Genome-wide impact of a recently expanded microRNA cluster in mouse

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1112772108">http://dx.doi.org/10.1073/pnas.1112772108</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Dec 22 04:28:52 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/71158">http://hdl.handle.net/1721.1/71158</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Genome-wide impact of a recently expanded microRNA cluster in mouse

Grace X. Y. Zheng\textsuperscript{a,b,c,1}, Arvind Ravi\textsuperscript{a,b,1}, Genevieve M. Gould\textsuperscript{d}, Christopher B. Burge\textsuperscript{b,4}, and Phillip A. Sharp\textsuperscript{a,c,3}

\textsuperscript{a}MIT Koch Institute for Integrative Cancer Research, Cambridge, MA 02139; \textsuperscript{b}Computational and Systems Biology Graduate Program, Cambridge, MA 02139; \textsuperscript{c}Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and \textsuperscript{d}Harvard-MIT Health Sciences and Technology Program, Cambridge, MA 02139

Contributed by Phillip A. Sharp, August 10, 2011 (sent for review April 6, 2011)

Variations in microRNA (miRNA) gene and/or target repertoire are likely to be key drivers of phenotypic differences between species. To better understand these changes, we developed a computational method that identifies signatures of species-specific target site gain and loss associated with miRNA acquisition. Interestingly, several of the miRNAs implicated in mouse 3′ UTR evolution derive from a single rapidly expanded rodent-specific miRNA cluster. Located in the intron of \textit{Sfmbt2}, a maternally imprinted polycomb gene, these miRNAs (referred to as the \textit{Sfmbt2} cluster) are expressed in both embryonic stem cells and the placenta. One abundant miRNA from the cluster, \textit{mir-467a}, functionally overlaps with the mir-290-295 cluster in promoting growth and survival of mouse embryonic stem cells. Predicted novel targets of the remaining cluster members are enriched in pathways regulating cell survival. Two relevant species-specific target candidates, \textit{Lats2} and \textit{Dedd2}, were validated in cultured cells. We suggest that the rapid evolution of the \textit{Sfmbt2} cluster may be a result of intersex conflict for growth regulation in early mammalian development and could provide a general model for the genomic response to acquisition of miRNAs and similar regulatory factors.

The emergence of novel regulatory interactions provides a critical means of evolutionary change (1). By introducing new regulatory elements, or simply rewire existing ones, organisms can adapt to alterations in their environment. In the case of protein coding genes, a number of precedents for these principles have been established. For instance, the transcriptional cofactor TAFII105 emerged in mammals to specifically direct expression of a subset of genes in ovarian follicle cells (2). Similarly, the remapping of existing transcription factor networks through promoter evolution is thought to be widespread, even between similar species (3, 4). However, given the relatively small increases in protein coding genes across more complex organisms, many have turned their attention to the roles of noncoding RNAs in explaining evolutionary changes (5).

Among noncoding RNAs, miRNAs are thought to be particularly relevant to phenotypic differences between species, with some claiming that miRNA gene number scales roughly with organismal complexity (6). Although only approximately 22 nucleotides in length, miRNAs can repress the expression of hundreds of genes posttranscriptionally, making them ideal candidates for the establishment or alteration of large regulatory networks. Indeed, it has been suggested that miRNAs are in fact more “evolvable” elements than transcription factors because targeting of a novel sequence requires changing only one or a few bases rather than a complex set of amino acids (4). However, the constraints of processing require that precursors be present as hairpin structures in the genome, therefore favoring their emergence via certain evolutionary routes. Three mechanisms in particular have been hypothesized for miRNA generation: (i) duplication of existing miRNAs; (ii) processing of transposable elements with terminal inverted repeats; or (iii) processing of hairpin structures generated by mutation (7). Following gene duplication in the first mechanism, targeting could be altered by subsequent base substitutions, particularly those corresponding to positions 2–7 from the 5′ end of the miRNA, known as the miRNA “seed” (8).

In contrast to miRNA evolution, which can simultaneously introduce multiple novel interactions, mutations in single target sites can provide more modest increments of evolutionary change. Nevertheless, these changes could also be important drivers of organismal differences. For instance, variations in Nodal family targeting by the miR-430/427/302 family guide differences in germ layer specification during development across a range of vertebrates (9). Even within a species, presence or absence of even a single target interaction may have notable effects. In the case of Texel sheep, a forward genetic screen identified a single base change in the myostatin 3′ UTR that creates a miRNA target site and confers muscular hypertrophy (10). In humans, a similar case has been reported for Tourette syndrome, where changes in the SLITRKL transcript enhance repression by hsa-miR-189 (11). A more global assessment using human SNP datasets identified a number of variants that may alter miRNA binding (12). This study suggested that many target sites in the human genome are under purifying selection to maintain the presence or absence of a miRNA target site but only identified a single instance of positive selection, presumably because few SNPs passed the high heterozygosity thresholds required for informative intraspecies analysis.

Here we describe a method for identifying species-specific changes in miRNA-target relationships and use it to identify miRNA innovations in the mouse genome. We find that many of the mouse-specific changes correspond to a single genomic locus, located in the intron of an imprinted polycomb group gene (13). The miRNAs in this cluster, which we refer to as the \textit{Sfmbt2} cluster, appear to have expanded through a duplication-divergence mechanism, generating both novel seeds and seeds corresponding to earlier miRNA families. Finally, predictions suggest that these miRNAs may in part regulate targets involved in growth and survival, in line with predicted roles for the mir-290-295 cluster, the dominant cluster in mouse ES cells. As the expression patterns of these two clusters appear to mirror one another, we suggest that the \textit{Sfmbt2} cluster may promote proliferation in extraembryonic tissue, serving as a counterpart to the mir-290-295 cluster in early murine development.

Results

Positive Selection Acts on Target Sites of Many Species-Specific miRNAs. To better understand how gene networks respond to miRNA
introduction, we examined the 3′ UTR landscape of the mouse genome for signatures that represent species-specific responses to different miRNA repertoires, using relative sequence conservation with humans as a background for comparison. Comparative analysis of 3′ UTRs should reveal two groups of genes associated with species-specific miRNAs: (i) genes that gained binding sites to novel miRNAs and whose downregulation improved fitness; and (ii) genes that were inadvertently targeted by novel miRNAs and that subsequently lost target sites to maintain expression.

To test for the presence of each group of genes, aligned 3′ UTRs from the mouse and human genomes were analyzed for signatures of target site gain and loss (Fig. 1A). For target site gain, we considered seed match sites present in mouse 3′ UTRs, whereas for site loss, we considered the reverse scenario, requiring a seed match site in human but not necessarily in mouse UTRs. While these putative assignments of site gain or loss in a given species are the most evolutionarily parsimonious, it is possible that some resulted from loss and gain, respectively, in the second species used for comparison. Despite this caveat, functional species-specific miRNAs would be expected to generate both gain and loss signatures at the genome-wide scale. Although sites with mismatches in the seed region have been described, in-centering of pairing sites (14), we defined predicted targets using the traditional method of seed complementarity (7mer-M8 type sites), as it is the most common type of target site and the best characterized to date (8). For miRNA targets that have been positively selected to gain a target site, we would expect higher nucleotide variation between species at the miRNA binding site than at neighboring sequences in the same 3′ UTR (Fig. 1A and B, Left).

To measure this difference, we defined the statistic d as the difference in average per base mutation frequency between the miRNA target site and the frequency in the adjacent 42 nt upstream and 42 nt downstream of the target site (Fig. 1A and B, Middle). Significance of selection was then assessed by comparing d for a given seed to that of control oligonucleotides with similar abundance and composition in mouse and human 3′ UTRs (Fig. 1A and B, Right and SI Appendix). Using this approach, 25 of 201 mouse-specific miRNA seeds (12.4%) showed a signal for gain of target sites (Fig. 2A and SI Appendix), roughly twice the rate observed for all heptamers (6.02%, p = 0.004, Chi-square test). Thus, our analysis detected evidence of selection for acquisition of seed matches to these mouse-specific miRNAs in the mouse genome. As an additional control, evaluation of the reverse complements of these seeds failed to show enrichment in any of our tests, consistent with selection being sensitive to the transcribed strand.

We additionally tested for site loss (Fig. 1C). Again, we found a number of miRNAs (16/201 = 7.96%) whose target sites were lost at greater frequencies in mouse 3′ UTRs than controls, but this fraction was not significantly higher than the fraction observed for control heptamers (6.02%) (Fig. 2A and SI Appendix). Of these 16, a significant fraction, 13, overlapped with those seeds

---

**Fig. 1.** Positive and purifying selection analysis for miRNAs in mouse and human. (A) A flow chart of the analysis: (i) seed match sites in aligned mouse and human 3′ UTRs were identified; (ii) the difference in average per base mutation rate, noted as d, was compared between the miRNA binding site and adjacent sequences; and (iii) d was compared to that of control heptamers. (B) Illustration of the analysis of miR-466a-5p, an example of a mouse-specific miRNA. Left: the alignment of mouse and human 3′ UTRs of a target of miR-466a-5p, where the target site is gained in mouse. Seed binding site is colored in dark red. Middle: mutation rate of the 84-nt sequence spanning the miR-297a-5p match site. Right: distribution of d’s from control heptamers. The black arrow points at the d for miR-297a-5p match sites. (C) Illustration of the analysis of miR-297a-5p, where the target site is lost in mouse. (D) Illustration of the analysis of mir-1, a conserved miRNA between mouse and human. Frequency of mutation denotes the fraction of mismatched nucleotides in the 7-nt window. “-42” indicates the miRNA binding site, “+42” refers to 42-nt downstream of the miRNA binding site, “-42” refers to 42-nt upstream of the miRNA binding site.
Target Sites of Many Mouse-Specific Sfmbt2 Cluster miRNAs Are Under Positive Selection. We next evaluated whether the above 28 seeds (the unique set showing either gain or loss signals above) had any interesting contextual relationships in the mouse genome so that we could better discern evolutionary “hot spots” for miRNA diversity. Interestingly, when the 28 miRNAs with significant target site selection were plotted on a chromosome map of the mouse genome, a prominent clustering was apparent on chromosome 2. Five additional regions on chromosomes 2, 3, 4, 12, and X also contained miRNAs significant for both gain and loss of target sites, suggesting that these miRNAs may be functional (see SI Appendix). The chromosome 12 miRNAs are part of a cluster of 40 miRNAs near the imprinted Dlk-Gtl2 region (17). These miRNAs show preferential expression in the placenta and brain and have been suggested to have key roles in embryonic development and neurogenesis (18). Examination of miRNA expression data revealed that the remaining miRNAs outside of chromosome 12 with both gain and loss signals are expressed in germ tissue (mir-511) or during early embryonic development (mir-1198-5p, mir-3094, mir-380, mir-466k) (19), although detailed characterization of their expression and functions is minimal.

Closer inspection of the largest cluster of hits, located proximally on chromosome 2, revealed that they were derived from the 10th intron of Sfmbt2, a polycomb group gene (20). In total, this intron contains 36 distinct miRNAs, which we refer to as the Sfmbt2 cluster miRNAs (Fig. 2B and SI Appendix). Although the coding region of the Sfmbt2 gene is highly conserved among vertebrates, the intron that harbors the miRNA cluster bears little similarity to those outside of rodent species. Two of the 36 mouse Sfmbt2 miRNAs can be mapped to the homologous intron in rat, but none can be aligned to the corresponding intron in human. Consistent with these findings, expression of the Sfmbt2 cluster—cloned from both T cells (21) and embryonic stem cells (22)—has only been detected in mouse tissue.

Homology of the intronic sequence to the SINE B4 repeat suggests that the cluster derived from a seeding transposition event followed by segmental duplications (see SI Appendix). Consistent with this model, the miRNAs contained within the cluster can be classified into a handful of broad groups based on sequence similarity: miR-297s, miR-466s, miR-467s, and miR-669s (Fig. 2B). Of these four classes, members of the miR-297s and miR-467s can be considered miRNA “families” because they largely consist of the seeds “UGUAUG” and “AAGUGC,” respectively, whereas the miR-466s are a “superfamily” of the shifted seeds “GAUGUG,” “AUGUGU,” and “UGUGUG,” and the miR-669s contain a diversity of seeds. Plausible reconstruction of the evolutionary history of this cluster suggests that miR-669 elements were most similar to the ancestral sequence and that subsequent duplications gave rise to a miR-466 precursor, from which both the miR-467s and miR-297s are derived (see SI Appendix). Sequence variation within these four classes in the Sfmbt2 cluster is sufficient to generate 23 distinct miRNA seeds, the majority of which are not found in humans.

Experimental characterization of the Sfmbt2 miRNA cluster confirmed that they function like canonical miRNAs because they are Dicer dependent (see SI Appendix), repress luciferase targets (see SI Appendix), and appear to globally destabilize messages containing seed complementarity (see SI Appendix). Interestingly, these repressive signatures were evident even for miRNAs without significant evolutionary signatures, although those miRNAs that were identified in our earlier analysis with greater seed conservation (see SI Appendix) showed increased significance, suggesting they may have a greater number of functional targets.

Fig. 2. Target sites of 28 mouse miRNAs show significant positive selection signals, and many fall into the Sfmbt2 locus. (A) Summary statistics of positive selection signals detected from mouse miRNA M8 heptamer binding sites in mouse and human 3’ UTRs. “Count” represents the number of heptamers that were significant in the analysis. “Total” represents the total number of possible heptamers in the specific test category. P-values were the result of Chi-square tests. There are 16,384 heptamers in total. There are 201 mouse-specific miRNA seeds (dog and horse miRNAs were used as an outgroup to determine the species specificity). (B) Genomic structure of the Sfmbt2 miRNA cluster. Precursors of Sfmbt2 miRNAs were mapped to the 10th intron of the Sfmbt2 gene based on the coordinates of precursors. The intron spans a region from roughly 10.39 Mbp to 10.44 Mbp on Chromosome 2. The Sfmbt2 miRNAs are color coded as follows: mir-297s, green; mir-466s, black; mir-467s, blue; mir-669s, magenta.
miR-467 family is best described as species-specific. Given the previously described cell cycle regulatory roles of miR-290-295 and its human counterpart, miR-371-373 (26, 27), we tested whether miR-467 family members might function similarly. Indeed, miR-467a was able to repress luciferase reporters for previously validated miR-295 targets, such as p21 and Lats2 (see SI Appendix) and was comparable to miR-295 in rescuing the G1/S delay of Dicer null cells (Fig. 3A). In addition, miR-467a recapitulated a novel antiapoptotic phenotype we have recently described for miR-295 following genotoxic stress (28) (Fig. 3B and C). Taken together, these results suggest that miR-467 family members could reinforce proliferative cellular programs in parallel with miR-295.

Examination of available global miRNA expression data suggested that in addition to ES cells, placental tissue shows high expression of Sfmbt2 miRNAs (29). Comparing the relative expression of miR-467a by Northern blot analysis, there appeared to be a fourfold upregulation in placental tissue relative to ES cells (Fig. 4A and SI Appendix). Similar expression trends were observed with representatives of the miR-297 (sevenfold increase) and miR-466 (2.5-fold increase) classes (Fig. 4D and SI Appendix). These findings are consistent with miRNA expression data for Sfmbt2 (30), which has been shown to be maternally imprinted (and therefore paternally expressed) in the placenta (20). Therefore, it is likely that expression of the intronic miRNA cluster is dependent on transcription of the entire gene, as has been suggested for many miRNAs similarly positioned within host genes (31). In contrast to Sfmbt2 miRNAs, miR-295, which shares the AAGUGC seed with miR-467a and is highly expressed in ES cells, was barely detectable in the placenta (Fig. 4A and SI Appendix). The differing expression patterns of these miRNAs, also confirmed by a global miRNA profiling dataset (SI Appendix) (29), suggest that the miR-295 family may be principally responsible for the proliferation of ES cells while some members of the Sfmbt2 miRNAs may regulate placental growth.

**Targets of Sfmbt2 miRNA Families Are Enriched in Pathways that Regulate Growth.** Because the remaining members of the Sfmbt2 cluster that are associated with genome-wide signals of selection have seeds unique to mice, we were able to predict potential mouse-specific target transcripts. As defined above, these target sites should have significantly higher mutation rates than adjacent sequences. We assessed their significance with the binomial test and identified 511 placently expressed genes whose target sites have probably been specifically selected for in mouse (p ≤ 0.05, FDR < 0.40) (see SI Appendix). Enrichment in annotated pathways and functions was tested for all 511 genes using ingenuity pathway analysis, with the top statistically significant categories including “Cell Death” (p = 6.7 × 10⁻⁵ to 4.6 × 10⁻² and “Cell Cycle” (p = 2.1 × 10⁻⁴ to 4.7 × 10⁻²). Given the similarity between these annotated functions and known functions of the miR-290-295 cluster, we compared this list to AAGUGC targets to determine whether significant cotargeting was evident. Indeed, a statistically significant overlap of 157 common genes (p = 3.7 × 10⁻⁵, hypergeometric test) was present, supporting the notion that these two miRNA clusters were selected in part for similar functional roles (Fig. 4B).

To explore targets relevant to our pathway analysis results, we tested whether two previously well characterized proliferation-related genes contain functional mouse-specific target sites: Dedd2 (p = 0.16, FDR < 65%) and Lats2 (p = 0.23, FDR < 57%). Dedd2 is a well conserved inducer of apoptosis across a variety of cell types, known to associate with Caspase 8 and Caspase 10 via its Death Effector Domain (DED) (32). Lats2, a target of the miR-290-295 cluster (26) and its homologous cluster in human (33), participates in both cell cycle progression (34) and apoptosis (35), in part through a positive feedback loop with p53 (36). Both genes show acquisition of target sites for Sfmbt2 cluster miRNAs specifically in mouse with a mutation profile higher than surrounding 3' UTR regions (Fig. 4C). Reporters containing the 3' UTRs of these genes fused to luciferase were transfected into Dicer KO cells with and without miRNAs predicted to have gained target sites in mice. Of the two potential target sites acquired by Lats2, the miR-466f-3p site showed an approximately 25% repression relative to a site mutant construct, while the miR-669-3p site appeared to be nonfunctional (Fig. 4D). In the case of Dedd2, miR-297a led to an approximately 33% repression (Fig. 4D). Thus, mouse-specific sequence variations in the 3' UTRs of these genes gave rise to functional target sites. We additionally tested Caspase 8, a recently described miR-295 target with partially homoduped target sites (28) and observed an 80% repression in response to transfection of miR-467a (Fig. 4D), indicating that these miRNAs can exert strong as well as moderate repression of targets.

**Discussion**

Here we present a previously undescribed method of studying species-specific changes in miRNA-target relationships and apply it to uncover the genome-wide response to miRNA acquisition.
In contrast to many current methods of target prediction, which detect conserved relationships (37–39), our method emphasizes relative divergence to identify positive selection for site gain and loss. Our results indicate that introduction of a novel miRNA cluster can be associated with genome-wide adaptation.

The detection of these genome-wide responses was likely enabled by a number of underlying factors. First, the recency of the cluster expansion and subsequent time window of selection may have aided our ability to observe these responses. In general, evolutionary models of positive selection are likely to have a biphasic mutation profile: Early in the evolutionary window, one would expect an increase in mutation rate as favorable changes are fixed, while afterward a subsequent decrease in mutation rate should be observed as these mutations are under pressure to be maintained. The positive selection signal observed here therefore suggests that the time frame of positive selection was short enough to prevent the second conservation phase from overshadowing the early rate increase (see SI Appendix).

In addition to the timing of the selection, the nature of the sites examined may have affected the signals we observed using this method. In general, the statistical significance of signals described here for species-specific miRNAs is less dramatic than signals observed for conserved miRNAs. This difference is likely a consequence of the relatively greater time over which selective pressure has acted on targets of conserved miRNAs, many of which have been conserved quite extensively across mammals and more broadly in vertebrates. Analysis of human-specific miRNAs failed to show a significant enrichment for positively selected targets, which likely results from the differing quality of miRNA annotation in human. Stringent criteria have been used to define mouse miRNAs, unlike the human annotations. As a result the human set contains roughly 300 more miRNAs, some of which may be spurious and therefore reduce the statistical power of this test. Even within the mouse-specific targets, these signals showed some variation, with more significant changes being detected for 7mer-M8 seed matches as compared to 7mer-A1 sites. This result is consistent with the generally stronger activity and improved ability of 7mer-M8 sites to predict authentic targets (15, 40).

While many of the identified target interactions are likely to be novel, a subset may functionally overlap with those of the miR-290-295 cluster which promotes proliferation (26) and survival (28). In some cases, single 3′UTRs may be targeted by multiple cluster members, as was observed for Lats2 (a target of miR-467a and miR-466f-3p), a gene known to oppose proliferation and growth (32, 34, 35). Because of the less than twofold repression observed by luciferase assay for the targets Lats2 and Dedd2, we favor the hypothesis that these genes are part of a much larger network of targets that together mediate the phenotypic consequences of these miRNAs. Importantly, however, these examples do validate the notion that a subset of relatively rapid mouse-specific mutations has resulted in functional targeting.

Inhibition of Lats2 and our other validated target, Dedd2, fits well with the position of the Sfmbt2 cluster in a paternally expressed placental gene (20), because such genes are thought to be commonly involved in redistributing resources from mother to offspring under the parental sex conflict model (41). In this model, the father has an evolutionary incentive to promote fetal growth at the expense of maternal resources as only the fetus shares half of his genetic material. On the other hand, the mother has a mixed incentive because her reproductive success is tied to her ability to generate more offspring in the future, all of whom will share half her genome (41, 42). In support of this notion, numerous genes with proprogrowth properties such as Igf2 tend to be exclusively expressed by the paternal allele, while growth-inhibiting genes such as Igf2r are exclusively maternally expressed (43). Consistent with this idea, paternal duplication of proximal chromosome 2 (which includes the Sfmbt2 gene) results in placental growth enhancement, whereas maternal disomy results in fetal and placental growth reduction (44). Our data suggest that some of these growth effects may result from contributions of the miRNA cluster in addition to the coding gene. Given the frequently paired relationships between maternally and paternally imprinted...
genes, the identification of a maternally expressed gene that counteracts the role of this cluster would not be surprising.

The Sfmb2 cluster expansion in mouse may additionally provide a useful model for other species- and lineage-specific miRNA expansions. For instance, in humans, an analogous primate-specific miRNA expansion is present on Chromosome 19 (C19MC), with a total of 46 miRNAs within a 100-kb interval being processed from the highly repetitive intron of a maternally imprinted placental noncoding RNA (45, 46). In addition to a being processed from the highly repetitive intron of a maternally mate-specific miRNA expansion is present on Chromosome 19 provide a useful model for other species- and lineage-specific

Zheng et al. PNAS

Materials and Methods
Oligos and siRNAs used in all the experiments can be found in SI Methods in SI Appendix. Microarray data was obtained from and processed according to Zheng et al. (28). Human, dog, and horse mature miRNA sequences were obtained from miBase Release 15 (50). Mouse mature miRNA sequence data were obtained from Chiang et al. (19). Aligned human, mouse, and dog 3’ UTRs were obtained from TargetScan 5.1 (15, 16, 37). For detailed methods, see SI Methods in SI Appendix.

ACKNOWLEDGMENTS. We thank Joel Neilson for reading of the manuscript and the Sharp and Burbage labs for helpful advice and discussions. A.R. was funded by a Fannie and John Hertz Foundation Fellowship. This work was supported by National Institutes of Health grants R01-GM34277, National Cancer Institute (NCI) Grant P01-CA42063, and the NCI Cancer Center Support (core) Grant P30-CA14051. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.