microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes

Yirong Wang,[1,2] Junjie Luo,[1] Hong Zhang,[1] and Jian Lu*[,1]

[1]State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

[2]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China

***Corresponding author:** E-mail: luj@pku.edu.cn.

Associate editor: Jianzhi Zhang

Abstract

MicroRNAs (miRNAs) are endogenously expressed small noncoding RNAs. The genomic locations of animal miRNAs are significantly clustered in discrete loci. We found duplication and *de novo* formation were important mechanisms to create miRNA clusters and the clustered miRNAs tend to be evolutionarily conserved. We proposed a "functional co-adaptation" model to explain how clustering helps newly emerged miRNAs survive and develop functions. We presented evidence that abundance of miRNAs in the same clusters were highly correlated and those miRNAs exerted cooperative repressive effects on target genes in human tissues. By transfecting miRNAs into human and fly cells and extensively profiling the transcriptome alteration with deep-sequencing, we further demonstrated the functional co-adaptation between new and old miRNAs in the *miR-17–92* cluster. Our population genomic analysis suggest that positive Darwinian selection might be the driving force underlying the formation and evolution of miRNA clustering. Our model provided novel insights into mechanisms and evolutionary significance of miRNA clustering.

Key words: **miRNA clusters, functional co-adaptation, coordinated regulation, evolution, natural selection, mRNA-Seq, miR-17–92 cluster.**

Article

Introduction

Genetic novelties are the primary sources of new phenotypes (Haldane 1933). The fates of novel genetic elements are usually affected by various evolutionary forces such as selection, drift, and demography (Fay and Wu 2003; Nielsen et al. 2007). The fitness effect of a novel genetic element is also influenced by its epistatic interactions with other genomic factors (Phillips 2008). Among various interactions between genetic novelties and genomic contexts, the origin and evolution of microRNAs (miRNAs) stand out as a paradigm to deepen our understanding of co-evolution between genomic contexts and novel non-coding RNAs.

MiRNAs are a class of endogenously expressed small non-coding RNAs (~22 nt in length) that regulate the expression of target genes at the post-transcriptional level. In animals, a miRNA targets 3′ UTR of target mRNAs by seed (positions 2–8 of the mature miRNA) pairing to cause mRNA degradation and/or translational inhibition (Ambros 2003; Kim and Nam 2006; Bartel 2009; Ghildiyal and Zamore 2009). A single miRNA usually concurrently regulates a large number of target genes, and one gene might be regulated by multiple miRNAs (Enright et al. 2003; Lewis et al. 2003, 2005; John et al. 2004; Rajewsky 2006). It was estimated that ~60% of mammalian protein-coding genes were conserved targets of miRNAs (Friedman et al. 2009). The comprehensive interactions between miRNAs and protein-coding genes often interplay to compose complex genetic networks and fine-tune diverse cellular functions, such as development, differentiation, apoptosis, and metabolism (Hornstein and Shomron 2006). Mutations related to miRNA dysregulation often lead to developmental defects and pathological events (Ambros 2003; Kim and Nam 2006; Bartel 2009; Ghildiyal and Zamore 2009). Therefore, sequences of established miRNAs are usually highly conserved due to functional constraints (Bartel 2004). For example, the mature sequence of *let-7*, one of the first discovered miRNAs, are highly conserved across vertebrate and invertebrate species (Pasquinelli et al. 2000).

The repertoire of animal miRNAs gradually expanded during long-term evolution (Pasquinelli et al. 2000), however, excessive novel and lineage-specific miRNAs have been identified in various taxa (Berezikov et al. 2006; Fahlgren et al. 2007; Zhang et al. 2007; Lu, Fu, et al. 2008; Lu, Shen, et al. 2008; Zhang et al. 2008; Liang and Li 2009; Berezikov et al. 2011; Gangaraju et al. 2011; Lyu et al. 2014; Mohammed, Bortolamiol-Becet, et al. 2014; Mohammed, Siepel, et al. 2014; Fromm et al. 2015). Based on observations in *Drosophila* and other taxa, we along with others proposed a birth and death model of miRNA evolution, which well explained the vast flux of evolutionarily young miRNAs in multiple lineages (Berezikov et al. 2006; Rajagopalan et al. 2006; Lu, Shen, et al. 2008; Lu et al. 2010). Another salient feature is that animal miRNAs are significantly enriched in clusters in discrete genomic regions (Lagos-Quintana et al.

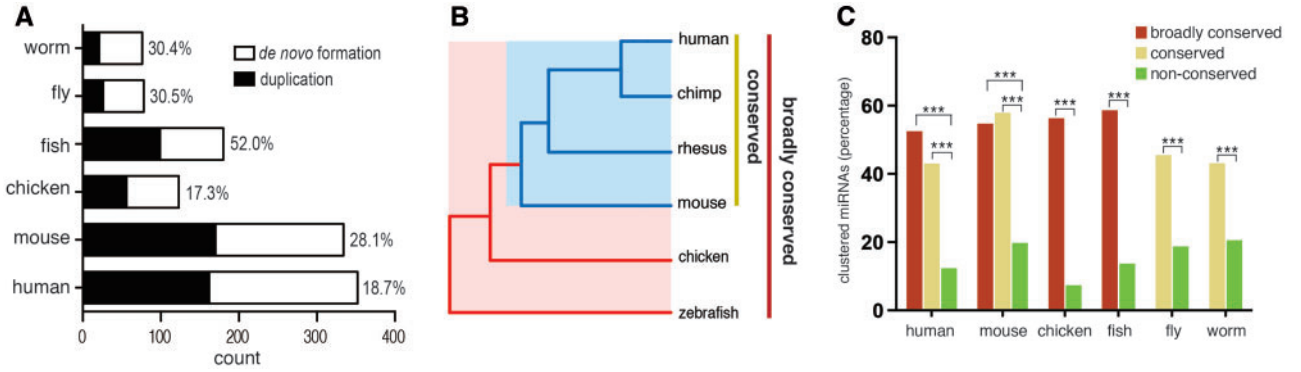

Fig. 1. miRNAs are significantly enriched in clusters in six animal species. (*A*) The number of miRNA precursors that are located in clusters via duplication or *de novo* formation. The percentage of the clustered miRNAs out of the total number of miRNAs annotated in miRBase (V21) is presented beside each bar. (*B*) The classification of broadly conserved, and conserved miRNAs in vertebrates. The information was extracted from whole genome alignments of 100 vertebrate species. (*C*) Percentage of clustered miRNAs in different conservation group. In each species, the (broadly) conserved miRNAs are significantly enriched in clusters. The *y*-axis is the percentage of miRNAs that are located in clusters in that conservation group.

2001; Lau et al. 2001; Lai et al. 2003; Altuvia et al. 2005; Ruby et al. 2007; Marco et al. 2013; Mohammed, Siepel, et al. 2014). The clustering patterns suggest that miRNAs in the same cluster might be transcribed in a polycistronic manner (Baskerville and Bartel 2005; Saini et al. 2007; Ozsolak et al. 2008; Wang et al. 2009; Ryazansky et al. 2011), similar to the operon regulation systems in prokaryotes (Lawrence 1999; Price et al. 2005). As genes located in the same operon often have relevant functions (Jacob et al. 1960), miRNAs in the same cluster were hypothesized to regulate functionally related genes (Ventura et al. 2008; Kim et al. 2009; Yuan et al. 2009; Wang et al. 2011).

The evolutionary principles and functional importance of miRNA clustering are still open questions. In this study, we found duplication and *de novo* formation were important mechanisms to create miRNA clusters and the clustered miRNAs tend to be evolutionarily conserved. We proposed a "functional co-adaptation" model to explain how clustering helps new miRNAs survive and develop functions related to other members of that cluster. We tested our hypothesis by transfecting miRNAs of the *miR-17–92* cluster into human and fly cells and extensively profiling the transcriptome alteration with deep-sequencing. We presented experimental evidence to support the functional co-adaptations between new and old miRNAs in the *miR-17–92* cluster.

Results

miRNAs Are Significantly Enriched in Clusters Via Duplication or *De Novo* Formation

Previous studies have revealed that miRNAs tend to be clustered in introns or intergenic regions (Lagos-Quintana et al. 2001; Lau et al. 2001; Lai et al. 2003; Altuvia et al. 2005;Ruby et al. 2007; Marco et al. 2013; Mohammed, Siepel, et al. 2014). Since the characterizations and annotations of miRNAs have been greatly expanded after the original studies, herein we revisited the clustering patterns of miRNAs with the updated information. We conducted analysis on miRNAs from human, mouse, chicken, zebrafish, fly, and worm, which had high-quality genome assemblies and extensive miRNA expression and target prediction results. In each species, we grouped the miRNA genes into distinct clusters following the procedures described in previous studies (Altuvia et al. 2005; Griffiths-Jones et al. 2008; Marco et al. 2013). Specifically, clustering of miRNA genomic locations is determined if two neighboring miRNAs are located within 10 kb and are in the same strand. The proportion of clustered miRNAs varied across species: ~50% of the miRNAs were clustered in zebrafish and 17%–30% of the miRNAs were clustered in the other five species (fig. 1A). For example, among all the 1,881 miRNAs annotated in human, we identified 352 miRNA genes that were grouped into 99 distinct clusters, including 22 homo-seed clusters (miRNAs having identical "seed" sequences, e.g., the *miR-29c~29b-2* cluster), 62 hetero-seed clusters (miRNAs having distinct "seed" sequences, e.g., the *miR-23b~27b~24* cluster), and 15 homo-hetero-seed clusters (a combination of the former two classes, supplementary table S1, Supplementary Material online). By randomly permuting genomic locations of the miRNAs, in each species we found the observed number of clustered miRNAs was significantly higher than that under randomness ($P < 0.0001$ for all the cases, supplementary fig. S1A, Supplementary Material online). And these results are well congruent with previous studies (Lagos-Quintana et al. 2001; Lau et al. 2001; Lai et al. 2003; Altuvia et al. 2005; Ruby et al. 2007; Marco et al. 2013; Mohammed, Siepel, et al. 2014). For most miRNA clusters, the size ranged from 2 to 6 miRNA precursors. However, unusually large miRNA clusters were also observed in certain species (supplementary fig. S1*B*, Supplementary Material online).

Similar to protein-coding genes, the origin of young miRNA genes is usually achieved by duplication (Kim and Nam 2006; Bartel 2009; Marco et al. 2013) or *de novo* formation (Lu, Shen, et al. 2008; Chen et al. 2013; Long et al. 2013; Marco et al. 2013; Meunier et al. 2013). Here, we pursued the mechanisms by which the clustering patterns were shaped during evolution. We searched for homologous sequences in the miRNA precursors to identify the recent duplications (BLAST *E*-value <0.0001). For the ancient miRNA duplication

events, we delimitated the paralogous copies to have homologous mature sequences (with fewer than five mismatches) and identical seeds. After combining results from the two approaches, we found that in vertebrate species ~50% of the clustered miRNAs were caused by tandem or nonlocal duplications, while in fly and worm, ~30% of the clustered miRNAs were duplicated (fig. 1A). Taken together, duplication of pre-existing miRNAs contributed significantly to the formation of miRNA clusters. Interestingly, the unusually large miRNA clusters in vertebrates were generally shaped by duplications followed by sequence divergence, such as the human *c19mc* cluster (supplementary fig. S2*A* and *B*, Supplementary Material online) and the mouse *Sfmbt2* cluster (supplementary fig. S2C, Supplementary Material online). An intriguing observation is the *miR-430* cluster in zebrafish, which plays important roles in maternal mRNA clearance and translational regulation during development (Giraldez et al. 2006; Bartel 2009; Bazzini et al. 2012). The *miR-430* cluster formed by tandem duplications and all the 58 paralogous miRNAs preserved identical seed sequences (supplementary fig. S2*D*, Supplementary Material online).

About half of the clustered miRNAs do not share sequence similarity with other miRNAs. As hairpin structures are easy to form during RNA transcription, *de novo* formation is the most parsimonious mechanism for these miRNAs (fig. 1A). The homo-seed clusters are usually shaped by duplications; the hetero-seed clusters are mainly shaped by *de novo* formation of new miRNAs; and the hetero–homo-seed clusters are combinations of these two mechanisms (although a few cases might be caused by genome rearrangement). In summary, both duplications and *de novo* formation are important mechanisms to generate miRNA clusters.

Evolutionarily Conserved miRNAs Are Significantly Enriched in Clusters

Next, we investigated the conservation patterns of the miRNA clusters. We employed the genome alignments of 100 vertebrate species to identify conserved miRNAs with similar criteria used in previous studies (Friedman et al. 2009; Agarwal et al. 2015): A "broadly conserved" miRNA should exist in most vertebrates, while a "conserved" miRNA is only conserved in mammals (fig. 1*B*, see Materials and Methods for details). We employed whole genome alignments of 12 *Drosophila* species to identify conserved miRNAs by requiring a miRNA to be present in both *Drosophila melanogaster* and *Drosophila pseudoobscura* (or beyond, supplementary fig. S2*E*, Supplementary Material online). In worms, we required a conserved miRNA to be identified in both *Caenorhabditis elegance* and *Caenorhabditis briggsae* (supplementary fig. S2*F*, Supplementary Material online). For all the (broadly) conserved miRNAs, we further required the orthologous precursor sequences to form stable hairpin structures and the seeds to be identical across the examined clades (Materials and Methods).

An interesting observation is that the (broadly) conserved miRNAs are significantly enriched in clusters (fig. 1C). For the 319 human miRNA precursors that encode conserved miRNAs (219 broadly conserved in vertebrates and 100 conserved only in mammals), 158 (49.5%) of them are located in miRNA clusters. Whereas among the 1,562 human miRNAs that are not evolutionarily conserved, only 194 (12.4%) of them are located in miRNA clusters ($P = 1.46 \times 10^{-53}$, χ^2 test). Similar patterns were observed in mouse, chicken and zebrafish ($P < 1.0 \times 10^{-10}$ for all three species, χ^2 tests, fig. 1C). In *Drosophila*, 45.5% (51 of 112) of the conserved miRNAs are located in clusters, while only 18.8% (27 out of 144) of the nonconserved miRNAs are located in clusters ($P = 7.39 \times 10^{-6}$, χ^2 test, fig. 1C). An analogous pattern was observed in worms as well (fig. 1C).

One should note that the above analysis was based on all the miRNAs annotated in miRBase (V21) (Griffiths-Jones et al. 2006, 2008; Kozomara and Griffiths-Jones 2011, 2014) and many of the curated miRNAs might not be *bona fide* (Fromm et al. 2015). To address this concern, we extensively compiled the small RNAs sequenced with Argonaute (AGO) IP-Seq in human and fly from previous studies (supplementary tables S2 and S3, Supplementary Material online, Materials and Methods). We were able to verify 1,157 miRNA precursors (1,542 mature miRNAs) in 40 publicly available AGO IP-Seq small RNA libraries in human (supplementary table S2, Supplementary Material online) and 213 miRNA precursors (335 mature miRNAs) in 14 publicly available AGO IP-Seq small RNA libraries in fly (supplementary table S3, Supplementary Material online). With these high-confidence miRNA annotations, our conclusion that conserved miRNAs tend to be clustered remains intact (supplementary fig. S3*A*, Supplementary Material online). Moreover, when we searched for novel miRNAs with miRDeep2 program (Mackowiak 2011) in all the AGO IP-Seq small RNA libraries used in this study (see Materials and Methods for details), we also identified 12 novel miRNA candidates in human (supplementary table S4 and fig. S4, Supplementary Material online) and four in *Drosophila* (supplementary table S5 and fig. S5, Supplementary Material online). The human novel miRNAs are generally nonconserved and lowly expressed. Notably, we found seven out of the 12 novel miRNAs we detected in human were clustered to known miRNAs (supplementary table S4, Supplementary Material online). Two out of the four novel miRNAs in fly were caused by recent duplications (supplementary fig. S5, Supplementary Material online) and none of the novel miRNA candidates in *Drosophila* was clustered to known miRNAs (supplementary table S5, Supplementary Material online). After incorporating these novel miRNAs into analysis, our observations that evolutionarily conserved miRNAs are significantly enriched in clusters remain intact (supplementary fig. S3*B*, Supplementary Material online).

Evolutionary Mechanisms for Clustering of *De Novo* Formed miRNAs: Functional Co-adaptation Versus Selection Interference

Thus far our results revealed that miRNAs tend to be clustered, especially for the evolutionarily conserved ones. Several hypotheses have been proposed to account for the clustering pattern of miRNAs, such as tandem

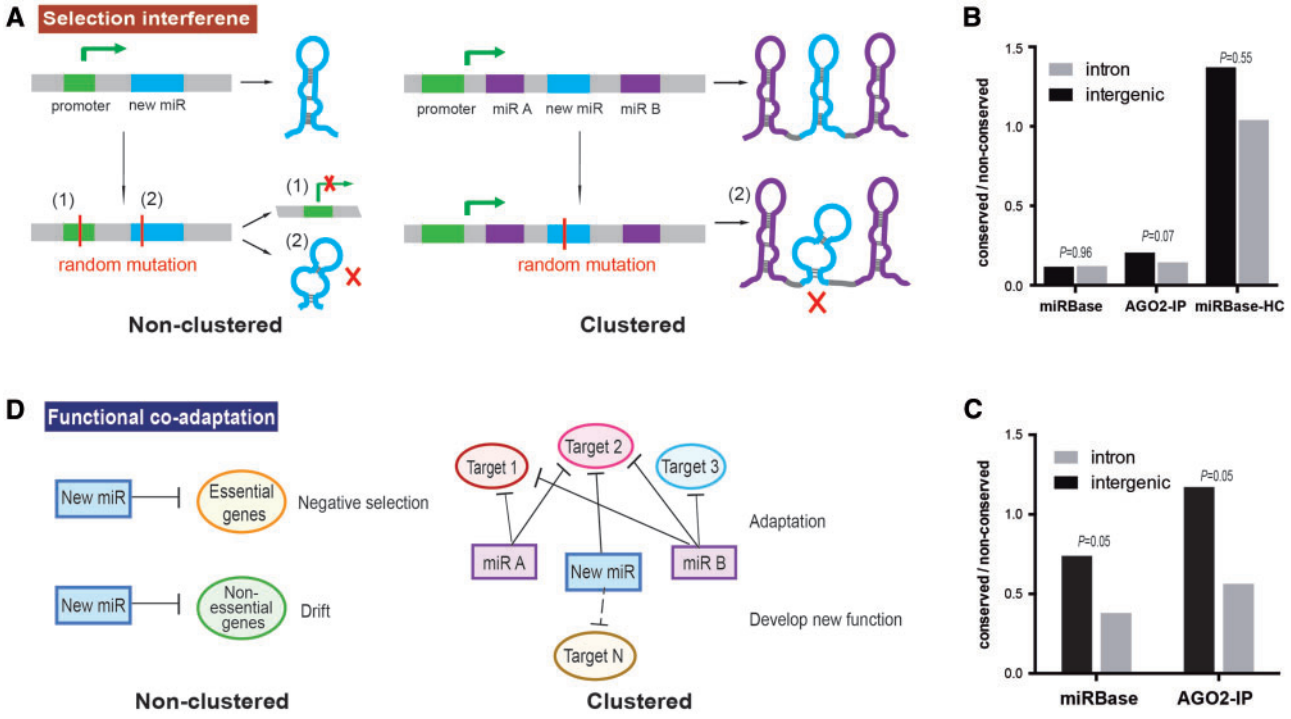

Fig. 2. Evolutionary mechanisms underlying miRNA clustering. (*A*) Selection interference (or drift-draft) model: If a *de novo* formed miRNA located in a discrete locus (nonclustered), it has a high probability to degenerate, either by mutations that abolish its transcription or by mutations that destroy the hairpin structure. If a *de novo* formed miRNA originates in a miRNA cluster, it will be sheltered by the pre-existing miRNAs since mutations that abolish the transcription of this miRNA cluster will be selected against. Therefore, new miRNAs in a cluster will have a higher chance to survive relative to the counterpart in a discrete locus. (*B*) The ratio of conserved/nonconserved nonclustered miRNAs in introns versus intergenic regions of human genome. The patterns are consistent for all the human miRNAs annotated in miRBase (V21), the miRNAs with evidence of AGO2-IP Seq, or the "high-confidence" miRNAs defined with miRBase (V21). "HC" stands for high-confidence. (*C*) The ratio of conserved/nonconserved nonclustered miRNAs in introns versus intergenic regions of *D. melanogaster* genome. The patterns are consistent with all the *D. melanogaster* miRNAs in miRBase (V21) and the miRNAs with evidence of AGO-IP Seq. (*D*) Functional co-adaptation model: For a *de novo* formed miRNA in a discrete locus, it might be selected against or drift since most animal novel miRNAs are not adaptive. On the other hand, the new miRNA in a cluster will be co-expressed with the pre-existing miRNAs spatially or temporally so that it might gradually develop function to regulate target genes of these pre-existing miRNAs. Thus clustering helps the new miRNA to develop function and the cluster is stabilized by natural selection once function of this cluster is fully established.

duplication, split or rearrangements of miRNA loci (Marco et al. 2013). Based on the observation that *Drosophila* new miRNAs often arose around the pre-existing ones to form clusters, Marco et al. (2013) proposed a "drift-draft" model by which the motifs of the pre-existing miRNAs would protect new miRNAs to be transcribed and processed properly since those motifs were already interacting with the miRNA processing machinery. Under such a model, the *de novo* formed new miRNAs are sheltered by the established ones in the same cluster because mutations that abolish the transcription or processing of the new miRNA will affect the pre-existing ones as well and are hence selected against (fig. 2*A*). On the other hand, if a *de novo* formed miRNA is located in a discrete locus, it will have a higher probability to degenerate, either by mutations abolishing its transcription or by mutations impairing its processing. Therefore, the linkage of a novel miRNA to other pre-existing miRNAs (or protein-coding genes) might shelter the new miRNA from degeneration in the initial stages; and later on a subset of these new miRNAs might develop function and becomes conserved across species. It is challenging to directly evaluate the protective effects of the pre-existing miRNAs on the newly emerged ones at this moment. However, if linkage attenuates the loss of a new miRNA by ensuing its transcription, we expect to observe a higher fraction of the miRNAs in introns of protein-coding genes to be conserved (and hence functional) since most miRNAs in introns are co-transcribed with the host genes (Rodriguez et al. 2004; Ruby et al. 2007). To test this hypothesis, we compared the conservation patterns of the nonclustered human miRNAs in introns versus those in the intergenic regions. Unfortunately, we did not find any evidence that the nonclustered miRNAs in introns were significantly more conserved comparing with their counterparts in intergenic regions (fig. 2*B*). Similar results were obtained if we used other miRNA filtering criteria to define the conserved miRNAs in vertebrates. And similar pattern were found in *Drosophila* (fig. 2*C*). Taken together, although the "drift-draft" model reasons that clustering (or genomic linkage) helps newly *de novo* formed miRNAs to develop function by

attenuating its loss, we were not able to find evidence to support this elegant model at this moment.

Furthermore, the majority of the *de novo* formed novel miRNAs are evolutionarily transient and will degenerate even after they are fixed in the populations if they are not maintained by functional constraints (Berezikov et al. 2006; Lu, Shen, et al. 2008), thus developing functions related to the pre-existing miRNAs will help the novel miRNAs to survive and stabilize. Herein, we proposed a "functional co-adaptation" model to account for the evolution and function of *de novo* formed new miRNAs in the clusters (fig. 2*D*). For the new miRNAs that eventually become functional, many adaptive changes might be needed to adapt to the genomic contexts (Lu, Fu, et al. 2008; Mohammed, Bortolamiol-Becet, et al. 2014). Since miRNAs in the same clusters are usually co-transcribed temporally or spatially (see below for details), the newly formed miRNAs might gradually develop functions to target genes that are related to the pre-existing miRNAs in the same cluster; or multiple *de novo* formed new miRNAs in the same cluster interplay to regulate overlapping sets of target genes. Therefore, although miRNAs in the same cluster have independent origins, they might regulate overlapping sets of target genes through convergent evolution. Thereafter the clustering patterns of miRNAs and the modular regulation of target genes will be stabilized by natural selection during long-term evolution. In the following sections we presented lines of evidence to support the "functional co-adaptation" model.

miRNAs in the Same Clusters Tend to Be Co-expressed and Regulate Overlapping Sets of Target Genes

To test the "functional co-adaptation" model, we first examined whether miRNAs in the same clusters were co-expressed among various tissues. Previous studies demonstrated that a miRNA cluster was usually transcribed as a single unit (Baskerville and Bartel 2005; Saini et al. 2007; Ozsolak et al. 2008; Wang et al. 2009; Ryazansky et al. 2011), however, mature miRNAs in the same cluster might not be highly correlated due to regulation in the maturation processes (Baskerville and Bartel 2005; Ryazansky et al. 2011). To examine whether the abundance of mature miRNAs from the same clusters tend to be correlated systematically, we conducted Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath 2008; Zhao et al. 2010) on 683 mature miRNAs that were highly expressed and unambiguously mapped on human genomes in 123 non-redundant small RNA libraries from different human tissues and cell lines (supplementary table S6, Supplementary Material online; the analysis was fully described in Materials and Methods). Many miRNAs are highly tissue specific (Lagos-Quintana et al. 2002; Guo et al. 2014; Mohammed, Bortolamiol-Becet, et al. 2014), therefore, the expression modulation analysis would effectively capture the signals of miRNA co-expression. The unsupervised WGCNA revealed the abundances of mature miRNAs were clustered into 17 distinct co-expression modules (fig. 3*A*). Strikingly, for the 47 miRNA clusters that are expressed above thresholds in the examined libraries, 39 of them have all the encoded miRNAs classified into the same expression modules (fig. 3*B*). Alternatively, the observed number of clustered miRNAs that are co-expressed in a module is significantly higher than the results obtained by randomly shuffling the miRNA clustering assignments ($P < 0.001$, fig. 3*C*). Overall, our WGCNA results demonstrated that abundances of mature miRNAs in the same clusters are significantly highly correlated.

Next we investigated whether miRNAs in the same clusters tend to target overlapping sets of genes. We obtained expression profiles of miRNAs and mRNAs from five tissues of human males (brain, cerebellum, heart, kidney, and testis) as determined in previous studies (Brawand et al. 2011; Meunier et al. 2013) (supplementary table S7, Supplementary Material online). We observed considerable variations in expression for both miRNAs (supplementary fig. S6*A*, Supplementary Material online) and mRNAs (supplementary fig. S6*B*, Supplementary Material online) across the five human tissues. We employed TargetScan (Friedman et al. 2009) to predict conserved target genes in mammals (P_{CT} cut-off is 0.5). Moreover, we also required a miRNA and its target genes to be co-expressed in the same tissue.

Our results demonstrated that miRNAs in the same cluster had the tendency to regulate overlapping sets of target genes. In total, 1,751 genes are regulated simultaneously by at least two distinct miRNA families (each family has a distinct seed) from one miRNA cluster; and this number is significantly higher than that obtained under random simulations ($P < 0.001$, fig. 3*D*). This pattern also held true when we examined the genes that were simultaneously targeted by at least three distinct miRNAs in a cluster ($P < 0.001$, fig. 3*E*). In the above analysis, we only focused on the hetero-seed or hetero-homo seed clusters and collapsed miRNAs in a cluster with the same seeds into one distinct miRNA family (the detailed analysis is described in the Materials and Methods). One should also note that in the permutation analysis, we only randomly shuffled the locations of miRNAs so that the length and conservation levels of 3′UTRs of each gene were fully controlled in this permutation procedure. The significant enrichment in overlapping targets among miRNAs in the same clusters strongly supports the "functional co-adaptation" model and cannot be explained by genetic drift related to miRNA regulation.

Genes simultaneously targeted by multiple (≥ 2) miRNA families from the same cluster (TM_C) have significantly lower expression levels than genes targeted by only one co-expressed miRNA (T1) in all the five human tissues we examined ($P < 0.05$ for all the five tissues, fig. 3*F*, TargetScan $P_{CT} > 0.5$ was used to predict target genes); whereas similar but weaker reduction was observed for genes targeted by multiple miRNA families but not from the same cluster (TM_N) ($P < 0.05$ for human brain, heart, kidney and testes, and $P = 0.15$ for cerebellum for TM_N versus T1 genes, fig. 3*F*). Thus, miRNAs from the same cluster have the tendency to regulate the same sets of targets and cooperatively repress expression levels of such genes. In the following we will present experimental evidence for the *miR-17–92* cluster to support this conclusion.

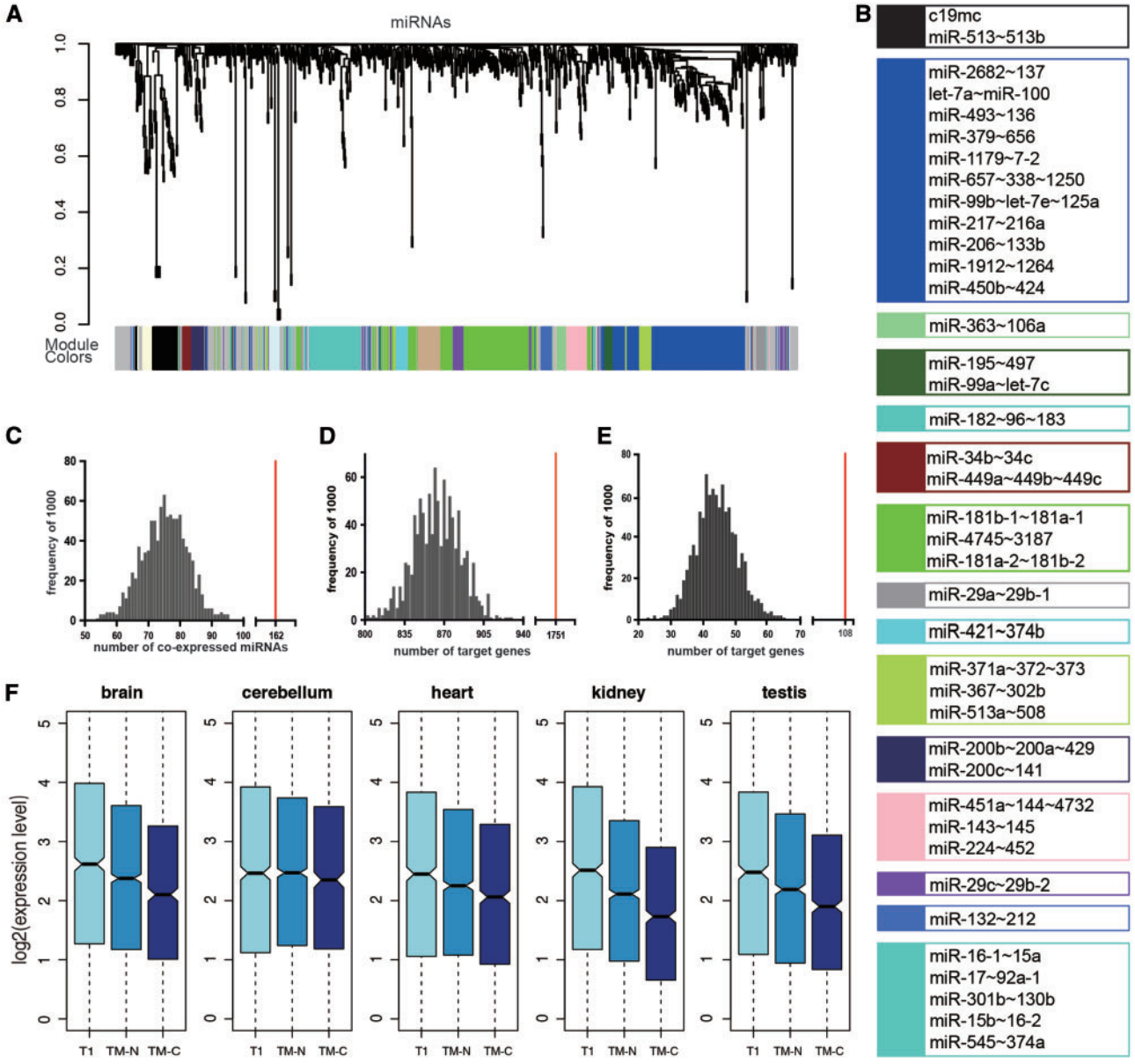

Fig. 3. miRNAs in the same clusters tend to be co-expressed and target overlapping sets of genes. (*A*) Hierarchical cluster tree showing co-expression miRNA modules identified using WGCNA. Modules correspond to miRNAs are labeled by colors. (*B*) List of miRNA clusters identified in co-expressed modules. miRNAs in the same color module are co-expressed and only clustered miRNAs are listed here. (*C*) The observed number of miRNAs that are co-expressed with other members in the same cluster (the red line) versus the numbers obtained by random permutation of miRNA genomic location (the grey histogram, 1,000 replications of permutations were performed). (*D*) Number of target genes regulated by at least two distinct miRNA seeds from the same cluster. (*E*) Number of target genes regulated by at least three distinct miRNA seed from the same cluster. In (D, E), histograms of 1,000 replicates of simulation results are shown in grey, and observed value is in red. (*F*) The expression levels of miRNA targets in five human tissues. (1) T1: genes that are targeted by only one co-expressed miRNA family in a tissue; (2) TM_C: genes that are targeted by at least two distinct conserved seeds from a miRNA cluster; and (3) TM_N: genes that are targeted by at least two distinct conserved seeds but none of them were from the same miRNA cluster (TargetScan $P_{CT} > 0.5$ were employed in the target prediction).

mRNA-Seq Analysis Reveals Members of *miR-17–92* Cluster Regulate Overlapping Sets of Target Genes

To test our hypothesis on the evolution and function of miRNA clustering, we transfected representative members of the *miR-17–92* cluster into human 293FT cells and quantified the transcriptome alteration by each miRNA with mRNA-Seq. The *miR-17–92* cluster is involved in a variety of biological processes (Ventura et al. 2008; Mogilyansky and Rigoutsos 2013) and it is deeply conserved across vertebrates (fig. 4*A* and *B*). This cluster comprises six miRNA precursors (*mir-17*, *18a*, *19a*, *20a*, *19b-1*, and *92a-1*) that encode four distinct seeds (fig. 4*A* and *B*). *miR-17* and *20a* encode the same seed (AAAGUGC) and *miR-19a* and *19b-1* encode the same seed (GUGCAAA). There is one nucleotide difference between the seeds of *miR-17* and *miR-18a* (AAAGUGC and AAGGUGC, respectively) and these two miRNAs are classified into two families; however, these two miRNAs might be caused by ancient duplications since their

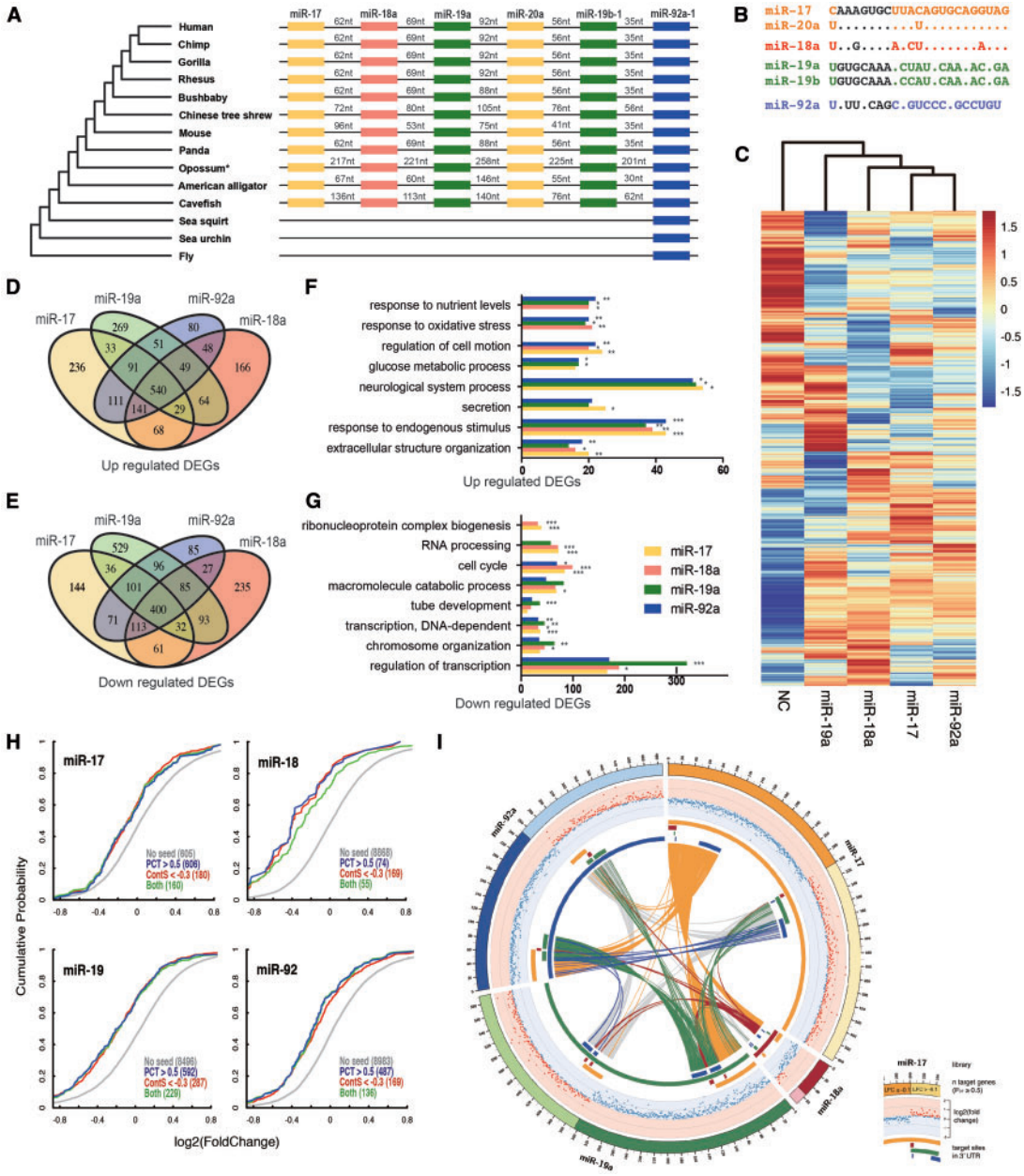

Fig. 4. The cooperative effects of miRNAs in the *miR-17–92* cluster revealed by miRNA transfection and deep-sequencing. (*A*) Evolution of the *miR-17–92* cluster in animals. Sequences and structures of the *miR-17–92* cluster are highly conserved in the vertebrate species. miRNAs with the same seeds are labeled with the same colors. *miR-92a* is the oldest member of this cluster and highly conserved in fly and some urochordate species. Distance between miRNAs in *miR-17–92* cluster in each species is listed. *: *miR-17–92* cluster is located on the "−" strand in opossum and on "+" strand in other species. (*B*) Mature sequence of miRNAs in *miR-17–92* cluster. The seed regions (positions 2–8) are labeled in black. (*C*) Clustering of mRNA expression profiles in cells transfected with NC, *miR-17*, *miR-18a*, *miR-19a*, and *miR-92a*. (*D*) Overlaps of the up-regulated genes in the miRNA transfected cells. (*E*) Overlaps of the down regulated genes in the miRNA transfected experiments. (*F*) Functional enrichment of up regulated genes in each miRNA transfection experiment. (*G*) Functional enrichment of down regulated genes in each miRNA transfection experiments. In (F, G), the *x*-axis denotes the gene count in each biological process. "*" means genes are significantly enriched in this BP (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). The color schemes for each miRNA experiment are the same. (*H*) Cumulative distribution of gene expression changes in miRNA transfected 293FT cells. The *x*-axis is log2(FoldChange) in cells transfected with miRNA versus those transfected with NC. The *y*-axis the cumulative fraction of genes. Blue: target genes with target scan $P_{CT} > 0.5$, Red: Context++ score < -0.3, Green: both, Grey: genes without any target site of the transfected miRNA in the entire mRNA. The number of genes in each category is indicated in parentheses. (*I*) Circos plot of conserved target genes ($P_{CT} > 0.5$) in miRNA transfection experiments. Each sector of the plot corresponds to a miRNA transfected sample. Down regulated target genes with log2(FoldChange) (LFC) < -0.1 are grouped in the darker portion of each sector. Expression level changes of the conserved targets are presented with scatter plots (up regulated genes labeled in red and down regulated in blue). Presence of predicted target genes ($P_{CT} > 0.5$) for the various components of the *miR-17–92* cluster is indicated by tiles plot in the inner circle. Target genes present in more than one library are joined by links. Links are colored if the target is down regulated and LFC < -0.1 in the correspondent library. Otherwise, they are shown in grey.

precursors share significant similarities (supplementary fig. S7, Supplementary Material online). It is notable that the nucleotide compositions are very similar in the seed regions of the four miRNA families (fig. 4B). The target prediction analysis revealed significantly higher number of genes co-targeted by at least two distinct members of the *miR-17–92* cluster than that expected under randomness (444 observed vs.. 220 expected under randomness, $P < 0.001$). Therefore, the *miR-17–92* cluster provides us a good system to explore function relevance between members of clustered miRNAs.

We selected four distinct mature miRNAs in the *miR-17–92* cluster (*miR-17*, *miR-18a*, *miR-19a*, and *miR-92a*) and transfected each miRNA mimic into human 293FT cells (see Materials and Methods). As a negative control, we also transfected the 293FT cells with miRNA Mimic Negative Control (NC) which was known not to directly target transcriptome (Zhao et al. 2013, 2014). We performed directional mRNA-Seqs to quantify the transcriptome of the NC and miRNA transfected cells (we also sequenced the untreated 293FT cells to further confirm the NC transfection does not cause specific target effects, supplementary fig. S8, Supplementary Material online; see Materials and Methods). We detected ~12,000 protein coding genes that were expressed at adequate abundance (RPKM ≥1). After normalization of expression levels, the clustering analysis indicated the transcriptome profiles of the four miRNA transfected cells showed high similarities with each other, while all the four miRNA transfected samples were dissimilar to the NC transfected cells (fig. 4C).

By comparing expression profiles of the miRNA versus NC transfected cells with NOISeq (Tarazona et al. 2012), we identified a large number of differentially expressed genes in each miRNA transfection experiment (fig. 4D and E, NOISeq $q = 0.9$ was used as cutoff). The four miRNAs differed in altering the transcritpomes, with transfection of *miR-19a* causing the largest number of genes differentially expressed, and *miR-17* affected the least mRNA alteration. Hundreds of significantly up- or down-regulated genes are overlapping across the miRNA transfection experiments (fig. 4D and E), and the observed numbers of overlapping genes are significantly higher than those obtained under randomness ($P < 0.001$ for both up- and down-regulated genes, supplementary fig. S9A and B, Supplementary Material online). The differentially expressed genes in each miRNA transfection experiment are significantly enriched in biological pathways such as "transcription", "cell cycle", or "response to endogenous stimulus" (fig. 4F and G). Therefore, the global profiling analysis demonstrated that members in the *miR-17–92* cluster generally exerted similar regulatory effects on transcriptomes despite of the difference in mature miRNA sequences.

In each miRNA transfection experiment, predicted target genes (TargetScan $P_{CT} > 0.5$) of the transfected miRNA are significantly more down-regulated [the median log2(FoldChange) relative to cells transfected with NC is −0.10, −0.38, −0.17, and −0.15 for *miR-17*, *18a*, *19a*, and *92a*, respectively] comparing with genes that does not harbor any target sites in the messengers ($P < 10^{-10}$ for all the four comparisons, Kolmogorov–Smirnov tests, fig. 4H). The down-regulation levels of the target genes in our transfection experiments are well consistent with previous results that the majority of the preferentially conserved target genes are repressed at modest levels (Bartel 2009). TargetScan Context++ Score does not require conservation of target sites but efficiently identifies the optimized genomic contexts of miRNA target (Agarwal et al. 2015). With the Context++ Score cutoff < −0.3, in each miRNA transfection experiment the predicted target genes are also significantly more down-regulated [the median log2(FoldChange) is −0.09, −0.27, −0.18, and −0.18 for *miR-17*, *18a*, *19a*, and *92a*, respectively] comparing with the genes that does not harbor any target sites ($P < 10^{-4}$ for all the four comparisons, Kolmogorov–Smirnov tests, fig. 4H).

To further confirm that miRNAs in the *miR-17–92* cluster directly target overlapping sets of genes, we identified the high-confidence targets by requiring the targets predicted by TargetScan ($P_{CT} > 0.5$) and the expression levels down-regulated with log2(FoldChange) < −0.1 in the transfected cells. With those criteria, we identified 301, 55, 345, and 268 high-confidence target genes for *miR-17*, *18a*, *19a*, and *92a*, respectively (totally 775 high-confidence genes after removing overlapping genes, fig. 4I). Among these 775 high-confidence target genes, 172 were targeted by at least two out of the four miRNAs (fig. 4I), significantly higher than the number obtained by randomness ($P < 0.001$, supplementary table S8, Supplementary Material online). Gene ontology analysis revealed the high-confidence targets of the *miR-17–92* cluster were significantly enriched in the category "GO:0045449, regulation of transcription" (173 genes, $P < 10^{-5}$). Notably, the high-confidence target genes simultaneously regulated by multiple members of the *miR-17–92* cluster were further significantly enriched in the "regulation of transcription" category (51 out of 173, or 29.5% comparing with 22.2% for all the targets, $P = 0.046$, Fisher's exact test). We also used log2(FoldChange) cutoff of −0.2, −0.3, and −0.5 on the evolutionarily conserved target genes in the transfection experiments, and we constantly observed the number of target genes repressed by at least two out of the four miRNAs was significantly higher than the number obtained by permutation tests ($P < 0.001$ for all cases, supplementary table S8, Supplementary Material online). In summary, our transfection and deep-sequencing experiments confirmed that miRNAs in the *miR-17–92* cluster target overlapping sets of target genes, which might reinforce the repression effects of this miRNA cluster. These results are also consistent with previous studies of this miRNA cluster in mouse (Ventura et al. 2008; Han et al. 2015).

Functional Co-adaption During the Evolution of the *miR-17–92* Cluster

The miRNA repertoire keeps increasing as the complexity of gene regulation increases (Fromm et al. 2015). Although the *miR-17–92* cluster is highly conserved among extant vertebrates (Mogilyansky and Rigoutsos 2013), deep phylogenetic analysis revealed *miR-92a* was the initial founding member of this cluster in vertebrates, since only *miR-92a* exists (and is highly conserved) in *Echinodermata* (sea urchin), *Urochordata*

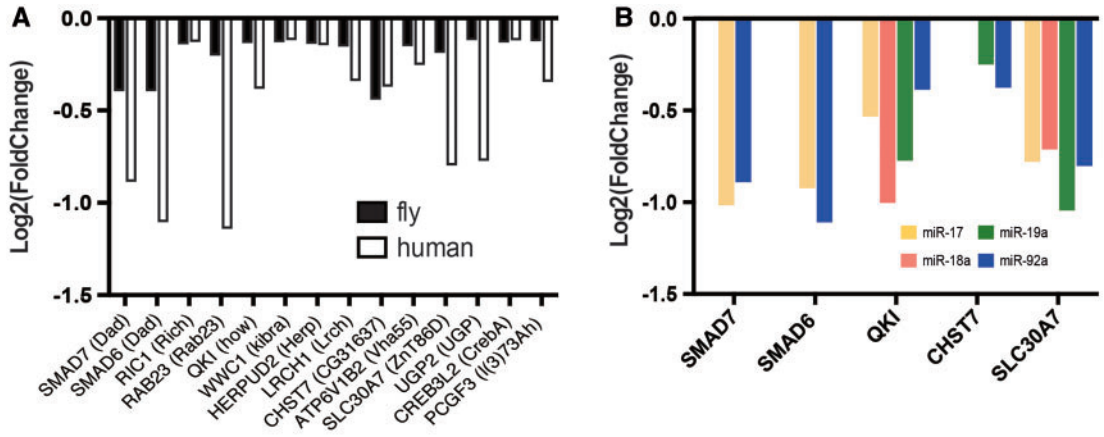

Fig. 5. Convergent evolution of the *miR-17–92* cluster. (*A*) Changes in expression level of conserved target genes of *miR-92a* in human and *Drosophila*. The *y*-axis is LFC in 293FT cells and S2 cells that were transfected with *miR-92a*. The *x*-axis are the gene names for human and fly (in parenthesis). (*B*) Expression level of four conserved target genes of *miR-92a* between human and fly that are also targeted and down regulated by other miRNAs in the *miR-17–92* cluster.

(sea squirt), and *Drosophila* (fig. 4*A*). The other members in the *miR-17–92* cluster, which does not share sequence similarities with *miR-92a*, emerged proximal to *miR-92a* in the common ancestors of vertebrates, possibly by *de novo* formation followed with duplications (fig. 4*A*).

To explore whether the newly derived members of the *miR-17–92* cluster (*miR-17*, *18a*, and *19a*) shared overlapping target genes of the founding member (*miR-92a*) during the ancient formation of this cluster, we first identified the evolutionarily conserved targets of *miR-92a* in both human and fly. We transfected the mature *miR-92a* mimics into the *Drosophila* S2 cells and deep sequenced the transcriptomes in the control and the cells 32-h post-transfection. For the 171 predicted conserved target genes of *miR-92a* that are expressed in S2 cells, they are significantly more down-regulated [the median log2(FoldChange) is −0.06] than genes without any target sites in the entire mRNAs ($P < 0.0001$, Kolmogorov–Smirnov test).

We totally identified 4,923 genes that were expressed in S2 cells and homologous to human genes expressed in 293FT cells. Our miRNA transfection and deep-sequencing experiments further revealed 14 of these genes beared canonical target sites of *miR-92a* and down-regulated by at least 10% in both human and flies after *miR-92a* transfection (fig. 5*A*). An interesting observation is that five of these conserved target genes of *miR-92a* (*SMAD7*, *SMAD6*, *QKI*, *CHST7*, *SLC30A7*) were also targeted and down-regulated by *miR-17*, *miR-18a*, or *miR-19a* (fig. 5*B*). Hence, the evolution and target regulation patterns of the *miR-17–92* cluster well support our functional coadaptation model of miRNA clustering: *miR-92a* was the ancient and founding member of this cluster, later on *miR-17*, *18a*, *19a* originated and regulated overlapping sets of target genes of *miR-92a*. Meanwhile, *miR-17*, *18a*, and *19a* also interacted with the genomic contexts and regulated overlapping sets of target genes (fig. 4*I*). Therefore, the function relatedness of the target genes facilitated novel members of the *miR-17–92* cluster to survive, develop related function, and stabilize as a regulatory module thereafter.

Discussion

The coordinated regulation of miRNAs in the same clusters can be manifested in different aspects. By examining expression levels of genes targeted by multiple miRNAs in the same clusters, we discovered that miRNAs in the same cluster generally have cooperative effects in repression target genes in various human tissues (fig. 3*D–F*). Furthermore, we presented experimental evidence that members in the *miR-17–92* cluster tend to target overlapping sets of target genes (fig. 4*I*). The co-operative effects of clustered miRNAs on the same sets of target genes can be generalized in figure 6*A*, by which a miRNA cluster exerts "reinforced" effects to repress a specific target gene.

Previous *ab initio* analysis identified large numbers of gene regulatory networks coordinately regulated by miRNAs in the same clusters (Lewis et al. 2003; Bartel 2009; Kim et al. 2009; Yuan et al. 2009; Wang et al. 2011). Consistent with those studies, our bioinformatic analysis based on the co-expression of miRNAs and mRNAs in various human tissues also suggests miRNAs in the same clusters tend to co-regulate genes in the same pathways (supplementary table S9, Supplementary Material online, see Materials and Methods). We further validated such patterns based on the high-confidence target genes of the *miR-17–92* cluster determined in this study. Our findings can be summarized as the following:

First, miRNA clustering expanded the transcription factor-miRNA co-regulatory feed-forward loops (TF:miRNA FFL). Previous studies demonstrated that miRNAs are significantly enriched in FFLs and the designing principles have several desired properties in regulating target genes (Herranz and Cohen 2010). By integrating the high-confidence targets of *miR-17–92* cluster and the TF:target gene regulation information obtained from previous studies (Wang et al. 2011; Han et al. 2015), we identified 61 FFLs reconstructed by a single miRNA in the *miR-17–92* cluster (fig. 6*B*). However, we identified 77 motifs with the TF and its target gene separately targeted by two miRNAs in the *miR-17–92* cluster (fig. 6*C*).

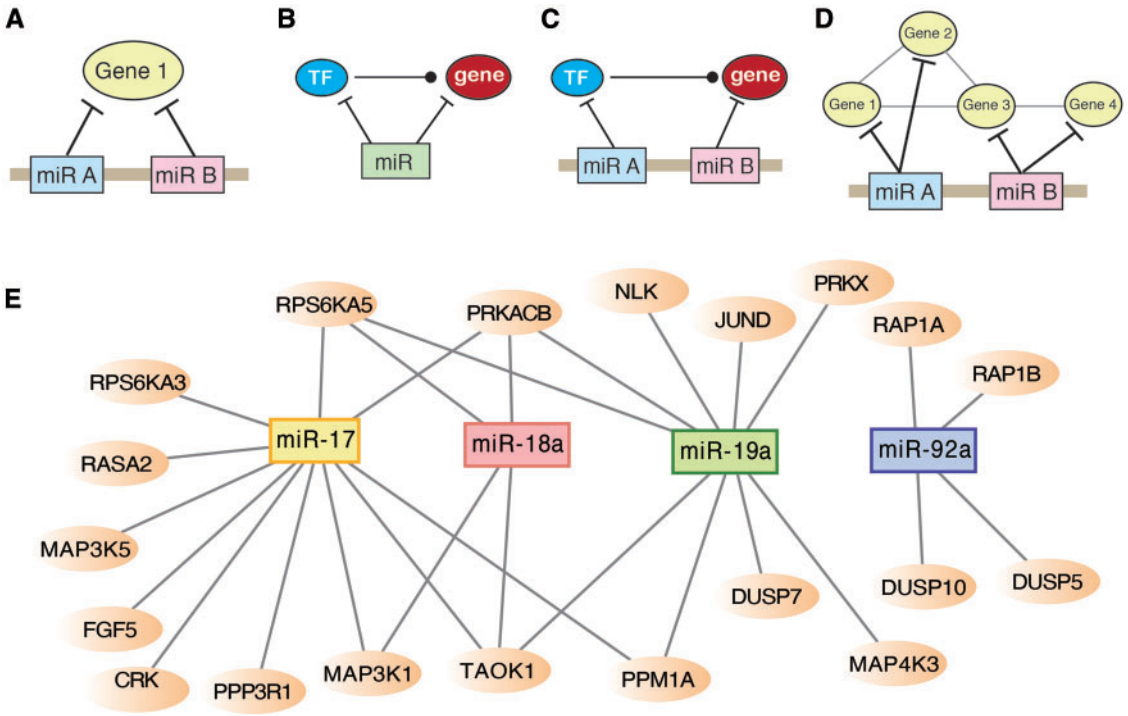

Fig. 6. Network motifs that are coordinately regulated by the *miR-17–92* cluster. (*A*) miRNAs in the same cluster target the same gene so that this miRNA cluster exerts "reinforced" repression effects on the specific target genes. (*B*) A single miRNA represses target genes through a feed forward loop. (*C*) A miRNA cluster represses target genes through a motif similar to a feed forward loop. (*D*) Clustered miRNAs broadly regulate biological pathways and help increase connectivity among nodes in regulatory network. (*E*) The high-confidence target genes of the *miR-17–92* cluster are significantly enriched in MAPK signaling pathway (the KEGG pathway accession hsa04010).

Since abundances of miRNAs in the *miR-17–92* cluster are highly correlated, the motifs in figure 6*C* might have similar properties as the motifs in figure 6*B*. Therefore, the miRNA cluster expands the scopes of the TF:miRNA regulatory motifs.

Second, miRNAs in the same clusters link the motifs into regulatory modules (or biological pathways). It was established that the target genes of miRNAs tend to be significantly enriched in certain biological pathways either to switch or to stabilize the gene regulatory networks (Vlachos et al. 2015). However, due to the limited number of target genes one miRNA can potentially regulate, the impact a single miRNA on the network might be limited as well. Clustered miRNAs can significantly increase the connectivity among the nodes in the regulatory networks (fig. 6*D*). By searching the target genes in the KEGG pathways, we found that at least 13 KEGG pathways that are significantly enriched with the high-confidence target genes of the *miR-17–92* cluster (fig. 6*E* and supplementary fig. S10 and table S10, Supplementary Material online). Overall, our experimental results confirmed previous observations that clustered miRNAs tend to target functionally relevant genes.

Our "functional co-adaptation" model of miRNA clustering requires the act of natural selection. Under our model, in the initial stage of cluster formation and expansion, positive Darwinian selection might drive the newly emerged miRNAs to develop function related to the pre-existing miRNAs in the same cluster or drive the evolution of all the new miRNAs in the same cluster to develop related function. Once the cluster is fully established, the miRNAs in the same cluster will be maintained by purifying selection and become highly conserved thereafter. Since many young miRNAs are undergoing function development, we expect to detect the signals of positive selection in the clusters comprised of such miRNAs. We calculated the divergence between human and other primate species (gorilla, orangutan, and macaque) for the ancient miRNAs as well as the young miRNAs that were clustered and non-clustered (see Materials and Methods for details). The ancient miRNA loci (those originated before the divergence of human and mouse) generally have significantly lower divergence levels than the intergenic sites which are evolutionarily neutral (fig. 7*A* for human and gorilla and fig. 7*B* for human and macaque comparisons), suggesting they are under strong selective constraints. However, the young miRNA loci (which originated after the radiation of primates but before the split of human and gorilla), whether clustered or nonclustered, generally have comparable divergence levels with the neutral sites between human and gorilla or between human and macaque (fig. 7*A* and *B*). Interestingly, when we conducted generalized McDonald–Kreitman's tests on the young miRNA precursors by comparing the fixed DNA changes across four primate species (human, gorilla, orangutan, and macaque) versus the polymorphisms in the human populations determined in the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015), we found the fixed/polymorphism ratio was significantly higher for the

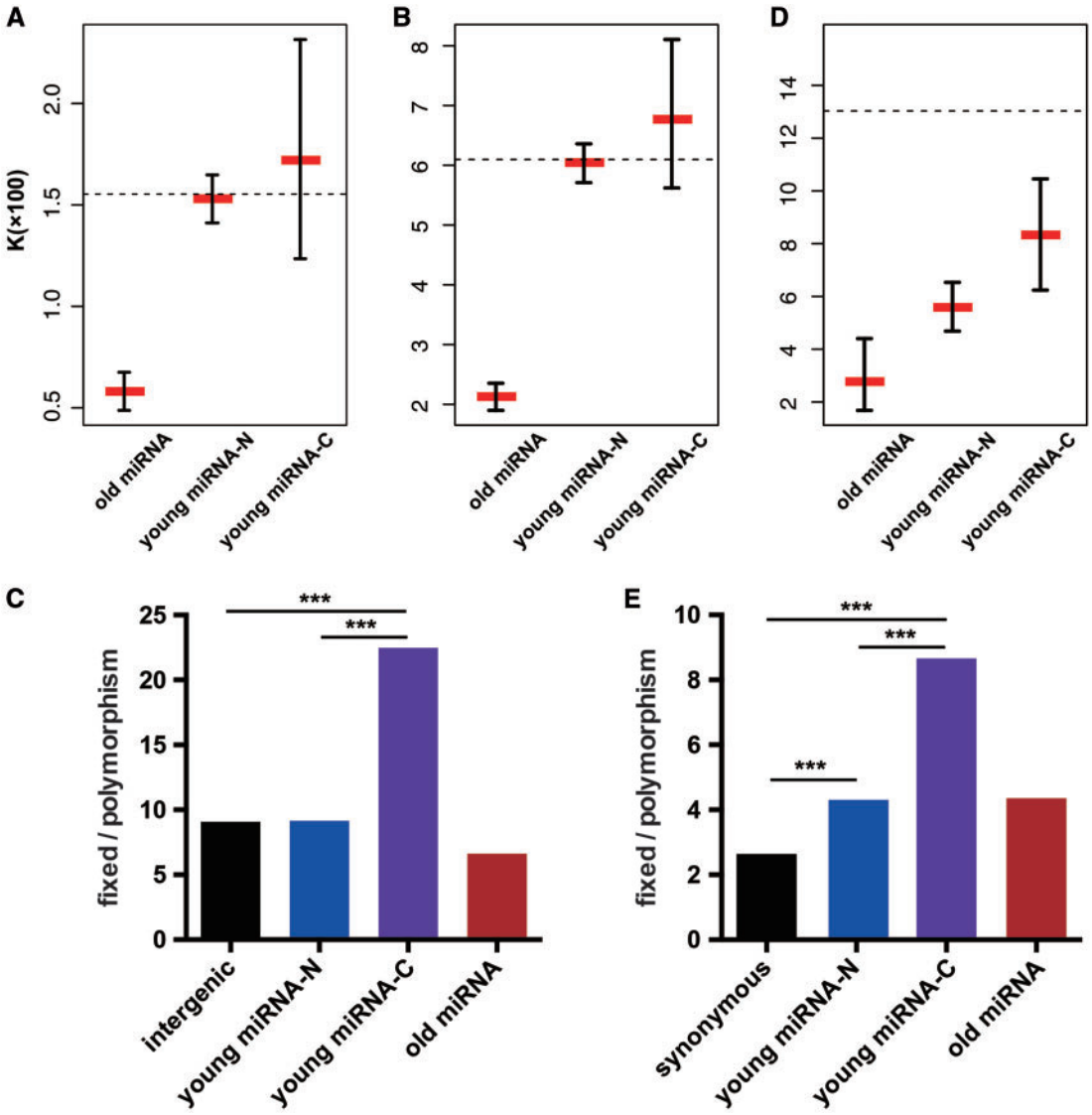

Fig. 7. Signatures of positive Darwinian selection in precursors of young miRNAs that are clustered. (*A*) Divergence in precursors of old miRNAs, young miRNAs that are nonclustered and clustered between human and gorilla. The dash line denotes divergence level of intergenic sites. The mean divergence is in red and the 95% CI were obtained by randomly sampling the sites in each category for 1000 replicates. (*B*) Divergence in precursors of old miRNAs, young miRNAs that are nonclustered and clustered between human and macaque. The dash line denotes divergence level of intergenic sites. The mean divergence is in red and the 95% CI were obtained by randomly sampling the sites in each category for 1,000 replicates. (*C*) The fixed/polymorphism ratio for intergenic region, precursors of young miRNAs that are not clustered, young miRNAs that are clustered, and old miRNAs in primates. The fixed DNA changes were counted across genomes of human, gorilla, orangutan, and macaque; the polymorphism data were retrieved from the 1000 Genomes Project. "***" denotes $P \leq 0.001$ (Fisher's exact test). (*D*) Divergence in precursors of old miRNAs, young miRNAs that are non-clustered and clustered between *D. melanogaster* and *D. simulans*. The dash line denotes divergence level of synonymous sites. The mean divergence is in red and the 95% CI were obtained by randomly sampling the sites in each category for 1,000 replicates. (*E*) The fixed/polymorphism ratio in synonymous sites, precursors of young miRNAs that are nonclustered, young miRNAs that are clustered, and old miRNAs for *D. melanogaster* and *D. simulans*. The fixed DNA changes were counted between *D. melanogaster* and *D. simulans*; the polymorphism data were retrieved from the DGRP project. "***" denotes $P \leq 0.001$ (Fisher's exact test). In (A, B, and D), "young miRNA-N" denotes precursors of young miRNAs that are nonclustered and "young-miRNA-C" denotes precursors of young miRNAs that are clustered.

young miRNA loci that were clustered than the young miRNA loci that were non-clustered (the ratio was 22.47 vs. 9.16, $P < 0.0001$, Fisher's exact test; in each class the sites were pooled together); strikingly, the fixed/polymorphism ratio for the clustered young miRNAs was even significantly higher than that of the intergenic sites (the ratio was 22.47 vs. 9.08, $P < 0.0001$, Fisher's exact test), suggesting positive selection might have been driving the clustered young miRNAs to develop function (fig. 7C). We also compared the divergence between *D. melanogaster* and *D. simulans* for the ancient miRNAs and the young miRNAs that were clustered versus nonclustered. All the three types have significantly lower divergence levels than the synonymous sites, which are putatively neutral in *Drosophila* (fig. 7D).

Furthermore, by comparing the fixed difference between *D. melanogaster* and *D. simulans* versus the polymorphic mutations in the DGRP project of *D. melanogaster* (Mackay et al. 2012), we detected strong signals of positive Darwinian selection on precursors of young miRNAs (originated after the split of *D. melanogaster* and *D. pseudoobscura*) that were clustered (the fixed/polymorphism ratio was 8.66, significantly higher than that of synonymous sites 2.64; $P = 4 \times 10^{-12}$, Fisher's exact test, fig. 7*E*) and the young miRNAs that were nonclustered (the fixed/polymorphism ratio was 4.30, also significantly higher than that of synonymous sites; $P = 4 \times 10^{-7}$, Fisher's exact test, fig. 7*E*). These results are consistent with previous studies that many adaptive mutations were needed for newly-emerged miRNAs to develop function (Zhang et al. 2007; Lu, Fu, et al. 2008; Zhang et al. 2008; Lyu et al. 2014; Mohammed, Bortolamiol-Becet, et al. 2014). Notably, the signals of positive selection on the clustered young mRNA precursors are significantly stronger than signals of the nonclustered young miRNA precursors ($P = 0.001$, Fisher's exact test, fig. 7*E*). Taken together, these new results suggest that positive selection might be the driving force underlying the formation and evolution of miRNA clustering, which well supports the "functional co-adaptation" model we propose for miRNA clustering.

Conclusions

In this study, we systematically investigated the origin and evolutionary patterns of miRNA clustering in animal species. Our analysis revealed that miRNAs are significantly enriched in clusters, especially for the evolutionary conserved ones. We found that *de novo* formation and duplication are two major mechanisms for creating new miRNAs in clusters; however, the fate of a new miRNA is greatly affected by its genomic location. We proposed a "functional co-adaptation" model to explain how clustering helps new miRNAs survive and develop functions related to the pre-existing miRNAs in that cluster. By miRNA transfection and extensively profiling the transcriptome alternation with mRNA-Seq, we confirmed that clustered miRNAs cooperatively target overlapping sets of genes. By identifying conserved targets of *miR-92a* in both human and fly, we further demonstrated that functional co-adaptation between miRNAs in the same cluster might be the driving force for clustering. Based on the high-confidence targets of *miR-17–92* cluster identified in this study, we found that miRNAs in this cluster have cooperative effects on overlapping sets of target genes to reinforce repression (fig. 6*A*). Furthermore, the network motif analysis indicated that miRNA clustering greatly expands the miRNA-mediated FFLs (fig. 6*C*) and can significantly increase the connectivity among the nodes in the regulatory networks (fig. 6*D*). Our results suggested that functional co-adaptation might be one driving force for new miRNA clusters to persist in the initial stage of formation and to be maintained by natural selection once their coordinated regulatory effects on target genes are established. Our study provides novel insights into the evolutionary principles and significance of genomic linkage of regulatory elements.

Materials and Methods

miRNA Annotation and Genomic Location

The genomic locations of miRNAs in human, mouse, chicken, zebrafish, worm, and fly were compiled from miRBase (miRBase V21, http://www.mirbase.org/, last accessed 16 May 2016) (Griffiths-Jones et al. 2006, 2008; Kozomara and Griffiths-Jones 2011, 2014). Information regarding the genome assemblies used in this study is as follows: *Homo sapiens* (GRCh38), *Mus musculus* (GRCm38), *Gallus gallus* (Gallus-gallus-4.0), *Caenorhabditis elegans* (WBcel235), *Danio rerio* (Zv9), and *D. melanogaster* (BDGP5.0).

Similar to previous studies (Altuvia et al. 2005; Griffiths-Jones et al. 2008; Marco et al. 2013), the clustering of miRNA genomic locations is determined if two neighboring miRNAs are located within 10 kb and are in the same strand. BEDtools (Quinlan and Hall 2010) was used to randomly shuffle the genomic locations of miRNA loci to test whether the observed number of clustered miRNAs was significantly higher than the simulated number of clustered miRNAs in each species. For each *in silico* analysis, 1,000 replicates of shuffling were performed in the simulations.

miRNA Duplication and Conservation

We employed two complementary approaches to identify miRNA duplications of different ages. For the recent miRNA duplication events, we searched the precursor sequences using a BLAST search with a word size of 6 and *E*-value cutoff of 0.0001. For the ancient miRNA duplication events, we delimitated the paralogous copies to have homologous mature sequences with fewer than five mismatches and identical seeds (positions 2–8 of the mature miRNA).

The genome alignments and phylogenetic information of 100 vertebrate species, 12 *Drosophila* species, and seven worm species were obtained from the UCSC Genome Browser (genome.ucsc.edu, last accessed 16 May 2016). The conservation information for the miRNAs was defined with similar criteria in the TargetScan package (V7.0, www.targetscan.org, last accessed 16 May 2016). Besides the criteria we mentioned before, we further required: (1) the orthologous sequences of miRNA precursors form stable hairpin structures ($\Delta G < -20$ kcal/mol), and (2) the seeds of orthologous miRNAs were identical across the examined clades.

Since many of the curated miRNAs in miRBase might not be *bona fide* (Fromm et al. 2015), we extensively compiled the small RNAs sequenced with AGO IP-Seq in human and flies from previous studies (supplementary tables S2 and S3, Supplementary Material online). We required an annotated human mature miRNA in miRBase have $\geq$10 raw reads in at least one AGO IP-Seq library. Same criteria were required in *D. melanogaster*.

Novel miRNAs Detection

We searched for novel miRNAs with miRDeep2 program (Mackowiak 2011) in all AGO IP-Seq small RNA libraries used in this study. To reduce the false positive rates in miRNA discovery, we employed sets of criteria based on features of the known miRNAs. The following criteria was used

for novel miRNA identification: (1) miRDeep2 score of novel miRNAs >4; (2) signal-to-noise ratio >15; (3) the length of the precursor >60 nt, and (4) the novel miRNAs are detected in at least two independent libraries.

Evolutionary Analysis on the Ancient and Young miRNA Precursors

We downloaded the reciprocal best whole genome alignments between human and gorilla, orangutan, and macaque from the UCSC Genome Browser. For each miRNA, we extracted the alignments of the miRNA precursors as well as the alignments of the intergenic sequences 100 kb flanking the miRNA. To estimate the divergence of miRNA precursors, miRNA loci from each category were pooled together and the confidence intervals of divergence were obtained by randomly sampling the miRNAs of equal numbers for 1,000 replicates. Human polymorphism data were downloaded from the 1000 Genomes Project (http://www.1000genomes.org/, last accessed 16 May 2016). SNPs with minor allele frequency <0.001 were discarded. The number of SNPs in human populations and the fixed DNA changes among human, gorilla, orangutan, and macaque were counted for the miRNA precursors and the flanking intergenic regions. The fixed/polymorphism ratios were calculated for the precursors of young miRNAs (originated after the radiation of primate species but before the split of human and gorilla) that were clustered and nonclustered separately. Similar analysis on the flanking intergenic regions were conducted. The CpG sites and the repetitive sequences in primates were removed in the generalized McDonald–Kreitman's tests. Similar analysis procedures were conducted in *Drosophila*. The whole genome alignments between *D. melanogaster* and *D. simulans* were downloaded from UCSC Genome Browser; the polymorphism data from DGRP project (dgrp2.gnets.ncsu.edu, last accessed 16 May 2016) was downloaded and the SNPs with minor allele frequency <0.05 were discarded. The synonymous sites in the coding regions flanking the miRNAs were treated as neutral regions. The K80 model (Kimura 1980) was used to estimate divergence levels in the miRNA precursors and the flanking non-coding regions. The statistical significance was determined with a Fisher's exact test for the fixed and polymorphism between two categories of sites.

Weighted Gene Co-expression Network Analysis (WGCNA)

To examine whether the abundances of miRNAs in the same clusters are correlated, we collected hundreds of miRNA expression libraries from different human tissue and cell lines from NCBI SRA database (www.ncbi.nlm.nih.gov/sra, last accessed 16 May 2016). Technical replicates and highly correlated datasets ($r > 0.95$) were filtered to avoid expressional bias and each miRNA was required to have ≥10 raw reads in each library. We also eliminated all duplicated miRNAs (e.g., hsa-miR-1-1/hsa-miR-1-2) to avoid inaccurate read counts. Finally, 683 miRNAs in 123 miRNA expression datasets were kept and WGCNA was performed as described in tutorials. The soft threshold for constructing a signed weighted correlation network ($\beta = 7$) was determined with the scale-free topology criterion applied to all 683 miRNAs ($r > 0.9$). Network construction performed by default dynamic tree-cutting method and smaller minimum module size was set to $n = 10$ genes.

miRNA and mRNA Expression Data and Target Prediction

The expression levels of miRNAs and mRNAs from five human tissues were extracted from a previous study by the Henrik Kaessmann lab (Brawand et al. 2011; Meunier et al. 2013). In each human tissue, we required that each miRNA could be detected with at least 50 raw reads. We employed the TargetScan algorithm to predict miRNA target genes with the following criteria: (1) the miRNA and its target genes are co-expressed in at least one human tissue, (2) the miRNA is conserved in mammals or broadly conserved in vertebrates, and (3) the target sites is conserved (aggregate $P_{CT} > 0.5$). For each human tissue, we separated the expressed genes into three distinct groups based on miRNA targeting patterns: (1) T1: genes that are targeted by only one co-expressed miRNA family in a tissue; (2) TM_C: genes that are targeted by at least two distinct conserved seeds from a miRNA cluster; and (3) TM_N: genes that are targeted by at least two distinct conserved seeds but none of them were from the same miRNA cluster (TargetScan $P_{CT} > 0.5$ were employed in the target prediction).

Cell Culture and miRNA Transfections

Human 293FT cells were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% fetal bovine serum at 37°C with 5% CO_2. Fly S2 cells were cultured in Schneider's Insect Medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum at 27°C. We selected four distinct mature miRNAs in the *miR-17–92* cluster (*miR-17-5p*, *miR-18a-5p*, *miR-19a-3p*, and *miR-92a-3p*, abbreviated as *miR-17*, *miR-18a*, *miR-19a*, and *miR-92a*, respectively) and transfected 50 nM miRNA duplex into human 293FT cells with Lipofectamine 2000 (Invitrogen). The guide and passenger sequence of negative control (NC) is 5′UUCUCCGAACGUGUCACGUUU3′ and 5′ACGUGACACGUUCGGAGAAUU3′, respectively. The untreated cells and cells 32-h post-transfection of miRNA mimics were harvested with TRIzol Reagent for total RNA isolation (Invitrogen).

Directional mRNA-Seq Library Preparation and Data Analysis

Total RNAs of cells were extracted using the TRIzol reagent (Thermo Fisher) and chloroform. For mRNA libraries, 10–20 μg poly(A)+ RNAs were selected on oligo-dT25 DynaBeads (Thermo Fisher) and fragmented at 70°C for 15 min in fragmentation reagent (Thermo Fisher). mRNA fragments within 40–80 nt were size selected by 15% TBE-Urea gel. After 3′ and 5′ ligation, size selected RNA fragments were reverse-transcribed with SuperScript III (Invitrogen). Sequence of 3′ adaptor is TGGAATTCTCGGGTGCCAAGG. All cDNAs were amplified by 12 PCR cycles with Phusion High-Fidelity DNA polymerase (NEB) and the products within the correct size ranges were collected from 20% TBE gels for quality tests (Fragment Analyzer, Agilent Technologies) and

sequencing (Platform: Illumina HiSeq-2500; Read length: 50 bp, single-end).

For all sequencing datasets, 3′ adaptor sequences were clipped by Cutadapt software (Martin 2011). The remaining reads were aligned to human reference genome (GRCh37) and fly genome (BDGP 5.0) using STAR (Dobin et al. 2013) (Mapping statistics were summarized in supplementary table S11, Supplementary Material online). The expression levels of transcripts were counted with eXpress (Roberts and Pachter 2013) and the most highly expressed transcripts of each gene in wild type condition were selected in the analysis. Only genes with RPKM >1 were preserved in the down-stream analysis. The NOISeq (Tarazona et al. 2012) package was used to normalize expression levels and detect differential expressed genes (q value cutoff is 0.9). All gene ontology analysis was performed by DAVID (Dennis et al. 2003).

miRNA Clustering and Functional Coordination in Biological Pathways

The human biological pathway annotations were extracted from the ConsensusPathDB database (Kamburov et al. 2013). For each human tissue, we first tested the enrichment of miRNA target genes in the biological pathways. miRNAs with identical conserved seeds were grouped into one miRNA family, and we employed the TargetScan algorithm to predict the conserved target genes ($P_{CT} > 0.5$ was used as the cutoff). For each conserved miRNA seed, we tested whether its target genes are significantly enriched in a pathway by Fisher's exact test. For each seed that has target genes significantly enriched in a biological pathway ($P < 0.001$ with Fisher's exact test), we also randomly sampled the same number of target genes in the co-expressed mRNAs, and then we pooled the randomly sampled target genes for each seed in the cluster together. We compared the observed number of genes targeted by a miRNA cluster with the simulated number of target genes of that cluster. This process was repeated 1,000 times, and a P value was calculated to test whether a miRNA cluster overall has a significantly higher number of target genes in a biological pathway than expected by random chance. We used the Benjamini & Hochberg method to adjust the P values for multiple testing.

Statistical Tests

All the statistical tests were performed under the open source statistics package R (www.r-project.org, last accessed 16 May 2016).

Data Availability

The NCBI SRA accession number for all mRNA-Seq data of human 293FT cells is SRP067876. The accession number for all mRNA-Seq data of fly S2 cells is SRP067904.

Supplementary Material

Supplementary tables S1–S11 and figures S1–S10 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Dr Fuchou Tang and Dr Yangming Wang for critical reading of this manuscript. We thank Dr Fuchou Tang for providing the 293FT cell line. We thank the Henrik Kaessmann lab for making the large-scale sequences of mRNA and miRNA expression data publicly available. This study was supported by grants from the National Natural Science Foundation of China (Nos. 91431101 and 31571333) and from the Peking-Tsinghua Center for Life Sciences to Jian Lu.

References

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S McVean GA. 1000 Genomes Project Consortium, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74.

Agarwal V, Bell GW, Nam J-W, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4:e05005.

Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.* 33:2697–2706.

Ambros V. 2003. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 113:673–676.

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

Bartel DP. 2009. MicroRNA target recognition and regulatory functions. *Cell* 136:215–233.

Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11:241–247.

Bazzini AA, Lee MT, Giraldez AJ. 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336:233–237.

Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung J-H, Okamura K, Dai Q, Bortolamiol-Becet D, Martin R, et al. 2011. Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res.* 21:203–215.

Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, Vos J, Verloop R, van de Wetering M, Guryev V, Takada S, et al. 2006. Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. *Genome Res.* 16:6–9. author reply 9–10.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348.

Chen S, Krinsky BH, Long M. 2013. New genes as drivers of phenotypic evolution. *Nat Rev Genet.* 14:645–660.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:P3.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2003. MicroRNA targets in Drosophila. *Genome Biol.* 5:R1.

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, et al. 2007. High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One* 2:e219.

Fay JC, Wu C-I. 2003. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet.* 4:213–235.

Friedman RC, Farh KKH, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 19:92–105.

Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, et al. 2015. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu Rev Genet*. 49:213–242.

Gangaraju VK, Yin H, Weiner MM, Wang J, Huang XA, Lin H. 2011. *Drosophila* Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat Genet*. 43:153–158.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet*. 10:94–108.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312:75–79.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 34:D140–D144.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 36:D154–D158.

Guo Z, Maki M, Ding R, Yang Y, Zhang B, Xiong L. 2014. Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Sci Rep*. 4:5150.

Haldane JBS. 1933. The part played by recurrent mutation in evolution. *Am Nat*. 67:5.

Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, et al. 2015. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep*. 5:11432.

Han Y-C, Vidigal JA, Mu P, Yao E, Singh I, González AJ, Concepcion CP, Bonetti C, Ogrodowski P, Carver B, et al. 2015. An allelic series of miR-17–92–mutant mice uncovers functional specialization and cooperation among members of a microRNA polycistron. *Nat Genet*. 47:766–775.

Herranz H, Cohen SM. 2010. MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev*. 24:1339–1344.

Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet*. 38(Suppl):S20–S24.

Jacob F, Perrin D, Sanchez C, Monod J. 1960. Operon: a group of genes with the expression coordinated by an operator. *C R Hebd Seances Acad Sci*. 250:1727–1729.

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human microRNA targets. *PLoS Biol* 2:e363.

Kamburov A, Stelzl U, Lehrach H, Herwig R. 2013. The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Res*. 41:D793–D800.

Kim VN, Nam JW. 2006. Genomics of microRNA. *Trends Genet*. 22:165–173.

Kim YK, Yu J, Han TS, Park SY, Namkoong B, Kim DH, Hur K, Yoo MW, Lee HJ, Yang HK, et al. 2009. Functional links between clustered microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. *Nucleic Acids Res*. 37:1672–1681.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 16:111–120.

Kozomara A, Griffiths-Jones S. 2011. MiRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 39:152–157.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 42:D68–D73.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294:853–858.

Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. 2002. Identification of tissue-specific MicroRNAs from mouse. *Curr Biol*. 12:735–739.

Lai EC, Tomancak P, Williams RW, Rubin GM. 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biol*. 4:R42.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294:858–862.

Lawrence J. 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr Opin Genet Dev*. 9:642–648.

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.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian MicroRNA targets. *Cell* 115:787–798.

Liang H, Li W-H. 2009. Lowly expressed human microRNA genes evolve rapidly. *Mol Biol Evol*. 26:1195–1198.

Long M, VanKuren NW, Chen S, Vibranovski MD. 2013. New gene evolution: little did we know. *Annu Rev Genet*. 47:307–333.

Lu J, Fu Y, Kumar S, Shen Y, Zeng K, Xu A, Carthew R, Wu C-I. 2008. Adaptive evolution of newly emerged micro-RNA genes in Drosophila. *Mol Biol Evol*. 25:929–938.

Lu J, Shen Y, Carthew RW, Wang SM, Wu C-I. 2010. Reply to evolutionary flux of canonical microRNAs and mirtrons in *Drosophila*. *Nat Genet*. 42:9–10.

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu C-I. 2008. The birth and death of microRNA genes in Drosophila. *Nat Genet*. 40:351–355.

Lyu Y, Shen Y, Li H, Chen Y, Guo L, Zhao Y, Hungate E, Shi S, WuChung ICI, Tang T. 2014. New MicroRNAs in *Drosophila*—birth, death and cycles of adaptive evolution. *PLoS Genet*. 10:e1004096.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* genetic reference panel. *Nature* 482:173–178.

Mackowiak SD. 2011. Identification of novel and known unit 12.10 miRNAs in deep-sequencing data with miRDeep2. *Curr Protoc Bioinforma* 2011:Chapter 12.

Marco A, Ninova M, Ronshaugen M, Griffiths-Jones S. 2013. Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic Acids Res*. 41:7745–7752.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J*. 17:10.

Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes. *Genome Res*. 23:34–45.

Mogilyansky E, Rigoutsos I. 2013. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ*. 20:1603–1614.

Mohammed J, Bortolamiol-Becet D, Flynt AS, Gronau I, Siepel A, Lai EC. 2014. Adaptive evolution of testis-specific, recently evolved, clustered miRNAs in *Drosophila*. *RNA* 20:1195–1209.

Mohammed J, Siepel A, Lai EC. 2014. Diverse modes of evolutionary emergence and flux of conserved microRNA clusters. *RNA* 20:1850–1863.

Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. *Nat Rev Genet*. 8:857–868.

Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE. 2008. Chromatin structure analyses identify miRNA promoters. *Genes Dev*. 22:3172–3183.

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P, et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408:86–89.

Phillips PC. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 9:855–867.

Price MN, Huang KH, Arkin AP, Alm EJ. 2005. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res*. 15:809–819.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev*. 20:3407–3425.

Rajewsky N. 2006. microRNA target predictions in animals. *Nat Genet*. 38(Suppl):S8–S13.

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* 10:71–73.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res*. 14:1902–1910.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* 448:83–86.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res*. 17:1850–1864.

Ryazansky SS, Gvozdev VA, Berezikov E. 2011. Evidence for post-transcriptional regulation of clustered microRNAs in Drosophila. *BMC Genomics* 12:371.

Saini HK, Griffiths-Jones S, Enright AJ. 2007. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci U S A*. 104:17719–17724.

Tarazona S, García F, Ferrer A, Dopazo J, Conesa A. 2012. NOIseq: a RNA-seq differential expression method robust for sequencing depth biases. *EMBnetJ*. 17:18.

Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, et al. 2008. Targeted deletion reveals essential and overlapping functions of the miR-17–92 family of miRNA clusters. *Cell* 132:875–886.

Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG. 2015. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res*. 43:W460–W466.

Wang J, Haubrock M, Cao K-M, Hua X, Zhang C-Y, Wingender E, Li J. 2011. Regulatory coordination of clustered microRNAs based on microRNA-transcription factor regulatory network. *BMC Syst Biol*. 5:199.

Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ. 2009. High-resolution human core-promoter prediction with CoreBoost_HM. *Genome Res*. 19:266–275.

Yuan X, Liu C, Yang P, He S, Liao Q, Kang S, Zhao Y. 2009. Clustered microRNAs' coordination in regulating protein–protein interaction network. *BMC Syst Biol*. 3:65.

Zhang R, Peng Y, Wang W, Su B. 2007. Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res*. 17:612–617.

Zhang R, Wang Y-Q, Su B. 2008. Molecular evolution of a primate-specific microRNA family. *Mol Biol Evol*. 25:1493–1502.

Zhao H, Zhu L, Zhu Y, Cao J, Li S, Huang Q, Xu T, Huang X, Yan X, Zhu X. 2013. The Cep63 paralogue Deup1 enables massive de novo centriole biogenesis for vertebrate multiciliogenesis. *Nat Cell Biol*. 15:1434–1444.

Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. 2010. Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat*. 20:281–300.

Zhao Y, Lin J, Xu B, Hu S, Zhang X, Wu L. 2014. MicroRNA-mediated repression of nonsense mRNAs. *Elife* 8:e03032.