al. [5] demonstrated that major expansions of the metazoan miRNA emergence coincides with the advent of bilaterians, vertebrates and mammals. However, in our study, major expansions of human miRNA genes coincide with the advent of vertebrates, mammals and primates. Furthermore, a burst expansion of human miRNA genes was observed in primates. Sempere et al. [6] identified three miRNAs that are unique to the human genome. Here we extended their results and identified 10 human-specific miRNAs whose homologs were not found in other animal genomes. These data suggested that most of human miRNA genes are younger in animal evolutionary history, and strongly support the idea that evolutionary young miRNAs are a common phenomenon [5,28,29]. Meantime, these results also suggested that emergence of miRNA genes is an ongoing progress and novel miRNA genes are continuously added to the metazoan genome over time with little secondary loss [5,6,22].

Moreover, we further studied the evolutionary distribution of human miRNA families (Fig. 1B and Table 1), and found that major expansions of the new human miRNA family emergence also coincides with the advent of vertebrates, mammals and primates. Heimberg et al. [8] found that a greatly expanded set of miRNAs have occurred in vertebrates as compared with bilaterian invertebrates and also identified 41 miRNA families evolved at the base of vertebrate. However, in our study, we revealed 72 miRNA families emerged from the vertebrate lineages, and further found that the number of the new human miRNA family reached a peak in the mammalian lineage. The evolution of these novel miRNA families is intimately tied to the evolution of novel cell types [6,8,22]. Moreover, we also observed that a small fraction of novel miRNA emerged at different time points are copies of members of ancestral families except for metazoan and human-specific. The emergence of these new miRNA genes which are copies of members of ancestral families could be due to genome duplication events (GDEs), including the occurrence of tandem duplications and non-local duplications, which increased diversity of miRNA genes [22]. Further analysis revealed that a large portion of novel human miRNA genes emerged at different time points are copies of new families, suggesting that novel miRNA genes often experience genome duplication events. We further examined the ratio of the number of miRNAs in new family to total number of miRNAs in each group (Fig. 1B). As shown in Fig. 1B, we also observed that members of new families emerged at different time points tend to be new miRNAs, suggesting after arising, novel miRNA genes often experience genome duplication events to forming gene families. However, a large portion of novel human miRNA appeared early in evolution are copies of new families. In contrast, most of miRNA genes appeared relatively late in evolutionary history are not copies of new miRNA families but are single miRNA genes (Table 1). For example, we found that 52.5% of new miRNA genes emerged from the hominidae lineage and 60% of human-specific miRNA genes are not copies of new miRNA families or ancestral families but are single miRNA gene, which is much higher than the average rate of 23.6%. These results suggested that new miRNA genes appeared more recently in evolutionary history have not yet experienced or less experienced GDEs, and the contribution of gene duplication for the formation of new miRNA genes is greater in early stage of evolution than in late stage. The generation of new miRNA genes appeared more recently in evolutionary history is likely due to random hairpin structures which accumulate nucleotide changes to become miRNA genes which are consistent with previous results obtained by Hertel et al. [5]. Hertel et al. [5] found that most of miRNA families in vertebrate genomes can not be traced back to an bilaterian, and concluded that miRNA can arise as de novo genes [5].

3.2. Sequence characteristics of different lineage-specific human miRNAs

The sequence characteristics of pre-miRNAs and mature miRNAs in plants and animals have been reported in recent studies [7,30,31]. To further address whether the sequence characteristics of pre-miRNAs and mature miRNAs differ among miRNA genes arising from different evolutionary lineages, we systemically examined the base composition of different human miRNAs (Fig. 2). Our analysis revealed that nucleotide divergence has occurred on pre-miRNA and mature miRNA sequences, and base composition of human miRNAs that emerged from different evolutionary lineages has different patterns. Our study further showed that base composition is related to the evolutionary age of miRNAs. The frequency of A and U residues in new pre-miRNA sequences was found to be gradually decreasing and the frequency of G and C residues was gradually increasing (Fig. 2A). Further analysis of Spearman's rank correlation suggested that there is a correlation between evolutionary age of miRNAs and base composition in pre-miRNAs sequences (Spearman's rank correlation, A: $Rs = -0.45$; U: $Rs = -0.9667$; G: $Rs = 0.833$; C: $Rs = 0.867$). The older human miRNA genes in evolutionary history often prefer to use A and U, while the younger tend toward higher GC content (Fig. 2B). The tendency of base composition in different mature miRNAs is similar to pre-miRNAs (Fig. 2C and D). Patterns of base composition of human miRNAs that emerged from different evolutionary time points may be generated by chance. In order to take this effect into account, we performed sampling approach and repeated 1000 times, as described in Supplementary Section S1.1. A summary of 1000 random sampling analysis is presented in Supplementary Figs. S1 and S2. The results of random sampling analysis indicated that there is a significantly correlation between evolutionary age of miRNAs and base composition in pre-miRNAs and mature sequences. Zhou et al. [31] reported that animal pre-miRNA sequences tend to contain more A+U nucleotides than

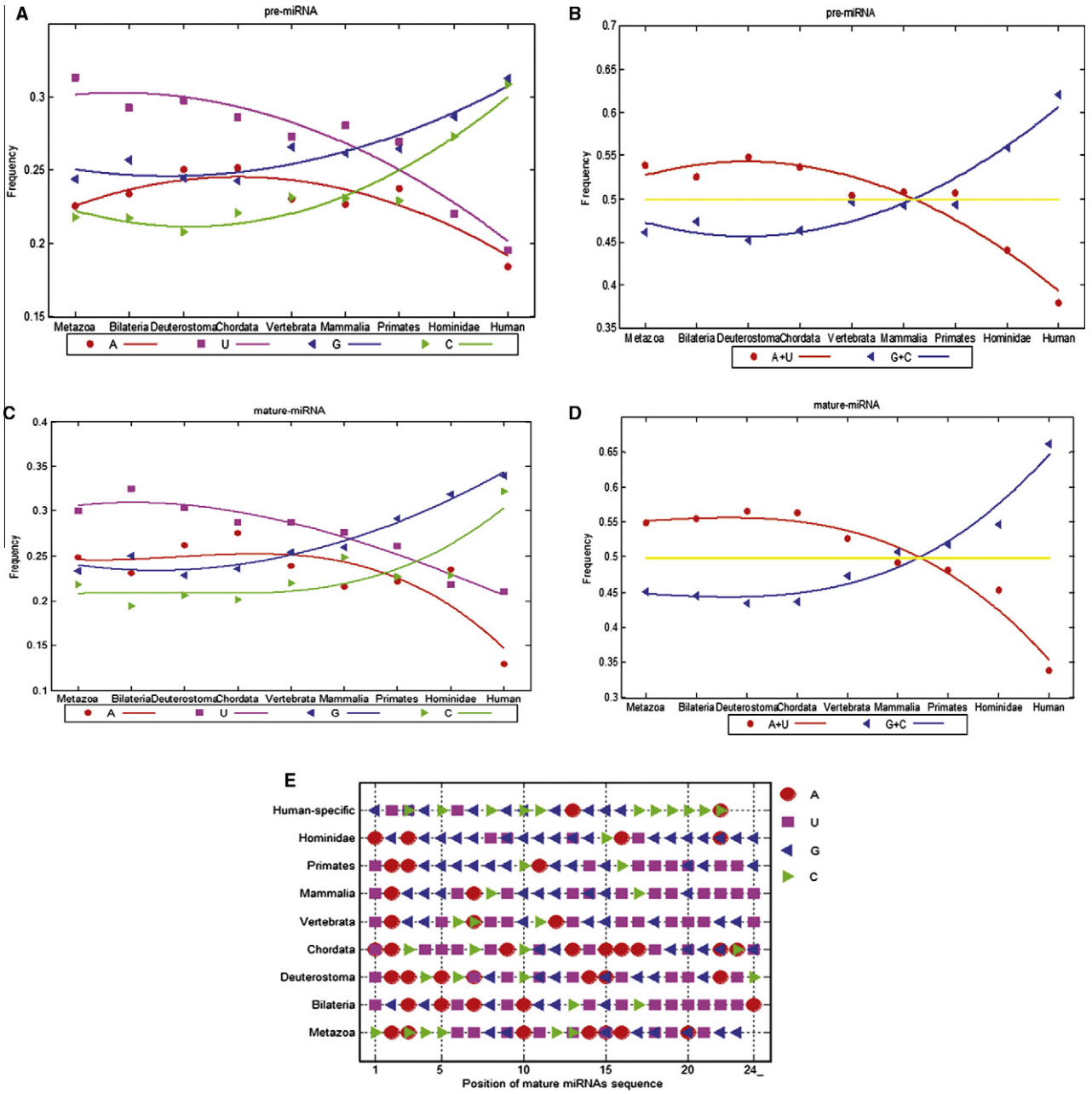

Fig. 2. Base composition divergence of human miRNAs emerged at different evolutionary time points. (A) Base composition of human pre-miRNAs. (B) GC content of human pre-miRNAs. (C) Base composition of human mature miRNAs. (D) GC content of human mature miRNAs. (E) Position-specific predominant nucleotide in different lineage-specific human mature miRNAs. The nucleotide used most frequently is defined as predominant nucleotide.

G+C nucleotides and such a pattern of base composition may be associated with the biosynthesis of mature miRNA. Pre-miRNA secondary structure containing a higher percentage of A+U nucleotides is less stable and more easily processed into mature miRNA by the RNA-induced silencing complex (RISC) [32]. We found that the older human miRNA genes in evolutionary history often prefer to use A and U, while the younger tend to use G and C. Meantime, newly human miRNA genes appeared relatively late in evolutionary history are likely to have been derived from random hairpin structures. These results seem to imply that the random hairpin sequences containing more A+U in genome are easier to evolve into new miRNA genes.

Liu et al. [9] suggested that as antisense regulators, alteration to the sequence of miRNAs will have an impact on their targeting capabilities. So, we investigated base composition at each position in the mature miRNA sequences arising from different evolutionary lineages and the predominant nucleotides at each position are summarized in Fig. 2E. As shown in Fig. 2E, the frequency of the four nucleotides (A, C, G and U) differed at each position. Previous studies have shown that animal mature miRNAs use U residues most frequently at their 5′ end [33,34]. Our analysis of sequence characteristics at each position suggested that this trend is obvious for most of human mature miRNA sequences. As shown in Fig. 2E, mature miRNA in six age groups (bilateria, deuterostoma, chordate, vertebrata, mammalian, and primates) prefer to use U residues most frequently at their 5′ end. However, miRNAs arose from metazoan and hominidae lineages prefer to use C and A residues most frequently at the 5′ end, and human-specific miRNAs tend to use G more often. The frequency of G residues in positions 2–8 was found to be gradually increasing and the frequency of the

other three residues (G, A, U) in positions 2–8 seed region was gradually decreasing, suggesting that miRNAs that appeared relatively late in evolutionary history use G residues with higher frequency between positions 2 and 8. For example, as shown in Fig. 2E, G residues was found to be predominant nucleotide in five sites and six sites in seed regions of mature sequences of miRNAs arose from primates and hominidae lineages. Interestingly, A residues was not found to be predominant nucleotide in any sites between positions 2 and 8 for human-specific mature miRNAs sequences. However, we also found that U residues more frequently at other positions after the seed region except for hominidae and human-specific. Zhou et al. [31] argued that in animals C residues were used less preferentially than the other three nucleotides at position 19, which also found in different human miRNAs. Therefore, it is possible that base composition play a certain role in origination and function of miRNA during evolution.

3.3. Genomic organization of human miRNAs that emerged at different evolutionary time points

Human miRNA genes are scattered in all chromosomes except for Y chromosome. The distribution of human miRNAs on chromosomes is non-random which potentially provides some indication of the role of human miRNA genes in specific diseases [35,36]. To explore the genomic organization of human miRNA genes arising from different evolutionary lineages, we therefore analyzed the chromosomal location of each human miRNA gene and performed genomic enrichment analysis based on the cumulative hypergeometric test. Our results showed divergence of chromosome distribution between different lineage-specific human miRNA genes (Fig. 3A and Table 2). Human miRNA genes that emerged early in evolutionary history tend to be preferentially encoded on specific chromosomes, such as miRNAs originated from bilateria enriched

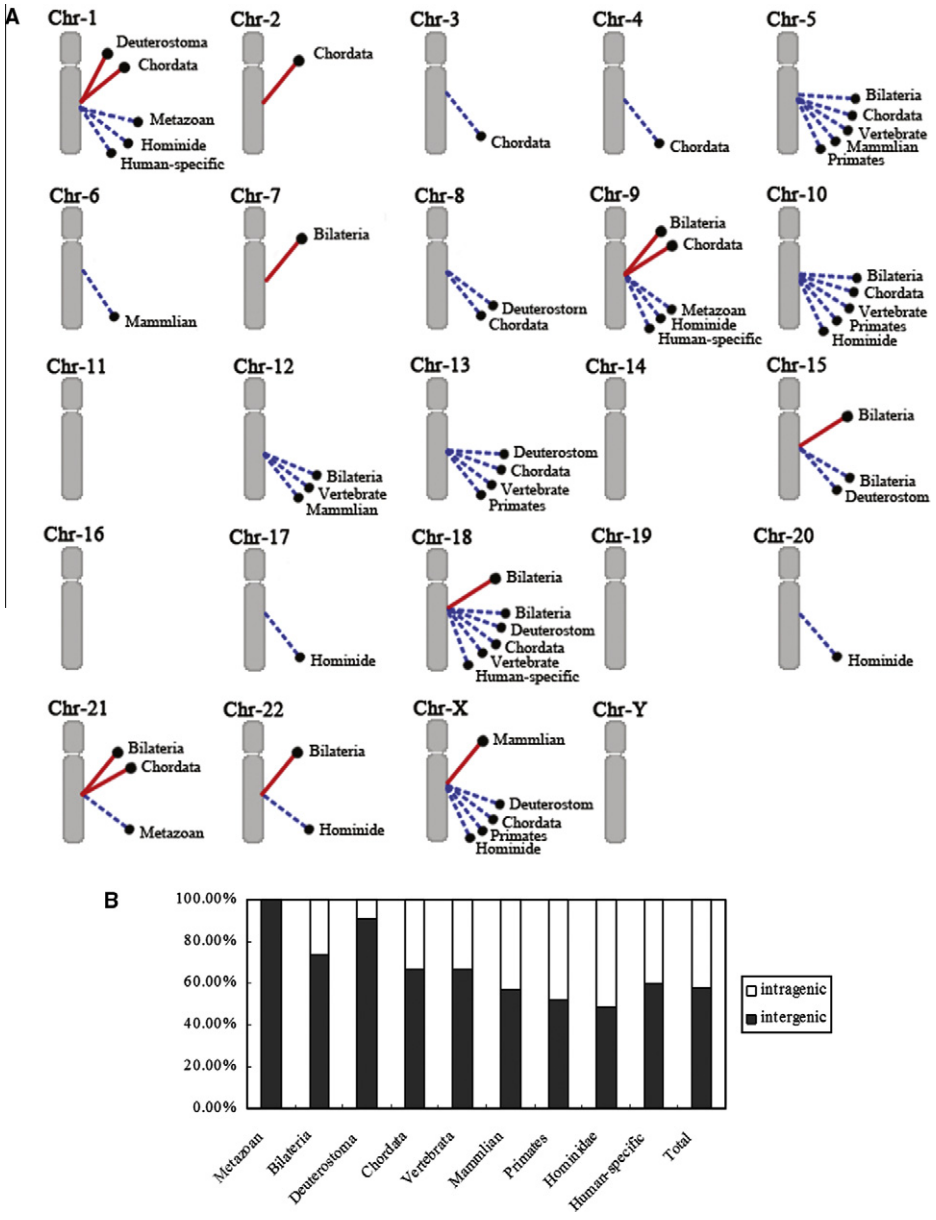

Fig. 3. Genomic organizations of different miRNAs and their targets in human. (A) Chromosomal preference for different human miRNAs emerged at different evolutionary time points and their target genes. Red line represents miRNA genes and blue line represents target genes. (B) Genomic distribution of miRNAs located in intergenic or intragenic regions. The genome coordinates of human miRNAs were determined from the miRBase Sequence Database. Protein-coding genes and introns were based on the refGene and knownGene tables in UCSC Genome Bioinformatics.

Table 2
Genomic organization of human miRNA genes arising from different evolutionary lineages and their targets.

	Metazoan	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalian	Primates	Hominidae	Human-specific
miRNAs[a]	NO	7,9,15,18,21,22	1	1,2,9,21	NO	X	NO	NO	NO
Target genes[b]	1,9,21	5,10,12,15,18	8,13,15,18,X	3,4,5,8,10,13,18,X	5,10,12,13,18	5,6,12	5,10,13,X	1,9,10,17,20,22,X	1,9,18

[a] The chromosomes on which miRNA genes arising from different evolutionary lineages tend to be preferentially encoded.
[b] The chromosomes on which target genes regulated by miRNA genes arising from different evolutionary lineages tend to be preferentially encoded.

on chromosome 7, 9, 15, 18, 21 and 22, while miRNAs originated from primates, hominidae and human-specific miRNAs were not observed to have miRNA enrichment phenomenon, suggesting that miRNA genes appeared more recently in evolutionary history have a more balanced distribution across different chromosomes. Also, miRNA genes and their target genes tend to be preferentially encoded on different chromosomes (Table 2). It has been reported that the density of miRNA genes on the X chromosome is higher than on autosomes in mammals [37]. Here, our analysis also revealed that miRNA genes that originated from the mammalian lineage enriched on chromosome X, supporting the idea that miRNA genes enriched chromosome X possibly contributed to speciation in mammals [37]. A recent study of the evolutionary emergence of miRNA expression in human hES cells showed that miRNA genes located on chromosome 19 were drastically increased in chimpanzees and humans, and argued that the evolution of these miRNAs and their chromosome clustering are primate-specific phenomena [38]. However, miRNA genes that originated from the primates are not observed significantly enriched on chromosome 19 in our study.

It has been reported that the majority of known mammalian miRNAs were found to locate in transcription units [39]. However, a recent study found that most human miRNAs lie between protein-coding genes and one-third are within the introns [40]. Here, miRNA genes were categorized into two types based on their genomic locations: intragenic sets and intergenic sets. Human miRNA genes that reside within the introns or untranslated regions of same-strand protein-coding genes were termed as intragenic sets. On the contrary, miRNA genes in the intergenic regions that did not overlap any same-strand protein-coding genes were termed as intragenic sets. We next studied the distribution characteristics in intragenic or intergenic regions for human miRNAs that emerged at different evolutionary time points (Fig. 3B). In general, 57.72% and 42.28% of human miRNA genes are found to locate in intergenic and intragenic regions respectively. However, comparison of distribution characteristics for different human miRNAs revealed that the frequency of ancient human miRNA genes located in intergenic regions is higher than younger human miRNA genes. Furthermore, more and more miRNA genes are coded in intragenic regions during animal evolution, suggesting that these intragenic miRNA genes may play important roles in animal development and evolution.

A previous study suggested that the target genes of major transcription factors have a strong preference to be encoded on specific chromosomes in yeast [41]. In our study, we also investigated whether the target genes regulated by different lineage-specific human miRNA tend to be preferentially encoded on specific chromosomes through genomic enrichment analysis. The results are summarized in Fig. 3A and Table 2. Our results revealed that miRNAs that emerged at particular evolutionary time points tend to have a strong preference to regulate genes on specific chromosome. Target genes regulated by miRNA genes having closer evolutionary origin were found to be enriched on the same chromosomes. Interestingly, genes located in five chromosomes (chromosome 5, 10, 13, 18, X) tend to be regulated by different lineage-specific human miRNA. However, we also found that genes located in chromosome 2, 7, 11, 14, 16, 19 and Y try to avoid regulating by different human miRNA genes, of which chromosome 11, 14, 16, 19 and Y also were not observed to have miRNA enrichment phenomenon. Our analysis showed that this non-random distribution of different lineage-specific human miRNAs and targets at specific chromosomes may be part of an evolutionary during species divergence and may have important implications for miRNA genes evolution. This non-random distribution potentially also provides some indication of the role of human miRNA genes in specific diseases.

3.4. The relationship between miRNA conservation and disease

A previous study on three species of Drosophila revealed that most young miRNA genes are weakly expressed whereas many older miRNA genes are highly expressed, suggesting that young miRNA genes have smaller effect in Drosophila [10]. In Arabidopsis, young miRNA genes showed weaker regulatory effects [42]. In humans, there is a significant association between a protein's connectivity and conservation [43]. Lu et al. [44] argued that if a gene is evolutionarily conserved, it will be more lethal when gene dysfunction occurs and revealed a significant association between human miRNA conservation and disease by simply grouping human genes into two types: conserved and specific. Here, we performed more detailed analysis to investigate the association between miRNA that emerged at different evolutionary time points and disease phenotype more closely (Fig. 4). Our results showed that highly conserved miRNAs are more likely associated with disease than lowly conserved and human-specific miRNA genes. A large proportion of the human miRNAs that existed early in evolutionary history were related to disease (40–50%), whereas only a small fraction of young human miRNAs are related to disease (5–20%). However, 20% of human-specific miRNA genes are related to disease which is higher than lowly conserved miRNA genes. Our results also suggested that miRNAs which existed early in evolutionary history are associated with a much larger number of diseases (on average 2.2–3.73 diseases per miRNA) than young miRNAs (on average 0.05–0.4 diseases per miRNA). Then we performed sampling approach to evaluate significance of the association between miRNAs and diseases, as described in Supplementary Section S1.2. A summary of 1000 random sampling analysis is presented in Figs. 4B and 4C. As a result, we found a significantly correlation between degree of conservation of miRNAs and the number disease (Spearman's rank correlation, $Rs = -0.883$ and $P < 0.01$) (Fig. 4B), and between degree of conservation of miRNAs and proportion of the miRNAs related to disease in each age group (Spearman's rank correlation, $Rs = -0.817$ and $P < 0.01$) (Fig. 4C).

3.5. Characterization of the protein–protein interaction subnetwork regulated by human miRNAs that emerged at different evolutionary time points

The relationship between the protein–protein interaction network (PPIN) and miRNA regulation has been studied in recent studies [45,46]. In this study, we focused on the roles that different lineage-specific human miRNAs play in the human PPIN. For each set of lineage-specific miRNA genes, we defined a subnetwork

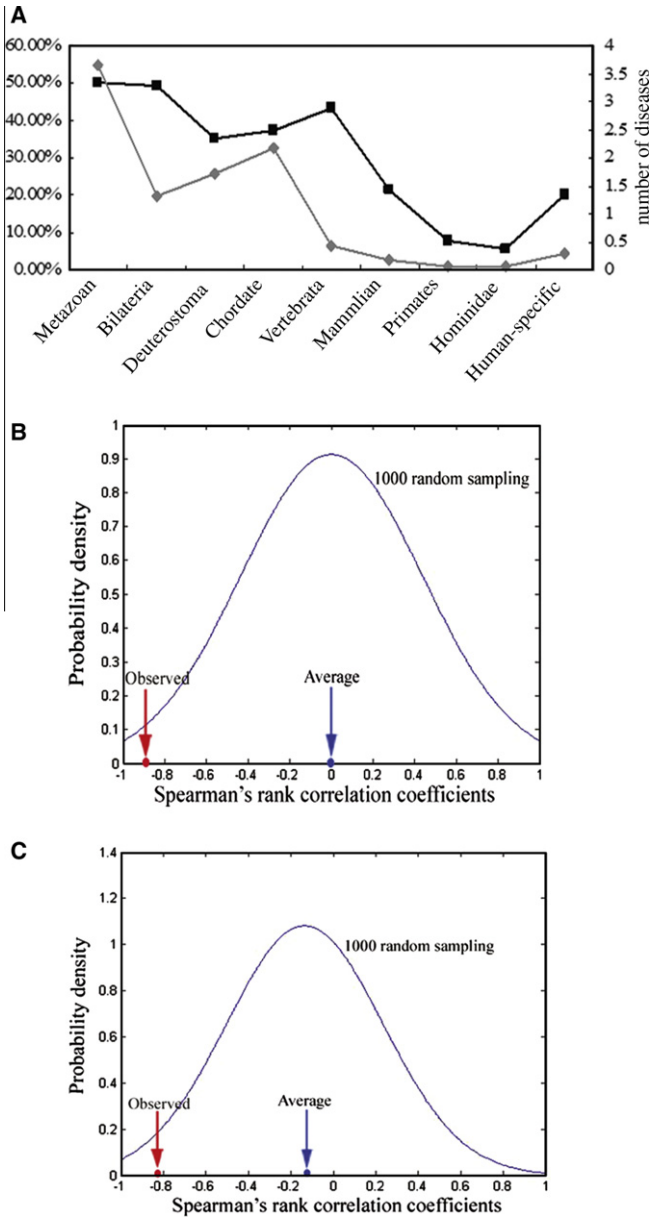

Fig. 4. The association between miRNAs and diseases based on depth of conservation. (A) Each gray pane represents the average number of diseases per miRNA. Each black pane represents the ratio of miRNAs related to at least one disease to total miRNAs for each group of lineage-specific human miRNAs. (B) The distribution of Spearman's rank correlation coefficients between degree of conservation of miRNAs and the number disease in 1000 random sampling analysis. (C) The distribution of Spearman's rank correlation coefficients between conservation of miRNAs and proportion of the miRNAs related to disease in 1000 random sampling analysis. The blue curve represents the distribution of Spearman's rank correlation coefficients in 1000 random sampling analysis. A red arrow represents the real Spearman's rank correlation coefficient.

consisting of all of the target proteins regulated by lineage-specific miRNA genes and all of the interacting proteins with target proteins. We investigated the characteristics of the PPIN regulated by human miRNAs that emerged at different evolutionary time points including: degree, clustering coefficient, betweenness centrality, closeness centrality and characteristic path length. The analysis results are summarized in Table 3. The number of interaction partners for each miRNA target in the network was calculated. Although the target proteins of the miRNAs were shown to interact with more proteins and have higher betweenness centrality than

Table 3
Characteristics of subnetwork formed by different miRNA targets (L0) and the extended subnetwotk containing interacting partners of miRNA targets (L1).

	HPRD	L0 network								L1 network							
		MET	BIL	DEU	CHO	VER	MAM	PRI	HOM	HSP	MET	BIL	DEU	CHO	VER	MAM	PRI	HOM	HSP
Degree	7.78	9.69	9.61	9.81	10.13	9.82	9.84	9.81	11.26	9.75	25.46	16.28	15.12	17.05	15.66	17.11	16.78	15.95	15.32
Betweenness centrality	0.0066	0.0053	0.0088	0.0084	0.0074	0.0085	0.008	0.008	0.0075	0.0143	0.0174	0.0174	0.016	0.0205	0.0174	0.0178	0.0184	0.0124	0.0157
Clustering coefficients	0.1462	0	0.2436	0.3051	0.2198	0.2841	0.2117	0.1940	0.2523	0.4426	0.2606	0.226	0.2205	0.2349	0.2241	0.2285	0.2331	0.2314	0.213
Characteristic path length	5.10	0.09	2.76	3.7	2.57	3.33	2.2	2.69	3.41	3.27	3.56	3.62	3.62	3.66	3.63	3.61	3.62	3.58	3.6
In-degree ratio	N/A	0.0008	0.0103	0.01	0.0106	0.0132	0.0099	0.0063	0.0086	0.0063	0.4021	0.4526	0.4889	0.511	0.4904	0.4373	0.4413	0.4908	0.5032

For each set of lineage-specific miRNA genes, we defined two types of microRNA-regulated PPI network, termed as L0 network and L1 network. Then the topological properties of networks regulated by lineage-specific miRNA genes such as degree, clustering coefficient, betweenness centrality, closeness centrality and characteristic path length was calculated to reveal the roles in which different lineage-specific miRNA genes play in protein interaction network. MET: metazoan, BIL: bilateria, DEU: deuterostoma, CHO: chordata, VER: vertebrata, MAM: mammalian, PRI: primates, HOM: hominidae, HSP: human-specific.

whole networks, which is consistent with a previous report [45]. The number of interaction partners of target genes regulated by different lineage-specific human miRNAs did not show significant difference which implies that target genes regulated by different lineage-specific human miRNAs tend to be hub proteins or bottlenecks. Liang et al. [46] argued that different types of hub proteins have different miRNA targeting propensity, and intermodular hub proteins are more likely to be under miRNA regulation than intramodular hubs. Our analysis showed that target proteins regulated by conserved miRNAs have lower clustering coefficient and betweenness centrality than those regulated by human-specific miRNAs (on average 0.2138 versus 0.4426 and 0.0077 versus 0.0143), suggesting that intermodular hub proteins are more likely to be regulated by conserved miRNAs than human-specific miRNAs. To further explore whether these subnetworks regulated by different human miRNAs form any modules, we defined "in-degree" of a node to represent the number of its within-subnetwork connections, and "in-degree ratio" to represent the ratio of in-degree to total degree of whole subnetwork [45]. We found that the in-degree ratio of L0 subnetworks regulated by highly conserved miRNAs except for metazoan is significantly higher than that of L0 subnetworks regulated by lowly conserved and human-specific miRNAs (on average 0.0109 versus 0.0071, $P < 0.01$ from a two-sample t-test), implying that L0 subnetwork formed by target genes which regulated by highly conserved miRNAs have higher modularity. Moreover, different L1 subnetworks exhibit significantly higher in-degree ratio than that of different L0 subnetworks (on average 0.0085 versus 0.4688, $P < 0.0001$ from a two-sample t-test). However, there is no evident different between in-degree ratio of L1 subnetworks regulated by different lineage-specific miRNAs. These results indicate that target genes regulated by miRNAs tend to form network modules together with their interacting neighbors in PPI network. Also, this mechanism is common behavior for different lineage-specific miRNAs. Our analysis offer new perspective to understand possible mechanisms about subnetworks regulated by human miRNAs that emerged at different evolutionary time points.

4. Conclusions

In this study, we report a comprehensive study of the phylogenetic distribution and evolutionary history of the currently know human miRNA genes by a combination of multiple homology search strategies. Then we further performed systematic analysis to overall uncover the divergence of human miRNA genes based on evolutionary emergence. Because of erroneous and incomplete genome assembly, the possibility that some human homologs may not be found by our sequence-based homology search can not be exclude. However, we believe that our study is a good starting point for uncovering genetic and functional divergence of human miRNA genes during evolution and provide more useful knowledge for further exploring origins and evolution of microRNA genes in human genome.

Acknowledgments

This work was supported in part by the National High Tech Development Project of China, the 863 Program (Grant No. 2007AA02Z329) and the National Basic Research Program of China, the 973 Program (Grant No. 2008CB517302).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.11.053.

References

[1] Zhao, Y. and Srivastava, D. (2007) A developmental view of microRNA function. Trends Biochem. Sci. 32, 189–197.
[2] Niwa, R. and Slack, F.J. (2007) The evolution of animal microRNA function. Curr. Opin. Genet. Dev. 17, 145–150.
[3] Axtell, M.J. and Bartel, D.P. (2005) Antiquity of microRNAs and their targets in land plants. Plant Cell 17, 1658–1673.
[4] Zhang, B.H., Pan, X.P., Wang, Q.L., Cobb, G.P. and Anderson, T.A. (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res. 15, 336–360.
[5] Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I.L. and Stadler, P.F. (2006) The expansion of the metazoan microRNA repertoire. BMC Genomics 7, 25.
[6] Sempere, L.F., Cole, C.N., McPeek, M.A. and Peterson, K.J. (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J. Exp. Zool. B: Mol. Dev. Evol. 306, 575–588.
[7] Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P. and Anderson, T.A. (2006) Conservation and divergence of plant microRNA genes. Plant J. 46, 243–259.
[8] Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C. and Peterson, K.J. (2008) MicroRNAs and the advent of vertebrate morphological complexity. Proc. Natl. Acad. Sci. USA 105, 2946–2950.
[9] Liu, N., Okamura, K., Tyler, D.M., Phillips, M.D., Chung, W.J. and Lai, E.C. (2008) The evolution and functional diversification of animal microRNA genes. Cell Res. 18, 985–996.
[10] Lu, J. et al. (2008) The birth and death of microRNA genes in Drosophila. Nat. Genet. 40, 351–355.
[11] Nozawa, M., Miura, S. and Nei, M. (2010) Origins and evolution of microRNA genes in Drosophila species. Genome Biol. Evol., 180–189.
[12] Heimberg, A.M., Cowper-Sal Lari, R., Semon, M., Donoghue, P.C. and Peterson, K.J. (2010) From the Cover: microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. Proc. Natl. Acad. Sci. USA 107, 19379–19383.
[13] Zhang, R., Wang, Y.Q. and Su, B. (2008) Molecular evolution of a primate-specific microRNA family. Mol. Biol. Evol. 25, 1493–1502.
[14] Tanzer, A. and Stadler, P.F. (2004) Molecular evolution of a microRNA cluster. J. Mol. Biol. 339, 327–335.
[15] Ibanez-Ventoso, C., Vora, M. and Driscoll, M. (2008) Sequence relationships among *C. elegans*, *D. melanogaster* and human microRNAs highlight the extensive conservation of microRNAs in biology. PLoS One 3, e2818.
[16] Liang, H. and Li, W.H. (2009) Lowly expressed human microRNA genes evolve rapidly. Mol. Biol. Evol. 26, 1195–1198.
[17] Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. (2008) MiRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158.
[18] Megraw, M., Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2007) MiRGen: a database for the study of animal microRNA genomic organization and function. Nucleic Acids Res. 35, D149–D155.
[19] Hsu, S.D. et al. (2008) MiRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. Nucleic Acids Res. 36, D165–D169.
[20] Sethupathy, P., Megraw, M. and Hatzigeorgiou, A.G. (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. Nat. Meth. 3, 881–886.
[21] Brameier, M. (2010) Genome-wide comparative analysis of microRNAs in three non-human primates. BMC Res. Notes 3, 64.
[22] Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S. and Peterson, K.J. (2009) The deep evolution of metazoan microRNAs. Evol. Dev. 11, 50–68.
[23] Rhead, B. et al. (2010) The UCSC Genome Browser database: update 2010. Nucleic Acids Res 38, D613–D619.
[24] Jiang, Q. et al. (2009) MiR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 37, D98–D104.
[25] Keshava Prasad, T.S. et al. (2009) Human Protein Reference Database – 2009 update. Nucleic Acids Res. 37, D767–D772.
[26] Grimson, A. et al. (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature 455, 1193–1197.
[27] Prochnik, S.E., Rokhsar, D.S. and Aboobaker, A.A. (2007) Evidence for a microRNA expansion in the bilaterian ancestor. Dev. Genes Evol. 217, 73–77.
[28] Bentwich, I. et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. 37, 766–770.
[29] Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H. and Cuppen, E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120, 21–24.
[30] Zhang, B., Pan, X. and Stellwag, E.J. (2008) Identification of soybean microRNAs and their targets. Planta 229, 161–182.
[31] Zhou, M. et al. (2009) In silico detection and characteristics of novel microRNA genes in the *Equus caballus* genome using an integrated ab initio and comparative genomic approach. Genomics 94, 125–131.
[32] Zhang, B.H., Pan, X.P., Cox, S.B., Cobb, G.P. and Anderson, T.A. (2006) Evidence that miRNAs are different from other RNAs. Cell Mol. Life Sci. 63, 246–254.
[33] He, P.A. et al. (2008) Identification and characteristics of microRNAs from Bombyx mori. BMC Genomics 9, 248.
[34] Yue, J., Sheng, Y. and Orwig, K.E. (2008) Identification of novel homologous microRNA genes in the rhesus macaque genome. BMC Genomics 9, 8.
[35] Calin, G.A. and Croce, C.M. (2006) MicroRNAs and chromosomal abnormalities in cancer cells. Oncogene 25, 6202–6210.

[36] Calin, G.A. et al. (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc. Natl. Acad. Sci. USA 101, 2999–3004.
[37] Guo, X., Su, B., Zhou, Z. and Sha, J. (2009) Rapid evolution of mammalian X-linked testis microRNAs. BMC Genomics 10, 97.
[38] Cao, H., Yang, C.S. and Rana, T.M. (2008) Evolutionary emergence of microRNAs in human embryonic stem cells. PLoS One 3, e2820.
[39] Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A. (2004) Identification of mammalian microRNA host genes and transcription units. Genome Res. 14, 1902–1910.
[40] Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11, 241–247.
[41] Janga, S.C., Collado-Vides, J. and Babu, M.M. (2008) Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. Proc. Natl. Acad. Sci. USA 105, 15761–15766.
[42] Fahlgren, N. et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 2, e219.
[43] Jordan, I.K., Marino-Ramirez, L., Wolf, Y.I. and Koonin, E.V. (2004) Conservation and coevolution in the scale-free human gene coexpression network. Mol. Biol. Evol. 21, 2058–2070.
[44] Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W. and Cui, Q. (2008) An analysis of human microRNA and disease associations. PLoS One 3, e3420.
[45] Hsu, C.W., Juan, H.F. and Huang, H.C. (2008) Characterization of microRNA-regulated protein–protein interaction network. Proteomics 8, 1975–1979.
[46] Liang, H. and Li, W.H. (2007) MicroRNA regulation of human protein protein interaction network. RNA 13, 1402–1408.