

RESEARCH ARTICLE

Evolutionary dynamics of microRNA target sites across vertebrate evolution

Alfred Simkin^{1,2*}, Rene Geissler¹, Alexa B. R. McIntyre^{1,3}, Andrew Grimson^{1*}

1 Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America, **2** Department of Biology, Elon University, Elon, North Carolina, United States of America, **3** Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York, United States of America

* asimkin@elon.edu (AS); agrimson@cornell.edu (AG)



Abstract

MicroRNAs (miRNAs) control the abundance of the majority of the vertebrate transcriptome. The recognition sequences, or target sites, for bilaterian miRNAs are found predominantly in the 3' untranslated regions (3'UTRs) of mRNAs, and are amongst the most highly conserved motifs within 3'UTRs. However, little is known regarding the evolutionary pressures that lead to loss and gain of such target sites. Here, we quantify the selective pressures that act upon miRNA target sites. Notably, selective pressure extends beyond deeply conserved binding sites to those that have undergone recent substitutions. Our approach reveals that even amongst ancient animal miRNAs, which exert the strongest selective pressures on 3'UTR sequences, there are striking differences in patterns of target site evolution between miRNAs. Considering only ancient animal miRNAs, we find three distinct miRNA groups, each exhibiting characteristic rates of target site gain and loss during mammalian evolution. The first group both loses and gains sites rarely. The second group shows selection only against site loss, with site gains occurring at a neutral rate, whereas the third loses and gains sites at neutral or above expected rates. Furthermore, mutations that alter the strength of existing target sites are disfavored. Applying our approach to individual transcripts reveals variation in the distribution of selective pressure across the transcriptome and between miRNAs, ranging from strong selection acting on a small subset of targets of some miRNAs, to weak selection on many targets for other miRNAs. miR-20 and miR-30, and many other miRNAs, exhibit broad, deeply conserved targeting, while several other comparably ancient miRNAs show a lack of selective constraint, and a small number, including mir-146, exhibit evidence of rapidly evolving target sites. Our approach adds valuable perspective on the evolution of miRNAs and their targets, and can also be applied to characterize other 3'UTR regulatory motifs.

OPEN ACCESS

Citation: Simkin A, Geissler R, McIntyre ABR, Grimson A (2020) Evolutionary dynamics of microRNA target sites across vertebrate evolution. *PLoS Genet* 16(2): e1008285. <https://doi.org/10.1371/journal.pgen.1008285>

Editor: Jianzhi Zhang, University of Michigan, UNITED STATES

Received: July 1, 2019

Accepted: January 2, 2020

Published: February 3, 2020

Copyright: © 2020 Simkin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by grant R01GM105668 from NIH to A.G. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Author summary

Gene regulation typically involves two components; the first is a regulator, and the second is a target that the regulator acts on. Examining the evolution of both components allows

us to make inferences regarding how gene regulation originated and how it is changing over time. One gene regulatory system, which is widespread in mammals and other animals, relies upon regulators known as microRNAs (miRNAs), which recognize target sites within mRNAs, leading to post-transcriptional repression of the targeted mRNAs. This system has been studied extensively from the perspective of the evolution of the miRNA regulators. Moreover, a subset of the target sites themselves are known to be deeply conserved. Here, we have systematically examined the rate at which target sites for individual miRNAs are created and destroyed across mammalian evolution. We find that mutations that strengthen or weaken existing target sites are strongly disfavored. For ancient microRNAs, even recently evolved target sites are under strong selective constraint, and as miRNAs age, they tend to initially experience selection only against loss of existing target sites, and later accumulate strong selection against gain of novel target sites.

Introduction

A fundamental property of biological systems is the regulation of gene expression. Diverse systems have evolved to generate controlled gene expression patterns across cell types and in response to different environments. As organismal complexity increases, the complexity of regulatory networks and mechanisms appears to have increased at a rate that is greater than any increase in the number of genes [1–3]. Thus, understanding the evolutionary pressures acting on different regulatory pathways is central to our understanding of the biology of complex organisms.

A major driver of evolutionary change is differential control of common gene complements. For example, humans and chimpanzees share 97% of their genomes [4], and yet exhibit strikingly divergent anatomical and behavioral traits. Primates, nematodes, and flies have similar numbers of protein-coding genes, many of which are well conserved in their primary sequence [5], and yet the three clades exhibit marked phenotypic differences. Different species of cichlid fish that shared a common ancestor within the past one million years display a remarkably diversified set of morphological and behavioral traits, despite virtually identical genomes [6]. Changes in gene regulatory networks, including those that appear modest, underlie many of the most striking phenotypic and developmental differences between life forms, and the study of these changes is essential to a complete understanding of the evolution of biological systems.

Regulatory networks are mediated by *trans* acting molecules (regulators) that interact with target genes or transcripts by way of *cis* recognition sites found on the targeted gene or gene product. Changes in gene regulatory networks can thus occur by two main mechanisms: either the *trans* factor changes, thus altering interactions with cognate *cis*-acting targets, or individual *cis* targets change to favor greater or lesser recognition by a static *trans*-factor, such that the targets of that *trans* factor shift over time. While changes in *trans*-acting factors can quickly change entire gene regulatory networks, gradual shifts in *cis*-acting targets may lead to significant changes in the function of a *trans*-acting factor without altering *trans*-factor sequence or the deep conservation of some *cis*-acting sites. Studying the rate at which *cis*-acting sites are lost and created is thus of comparable importance to studies of the rate at which *trans*-acting regulatory sequences change their binding affinities.

The evolution of gene regulatory systems is understood best for transcriptional regulation, which is mediated by *trans*-acting transcription factors and their corresponding *cis*-acting sites in DNA. Transcription factor affinities for their target sequences tend to be highly conserved [7], yet there are also examples of factors whose binding affinities change relatively rapidly

over time [8–10]. Some transcription factors are extremely conserved [7], while others are functional despite their recent evolutionary origins [11]. A similar process can be observed for *cis*-acting target sites; the rate at which conserved transcription factors acquire novel target sites or lose existing target sites is referred to as transcription factor binding site (TFBS) turnover [12]. Numerous approaches have been developed to model TFBS turnover, and have made significant contributions to our understanding of regulatory networks [13–15]. It is notable that some ancient transcription factors exhibit high levels of TFBS turnover [16], indicating that these deeply conserved transcription factors gain and lose targets relatively frequently. This finding demonstrates that the evolutionary age of *trans* regulators and their respective populations of *cis* targets are not always well correlated.

The approaches used in the study of transcriptional regulation by transcription factors can be applied to other regulatory systems. MicroRNAs (miRNAs) are a class of small regulatory RNAs found in almost all animal species, and many other organisms [17, 18]. In bilaterian animals, miRNAs promote mRNA degradation by recognizing ~6–8 base pair target sites, which are typically found within the 3' untranslated region (3'UTR) of mRNAs [19]. Thus, miRNAs represent a system of post-transcriptional gene regulation that largely follows the *trans-cis* paradigms established for transcription factors.

Following the discovery of miRNAs, the basic evolutionary characteristics of these regulatory molecules were studied extensively. Researchers discovered that certain miRNAs and a subset of their targets are conserved across deeply diverged phyla [20, 21], while others appear to be genus and even species specific [22, 23]. Using techniques similar to those applied to the study of transcription factors, models have been developed to track evolutionary changes in miRNAs [24–27]. By documenting the origins and decay of novel miRNAs, evolutionary studies of changes in these *trans*-acting factors have revealed the birth and death of regulatory sequences that impact hundreds of target genes at once [28–30].

The acquisition and destruction of *cis*-acting target sites of statically defined miRNAs (miRNA target turnover) has also been studied to some extent. Early studies revealed that some conserved miRNAs have deeply conserved targets and that the target sites of miRNAs tend to be themselves deeply conserved [31–35]. More recent studies suggest that target site evolution is influenced by the number of miRNA paralogs within a miRNA family [36]. Analyses of conserved miRNA target sites have been used as a metric of functionality to estimate that the majority of human protein coding genes harbor functional target sites for miRNAs in their 3'UTRs [37]. Population genetics models have been used to demonstrate that a fraction of miRNA target sites are likely to be under purifying selection [20], and in cichlid fish some studies have found that miRNA target sites appear to be under positive selection [38], although this conclusion has been contested [39].

While previous studies have demonstrated that miRNAs and their targets are under strong selective constraint, to date there are few studies [40–44] that have extended a study of turnover rate dynamics to miRNA target sites. Despite the demonstrated insights gained through examining turnover rates for transcription factor target sites [13–15], we know of few methods that have systematically examined turnover dynamics of the targets of individual miRNAs. Here, we have developed an approach that is capable of determining rates of miRNA target site turnover for individual miRNAs, and of quantifying the relative strength of selection acting on individual targets. We use this method both to characterize the evolutionary dynamics of the targets of specific miRNAs, and to further develop generalizable principles regarding the strength, prevalence, and distribution of natural selection acting on the targets of miRNAs. Our results identify remarkably different patterns of target site gain and loss between different groups of miRNAs. We further show that many of the specific patterns of miRNA target site evolution we find in vertebrates are also observed in insect phylogenies.

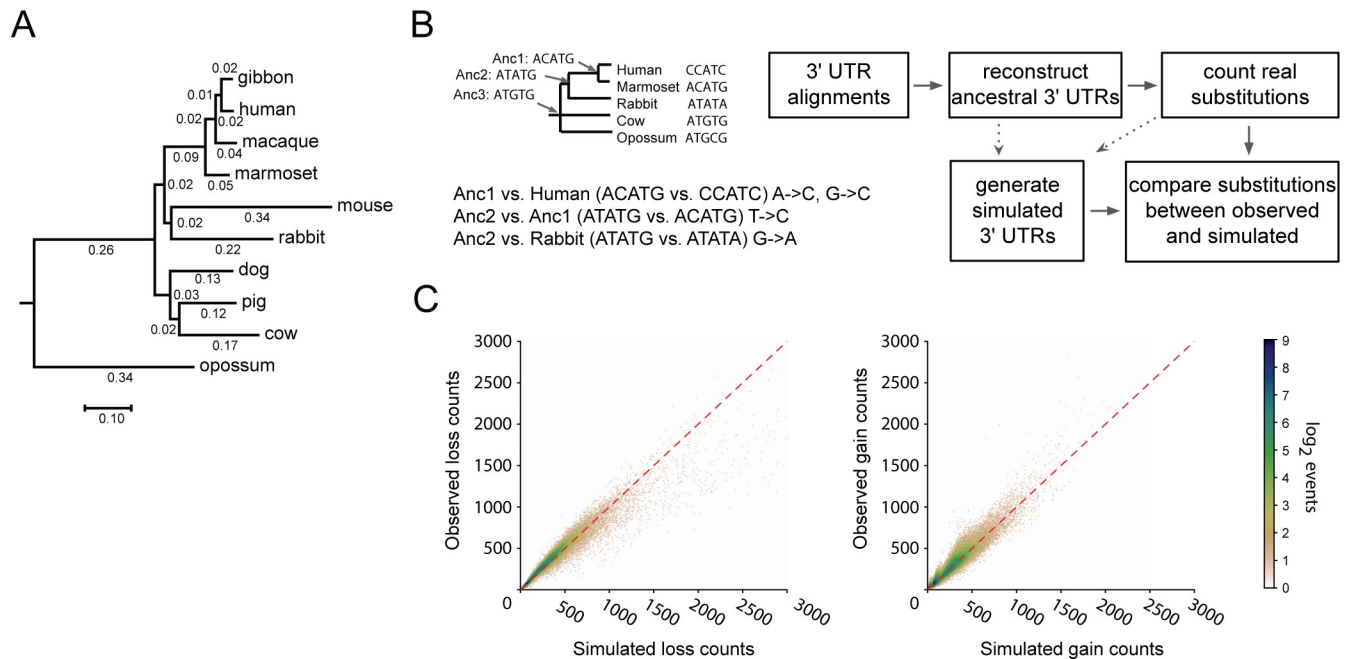


Fig 1. Simulated 3' UTR sequences recapitulate patterns of sequence evolution observed for sequence motifs. **A)** Mammalian phylogeny used in this study. Branch lengths are measured in substitutions per nucleotide site, as inferred from the UCSC hg38 100way whole genome syntenic alignment. Scale bar (and numbers) represent inferred number of substitutions per site of the original 100 way syntenic alignment. **B)** Schematic showing how ancestral nodes of a phylogeny are generated, and example comparisons between ancestral nodes and their direct descendants (left). Flowchart of experimental pipeline (right). **C)** Comparison of losses (left) and gains (right) of each possible 8mer across 3'UTR sequences, summed across species detailed in panel A, compared between simulated (x-axis; values are mean of 100 simulations) and observed (y-axis). Count densities (see log₂ scalebar), represent the number of points within a distance of 3 from the point in question, and have a pseudocount of 1 added to avoid taking the log of 0. Dashed red line indicates an idealized x = y relationship between simulated and observed counts.

<https://doi.org/10.1371/journal.pgen.1008285.g001>

Results

Simulating microRNA target site evolution

To model evolutionary changes in miRNA target sites, we developed a simulation-based approach to measure expected rates of nucleotide substitution events in the absence of selection, with the goal of using this as a benchmark to compare against the observed evolution of miRNA target sites. Among the species represented on the UCSC genome browser's 100 way multiz genome syntenic alignment [45], 20 represented mammals present in miRBase [22] that had at least 80% of their sequenced genome assigned to chromosomes. From these species, we selected ten representative mammalian species, chosen to be evenly distributed across the mammalian phylogeny, with genome sequences in at least their second revision, and with human and opossum representing the most distantly related species (Fig 1A). We reconstructed ancestral sequences using DNAML [46], which we used to estimate two parameters. First, by comparing inferred ancestral 8mer sequences to descendant 8mer sequences, we were able to directly infer changes in miRNA binding sites in the presence of natural selection. Second, by recording the substitution frequencies for every nucleotide given both of its flanking nucleotides, we were able to infer nucleotide substitution probabilities, which enabled us to create simulated sequences with randomly placed substitutions that account for the impact of flanking nucleotides on sequence evolution (Fig 1B). The simulated dataset, therefore, was comprised of 3'UTRs with randomly placed substitutions, and served as a proxy for neutrally evolving 3'UTRs. We compared this simulated neutrally evolving dataset against observed real 3'UTR sequences, and their DNAML inferred ancestors. This approach allowed us to compare

miRNA target site losses and gains in the observed dataset to those seen in sets of simulated neutrally evolving sequence.

The primary determinant by which miRNAs recognize their target sites is through base pairing between the 5' end of the miRNA (seed region) and a short 6–8 base pair target site [19]. Therefore, we began by examining our observed and simulated datasets with regard to 8 nucleotide motifs (8mers) to determine whether our simulations served as an effective model of neutral sequence change in 3'UTRs. For each 8mer, we counted the number of times that it was disrupted by point substitutions and the total number of times that the 8mer arose through point substitution, which we refer to as loss and gain rates, respectively. This examination was performed for each of the 65,536 possible 8mers, and repeated using 100 independent sets of simulated 3'UTRs across the ten species. When we compared the rates of loss and gain of 8mers between our set of simulated 3'UTRs and observed 3'UTRs, we found that the mean of our simulated rates was highly correlated with the observed rates (Pearson rank correlation, $R^2 = 0.93$ and 0.91 , for losses and gains, respectively, $P < 10^{-50}$; Fig 1C). Thus, with high consistency, 8mers that are predicted to undergo rapid substitution, as modeled in the simulated 3'UTRs, correspond to those undergoing the most rapid substitution rates in reality, and vice versa. We concluded that our model accurately captures substitution rates for the majority of 8mers, and that most of the fluctuation in observed 8mer substitution rates can be accurately predicted by modeling variations in the underlying substitution rates of constituent nucleotides.

Selection against loss and gain of miRNA target sites of ancient miRNAs

Having established that most 8mers in the observed dataset are lost and gained at rates that are equivalent to their simulated counterparts, consistent with their evolution under neutrality, we next sought to examine the loss and gain rates of 8mers corresponding to miRNA target sites. We concentrated on 8mer miRNA target sites rather than smaller sites, because as a group, 8mer sites exhibit the most pronounced evolutionary signatures in animals [19], presumably reflecting their increased efficacy [47]. We examined first the target sites of the 75 most deeply conserved miRNAs: those miRNAs whose 8mer 'seed' region (the region known to base pair directly with 3'UTRs) sequences are identical in human, mouse, and zebrafish miRNAs. We refer to these miRNAs in the remainder of this text as 'deeply conserved'. For each 8mer target site, we compared the observed loss or gain count relative to the mean of the simulated values, recording the number of standard deviations that the observed values deviated from the simulated. We then recorded the ranks of these standard deviation comparisons relative to all other 8mers (S1 and S2 Tables). We found that the target sites corresponding to the majority of the deeply conserved miRNAs (Fig 2A, in red) were lost far less frequently than expected absent selection, as judged by a comparison to all other 8mers (Fig 2A, in blue; Wilcoxon Rank test; $P = 2.4 \times 10^{-35}$; S1 Table). These results confirm that our technique has power to detect previously reported strong selection against loss of existing miRNA target sites [48]. It is important to note that our approach to determining selection against target site loss is distinct from previous approaches. Conventional approaches [35, 37, 49] capture selection that maintains conserved sites, examining the lack of change that occurs on the modest subset of sites that are under selective maintenance, a concept that contributes to miRNA target site prediction algorithms. Such approaches also often use several control kmers as a proxy for neutrally evolving versions of a kmer of interest. Our approach considers the observed rates at which all target sites—not just those that are conserved—are changing across an entire phylogeny, and compares these rates against explicitly modeled expected rates of change for that same kmer motif.

In addition to characterizing selection acting against the loss of existing binding sites, our approach also allows us to determine whether novel target sites are gained at neutral rates, an

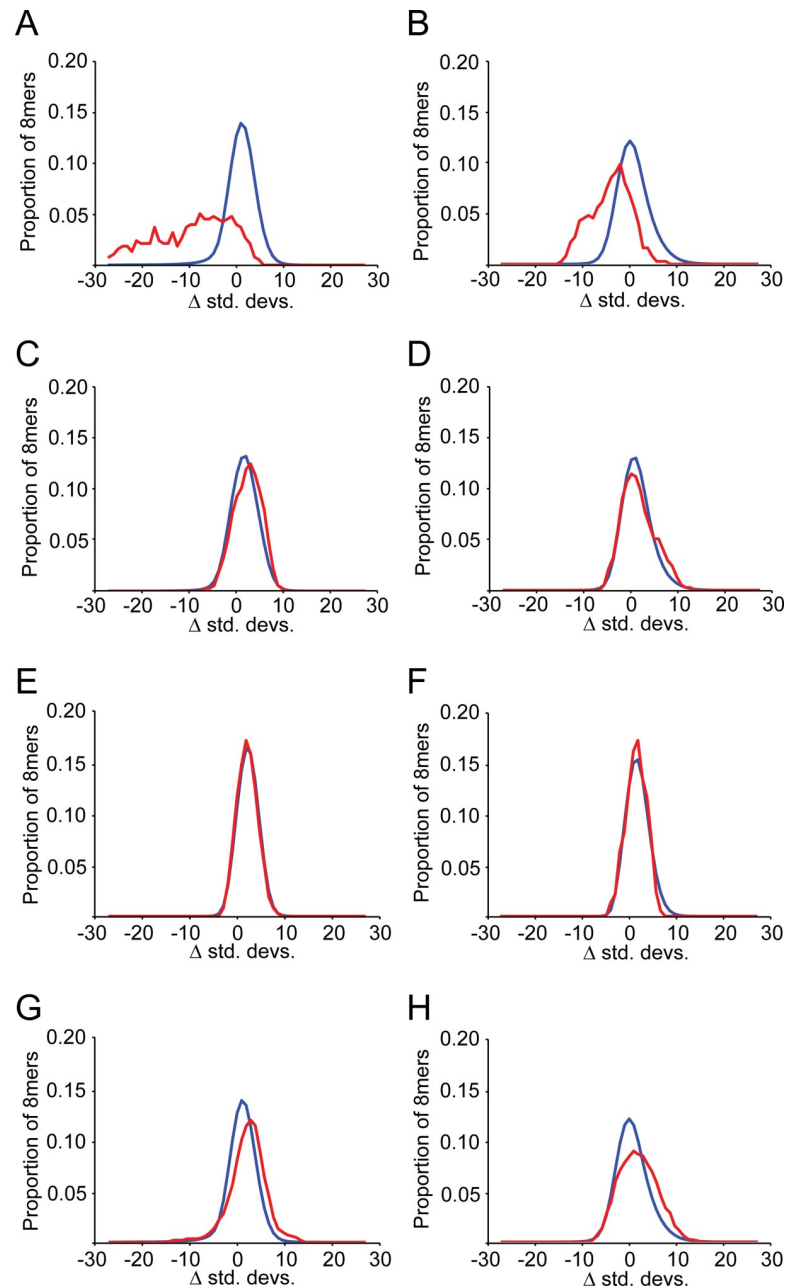


Fig 2. Target sites of deeply conserved miRNAs are lost and gained rarely in 3'UTRs. Each panel depicts the proportion of all 8mers (y-axis) that fall at a given number of standard deviations above or below the mean of simulations (x-axis). All 65,536 8mers are plotted in blue, and 8mer miRNA target sites in red. (A, B) Analysis of miRNA target site loss and gain events (respectively) in 3'UTRs, for 75 deeply conserved vertebrate miRNAs. (C, D) Analysis of miRNA target site loss and gain events (respectively) in intronic sequence, for deeply conserved miRNAs. (E, F) Analysis of miRNA target site loss and gain events (respectively) in shuffled 3'UTRs, for deeply conserved miRNAs. (G, H) Analysis of miRNA target site loss and gain events (respectively) in 3'UTRs, for moderately conserved miRNAs.

<https://doi.org/10.1371/journal.pgen.1008285.g002>

important additional aspect of miRNA target site evolution. We observed strong selection against the gain of binding sites for deeply conserved miRNAs (Fig 2B, in red) relative to all other 8mers (Fig 2B, in blue; Wilcoxon Rank test; $P = 3.1 \times 10^{-18}$; S2 Table). Thus, gain of target

sites for deeply conserved miRNAs is sufficiently deleterious to be observable across all 3'UTRs, despite the fact that many mRNAs are not co-expressed with their cognate miRNAs [31, 50], and therefore expected to evolve neutrally. It is worth noting that deviations in observed gain rates, compared to simulations, were significantly less pronounced than loss rates ($P = 0.017$), indicating that for deeply conserved miRNAs, creating novel targets appears to be somewhat less deleterious than destroying existing targets.

To verify the specificity of our results, we examined patterns of loss and gain of miRNA target sites in a neutrally evolving region of the genome that does not participate in miRNA-mediated regulation. MiRNAs are active predominantly in 3'UTRs; we therefore chose to examine the central regions of introns as a negative control, reasoning that by omitting the canonical splice sites at intron borders, the remaining sequence is predominantly evolving at a near-neutral rate [51] and is unlikely to participate in miRNA targeting. In this intronic dataset, we observed that target sites for deeply conserved miRNAs lose and gain miRNA target sites at approximately neutral or slightly above neutral rates (Fig 2C and 2D, S3 and S4 Tables, Wilcoxon Rank tests of 0.007 and 0.349, respectively). As an additional control, we generated shuffled 3'UTR sequences that maintain dinucleotide frequencies, and repeated our analyses of target site loss and gain. As expected, there is no signature of selection found for target sites corresponding to deeply conserved miRNAs in these artificial 3'UTRs (Fig 2E and 2F). Because losses and gains of miRNA target sites are only constrained in 3'UTRs, we conclude that our model has strong power to identify natural selection in motifs corresponding to miRNA target sites.

Returning to our 3'UTR dataset, we also examined target sites corresponding to less deeply conserved miRNAs, defined as those present both in humans and mice, but absent from zebrafish, which we refer to as moderately conserved miRNAs. Importantly, deeply conserved and moderately conserved miRNAs are present in all mammalian species for which detailed experimental miRNA annotations [22] are available, and thus are likely present throughout the entire mammalian phylogeny (Fig 1A) used to examine target site evolution. When considered as a set, these moderately conserved miRNAs showed evidence of marginally faster than expected target site turnover (Fig 2G and 2H, S5 and S6 Tables, Wilcoxon Rank tests of 3.2×10^{-14} and 6.0×10^{-9} , respectively). That is, rates of target site loss and gain for moderately conserved miRNAs are above those of most 8mers. It should be noted that more miRNAs were defined as moderately conserved than were defined as deeply conserved (299 versus 75) affording an increase in power to detect subtle selection acting on moderately conserved miRNAs as a group over deeply conserved miRNAs. Moderately conserved miRNAs therefore do not appear to exert detectable selective constraint against mutations within their targets when considered collectively, despite their origin prior to the common ancestor of our mammalian phylogeny.

To gain additional perspective on the extent to which deeply conserved miRNAs influence 3'UTR sequence evolution, we quantified the number of target site gains and losses observed, compared to their expected values estimated from simulated 3'UTRs. For miRNA target site losses, aggregated across the 75 deeply conserved miRNAs, we observed only 62% as many losses as expected (S1 Table). For site gains, we observed 78% as many events as expected (S2 Table). Collectively, this reduction in losses and gains corresponds to 1.9 fewer changes in miRNA target sites per 3'UTR than expected (1.4 fewer changes per 3'UTR than expected for losses, and 0.5 fewer gains per 3'UTR than expected, S1 and S2 Tables). We also investigated whether this deficit in losses and gains was related to the evolutionary distance between species. For pairs of species (from Fig 1A) that are most closely related, we observe a relatively small reduction in site losses and gains. This reduction in site loss and gains increases and is strongly correlated with evolutionary distance between the species pairs examined (S1 Fig;

green lines $R^2 = 0.88$, $P = 2.0 \times 10^{-7}$ and $R^2 = 0.87$, $P = 5.1 \times 10^{-7}$, for losses and gains, respectively), but only for branches shorter than approximately 0.25 substitutions per site, which represents the bulk of the phylogeny we considered. Thus, for deeply conserved miRNAs, there is an extensive evolutionary signature of target site constraint during mammalian evolution, which cumulatively amounts to tens of thousands of selective events.

Recently altered miRNA binding sites remain under purifying selection

It is well-established that miRNA target sites are among the most deeply conserved sequence motifs within mammalian 3'UTRs [33, 35, 37]. Yet, for deeply conserved miRNAs, their conserved target sites represent a small fraction of the sites capable of responding to their cognate miRNAs when co-expressed [31]. Importantly, the extent to which more recently evolved miRNA target sites contribute to global signatures of selection has not been thoroughly investigated. To determine whether the reduction in rates of target site gain and loss we observed (Fig 2A and 2B) is derived solely from selection acting on deeply conserved sites, we masked regions of the genome that are perfectly identical across all included species (species of Fig 1A) from our analysis, considering only the 8 nucleotide regions that have undergone substitutions. When we compared all 8mers containing at least one substitution against deeply conserved miRNAs whose target sites contain at least one substitution, we found that target sites of deeply conserved miRNAs were still undergoing loss and gain substitutions at rates significantly below other 8mers (Fig 3, losses and gains in parts A and B, respectively, Wilcoxon Rank test; $P = 3.0 \times 10^{-27}$ and 2.5×10^{-17} , respectively). This signature corresponds to thousands of constrained sites (S7 and S8 Tables). Importantly, these results indicate that purifying selection operates on miRNA target sites that are actively undergoing substitutions within mammals.

MicroRNA age correlates with selective constraint acting on target site turnover

Having established that target sites of deeply conserved miRNAs exhibit strong selection against target site turnover, and that target sites of moderately conserved miRNAs do not exhibit strong selection, we systematically investigated how miRNA age correlates with target site turnover. We used a 100 way vertebrate phylogeny [52], and intersected these 100 species with those found in miRBase 21, resulting in a phylogeny of 60 species. We estimated the age of each miRNA by measuring the summed branch length of species over which the same full-length mature miRNA sequence was present. We found that older miRNAs have lower loss

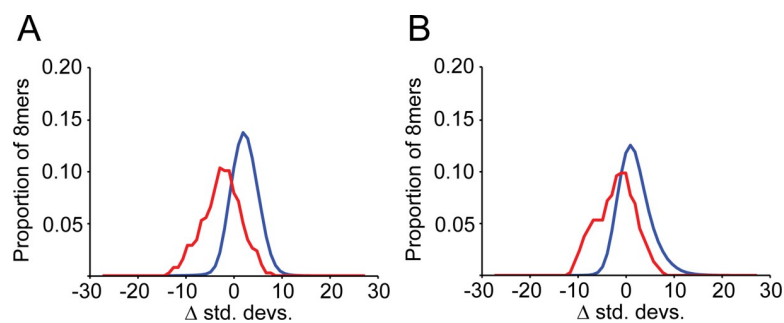


Fig 3. MiRNA target sites with recent substitutions are under strong selection. Analyses of the subset of target sites that have undergone at least one substitution within mammals are depicted (A, B; losses and gains, respectively), both for all 8mers (in blue) and for targets of deeply conserved miRNAs (in red), otherwise as described in Fig 2.

<https://doi.org/10.1371/journal.pgen.1008285.g003>

and gain rates of target sites than younger miRNAs (Fig 4A, Fig 4B, respectively; $P = 1.9 \times 10^{-28}$ and 4.8×10^{-16}), although the degrees of correlation were modest ($R^2 = 0.35$ and 0.21 , respectively). Because more ancient miRNAs tend to exhibit more constrained rates of target turnover, our observations are consistent with a model in which miRNAs begin with a relatively small subset of functional targets, and progressively acquire targets during evolution that then drive conservation of both targeting and primary miRNA sequence, consistent with results seen by others [42]. However, we also observed notable examples of miRNAs that are ancient, whose targets are gained and lost at approximately neutral rates. For example, miR-146 is deeply conserved, and yet its overall site turnover rate is more rapid than 74% of all other 8mers. Despite being ancient, such miRNAs appear to be rapidly evolving new target genes, as has been reported previously for the targets of some classes of deeply conserved transcription factors [16]. While our data provides direct evidence in support of the canonical view that conserved miRNAs generally demonstrate conserved targeting [48], our technique also identifies outlier miRNAs that are deeply conserved with rapidly evolving targets.

We next examined whether the miRNAs with the strongest selection against gain of deleterious target sites were also those with the strongest selection against loss of functional sites, focusing on our set of 75 deeply conserved miRNAs. We observed that many of the miRNAs with the strongest selection against gain of deleterious target sites were also those with the strongest selection against loss of existing target sites, and that there is a significant correlation between target site gain rates and target site loss rates (Fig 4C, $P = 3.1 \times 10^{-4}$). Nonetheless, many miRNAs showed limited association between loss and gain rates, decreasing the strength of the overall correlation ($R^2 = 0.16$). We classified each miRNA by considering a miRNA to have constrained shifts in targeting if its standard deviation score fell within the lowest 5% of all 8mers. Using this system to classify target site gain and loss rates as either 'constrained' or 'unconstrained' created four broad categories: constrained for both losses and gains (group 1), constrained for losses but not gains (group 2), constrained for neither losses nor gains (group 3), and constrained for gains but not losses (group 4). For comparison, we also repeated these analyses for our set of 299 moderately conserved miRNAs (Fig 4D). Examining panel 4C, within the 75 deeply conserved miRNAs alone, the first category (constrained loss and gain of targeting) was highly enriched (Fig 4C, 32 miRNAs observed versus 0 expected, binomial $P = 7.6 \times 10^{-63}$). We also found many deeply conserved miRNAs that exhibited strong selection against loss of existing target sites, but showed weak or no selection against gain of novel sites (group 2; 23 miRNAs versus 3.5 expected, binomial $P = 3.9 \times 10^{-13}$). We observed a group of deeply conserved miRNAs in the third category, such as miR-146, whose targeting was not conserved beyond the 5% threshold for either loss or gain of target sites. Consistent with our earlier analysis (Fig 2), we observed relatively few miRNAs within this third category (17 observed vs. 67 expected, binomial $P = 2.0 \times 10^{-43}$), which represents miRNAs with unconstrained target site loss and gain rates. Only three miRNAs showed constraint against gain without constraint against loss of target sites (group 4), a number of miRNAs which was not significantly different from that expected by chance (expected 3.5). Taken together, these results suggest the presence of three main groups of miRNAs (groups 1–3) within the deeply conserved miRNAs, with each group exhibiting pronounced differences in the evolution of their target site complement. There are those whose existing target sites are constrained against loss, and which avoid accumulation of novel target sites (Fig 4E and 4F, losses and gains, respectively; group 1, in red); there are those whose existing targeting is constrained, but with unconstrained accumulation of novel targets (group 2; in green); and there are those for which rates of target site loss and gain are not constrained (group 3, in black). The large number of miRNAs in group 3 is somewhat surprising given that these deeply conserved miRNAs are conserved in sequence across the span of vertebrates from human to mouse to zebrafish.

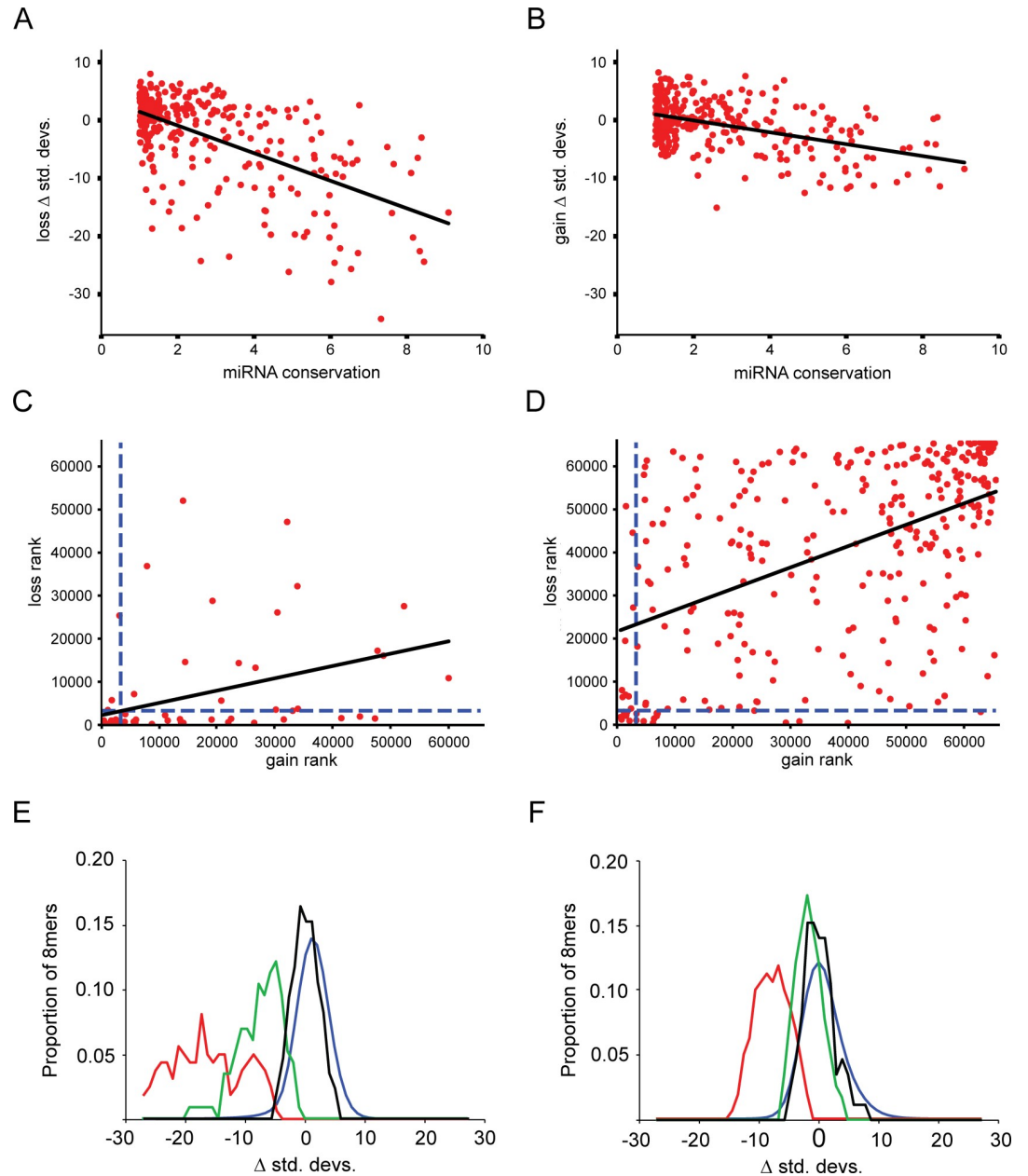


Fig 4. Selection against gain and loss of miRNA target sites correlates with miRNA age. The age of each miRNA, as measured by the summed branch length across which each full length miRNA is conserved (x-axis measured in genome-wide substitutions per site, larger numbers represent greater conservation), compared against the strength of selection acting on each target site (y-axis, standard deviations below the mean of simulated values), for losses and gains of sites (**A and B**, respectively). (**C, D**) miRNA gain ranks of individual miRNAs relative to all 8mers (x-axis, ranked by standard deviations above or below the mean of simulations) versus miRNA loss ranks relative to all 8mers (y-axis), analyzed for 75 deeply conserved miRNAs (**C**) and 299 moderately conserved miRNAs (**D**). miRNAs considered to be under negative selection are those below a rank of 3277 (5% of 65,536, dashed lines). (**E, F**) Skew of proportion of miRNA loss ranks (**E**) and gain ranks (**F**) at a given number of standard deviations from simulated values for 75 deeply conserved miRNAs, as described in **Fig 2**, separated into those in miRNA group 1 (constrained below 5% threshold for gains and losses in panel C; in red), group 2 (constrained for losses but not gains; in green), and group 3 (not constrained for gains nor losses; in black) as compared to the proportions of all 65,536 8mers (blue).

<https://doi.org/10.1371/journal.pgen.1008285.g004>

Selection against loss and gain of target sites of insect miRNAs

To gain additional perspective on the evolution of animal miRNAs and their target sites, we repeated our analyses on a *Drosophila* phylogeny including seven species (*melanogaster*, *simulans*, *sechellia*, *yakuba*, *erecta*, *biarmipes*, and *suzukii*; Fig 5A). In this dataset, we defined deeply conserved miRNAs as those present in *D. melanogaster*, *Triboleum castaneum*, and *Apis mellifera* (fly, beetle and bee, respectively), resulting in 43 deeply conserved insect miRNAs (S9 and S10 Tables). It should be noted that this definition of a 'deeply conserved' miRNA is somewhat more stringent than our earlier vertebrate definition, as flies, beetles, and bees are more diverged than humans, mice, and fish (*D. melanogaster*, *T. castaneum*, and *A. mellifera* have incurred a summed branch length of 3.8 substitutions per site since diverging from a common ancestor, whereas *H. sapiens*, *M. musculus*, and *D. rerio* have incurred a summed branch length of 2.6 substitutions per site since diverging from a common ancestor). Our analysis of insect miRNAs therefore contains a smaller set of more deeply conserved miRNAs than we used when examining deeply conserved animal miRNAs. Similar to our mammalian results, we found that deeply conserved insect miRNAs tend to lose and gain targeting at rates significantly below expected rates (Fig 5B and 5C, respectively, Wilcoxon Rank tests of 8.5×10^{-23} and 1.1×10^{-21} , respectively) We also found that the three groups of deeply conserved miRNAs we described in mammals also manifest in the *Drosophila* phylogeny (Fig 5D). Using ranks below 5% of all 8mers to indicate constrained target sites, 24/43 deeply conserved insect miRNAs exhibit reductions in both loss and gain of target sites (group 1, enrichment binomial $P = 1.9 \times 10^{-15}$). Nine of the *Drosophila* miRNAs were constrained for target site loss and unconstrained for target site gain (group 2, enrichment binomial $P = 1.6 \times 10^{-4}$). 9/43 highly conserved miRNAs were not strongly constrained against creation or destruction of binding sites and therefore appeared to be relatively promiscuous in their interactions with targets (group 3, depletion binomial $P = 9.7 \times 10^{-27}$), and only one highly conserved miRNA had strong selection against creation of novel sites, but weak selection against loss of existing sites (group 4, binomial $P = 0.39$). Thus, patterns of evolution of miRNA target sites in mammals are recapitulated in *Drosophila*.

Older miRNAs are broadly expressed and under stronger selection than younger, more narrowly expressed miRNAs

Following up on our results indicating that deeply conserved miRNAs fall into three major groups based on patterns of target site loss and gain, we next searched for parameters that might explain these differences. We examined miRNA and target expression data to determine whether differences between the efficacy of selection leading to creation or destruction of miRNA target sites correlates with differences in the expression patterns of these miRNAs. Using a miRNA expression atlas established for humans and rodents [53], we found that miRNAs with strong selection against both loss and gain of targeting (those in group 1) tend to be more broadly expressed (defined as expression at any level) than miRNAs with strong selection against loss of targeting and weak selection against gain of targeting (those in group 2, Wilcoxon rank sum test $P = 0.003$; S15 Table). miRNAs with weak selection against both gain of targeting and loss of targeting (in group 3) were generally expressed in fewer tissues than those in group 1 (S15 Table; group 1 vs. group 3 $P = 0.001$, group 2 vs. group 3 $P = 0.12$, Wilcoxon rank sum tests). miRNAs that are expressed broadly across the entire organism therefore appear to exhibit greater conservation of targeting than miRNAs that are narrowly expressed. However, we observed numerous exceptions to these trends; for example, miR-146 is broadly expressed and yet exhibits negligible constraints in gain and loss of target sites, while other miRNAs, such as miR-137, are narrowly expressed and yet exhibit strongly constrained

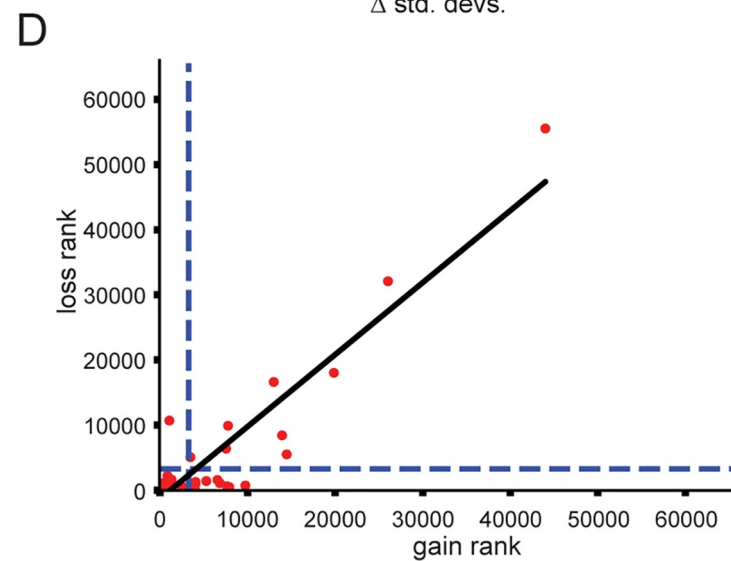
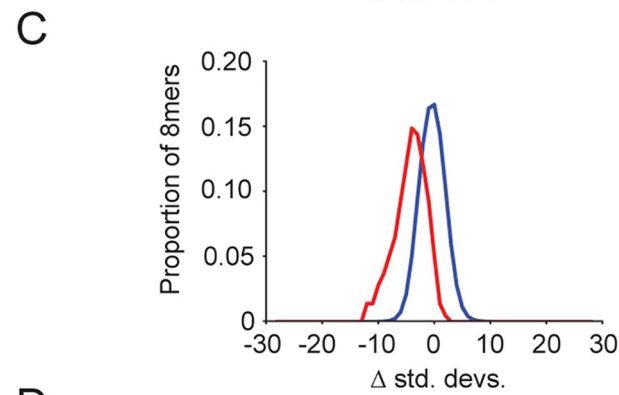
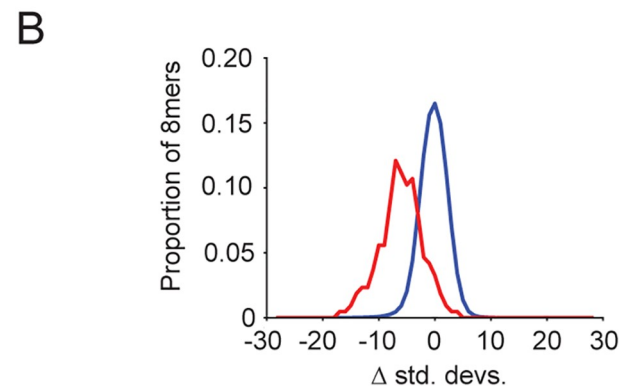
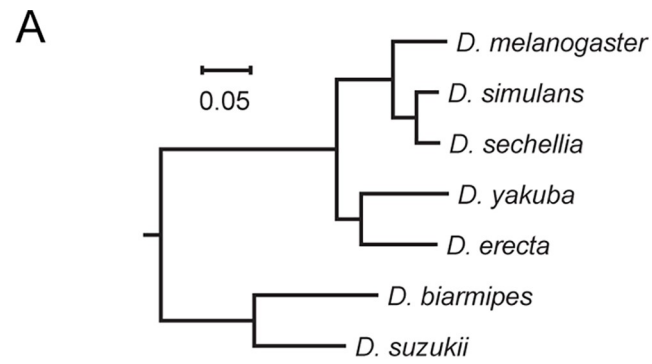


Fig 5. Patterns of selection observed in mammals are recapitulated in *Drosophila*. (A) Phylogeny of *Drosophila* species, as gathered from the UCSC dm6 multiz 15 way syntenic genome-wide alignment. Branch lengths are measured in units of inferred substitutions per site (relative to each ancestral sequence). (B, C) skews in proportion of 43 deeply conserved insect miRNAs a given number of standard deviations from the mean of simulated values, as described in Fig 2. (D) Individual gain and loss ranks of 43 insect miRNAs, as described in Fig 4C. Blue dotted lines represent a 5% threshold below which the turnover rates of the binding sites of a miRNA are considered 'constrained'.

<https://doi.org/10.1371/journal.pgen.1008285.g005>

patterns of target site loss and gain (S15 Table). We conclude that miRNA expression patterns are correlated with patterns of target site evolution, but there are unambiguous exceptions to the overall correlation, and therefore patterns of miRNA expression and target evolution are not likely to be deterministically linked.

Differences in patterns of miRNA target site evolution across the transcriptome

An outstanding question in the miRNA field concerns how many meaningful targets exist amongst the hundreds of possible target sites for a given miRNA. Possible answers to this question range from beliefs that each miRNA possesses a small handful of targets whose regulation is consequential [54, 55], to hypotheses that each miRNA regulates hundreds of targets [56]. Because many miRNAs are highly tissue specific [53], only a fraction of potential targets are co-expressed with their cognate miRNA. To gain additional perspective on this question, we examined the relative strength of selection acting on individual mRNAs for each miRNA. We began by analyzing each individual 3'UTR and comparing the number of observed substitution events against the mean of the simulated datasets for each possible 8mer. Some 3'UTRs incurred far fewer substitutions than expected, while other genes incurred far more. However, even in a randomly generated dataset, some points will fall far below the mean, and others far above. To control for this stochastic fluctuation, we repeated this process multiple times, choosing one of the simulations to compare against the mean of all others (Fig 6, purple lines; S1 Zipped File). This process revealed that for many of the deeply conserved miRNAs, the observed selective pressure against loss events is stronger than any of the simulated datasets. For some deeply conserved miRNAs, exemplified by let-7, we observed reduced rates of target site loss substitutions and reduced rates of target site gain substitutions across a large fraction of the 3'UTRs, indicating that almost all of these targets or potential targets are under selection (Fig 6A and 6B losses and gains, left and right, respectively). In contrast, other deeply conserved miRNAs, exemplified by miR-21, showed evidence of negative selection only for a small subset of the target space, suggesting functions that are restricted to a small portion of the possible set of regulated transcripts (Fig 6C and 6D, losses and gains, left and right, respectively), with the majority of the potential target set evolving at near neutral rates. Thus, the patterns of target site evolution for some miRNAs are consistent with broad selection acting across many targets, whereas for other miRNAs, selection is most apparent on a more restricted target set. Some miRNAs, exemplified by miR-122, showed very weak, if any, evidence of selection across most of the transcriptome (Fig 6E and 6F, losses and gains, left and right, respectively), while still other miRNAs showed subsets of highly constrained targets in addition to subsets of targets with weak evidence of rapid turnover of binding sites (Fig 6G and 6H, losses and gains, left and right, respectively). In this manner, our method demonstrates an ability to differentiate deeply conserved miRNAs that appear to operate on large subsets of genes from those that operate on small subsets of genes, and to differentiate between miRNAs that operate only in conserved processes, those that operate in both conserved and rapidly evolving processes, and those that appear to operate predominantly in rapidly evolving processes. For example, certain ancient miRNAs, including let-7, miR-9, and miR-1, exert

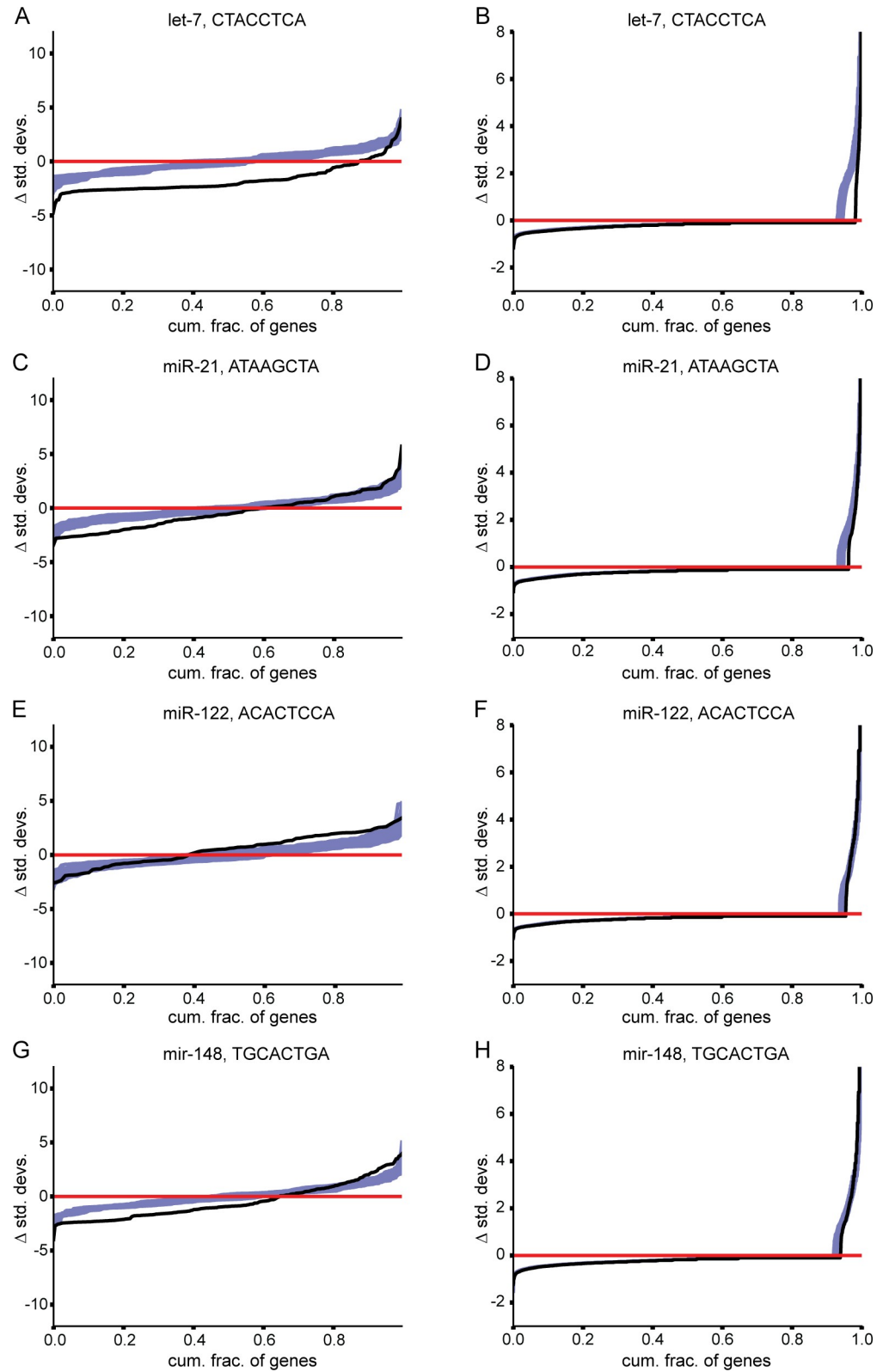


Fig 6. Target sites for individual miRNAs exhibit distinct patterns of selection. For each miRNA, all 3'UTRs that undergo substitutions that destroy or create an 8mer miRNA target site were analyzed. For each 3'UTR, the total number of such substitutions was evaluated relative to the mean of simulated substitutions for that 3'UTR, and the

number of standard deviations above or below the mean recorded (y-axis), ordering genes from those that fall the greatest number of standard deviations below the mean to those that fall the greatest number of standard deviations above the mean (x-axis). The black line represents scores for the observed dataset as compared to the mean of 99 simulated datasets. The purple lines represent scores for each of the 100 simulated datasets as compared to the mean of the remaining 99 simulations, and serve as a measure of expected stochastic fluctuations. Regions in which the black line is above all other lines represent an excess of substitution events (positive selection favoring substitutions), while regions in which the black line is below all other lines represent a depletion of substitution events (negative selection against substitutions). (A, C, E, G) Analysis of target site loss rates (left panels) for the 8mer target sites corresponding to the miRNAs let-7, mir-21, mir-122, and mir-148, respectively. (B, D, F, H) Analysis of target site gain rates (right panels), otherwise as for panels A, C, E, and G.

<https://doi.org/10.1371/journal.pgen.1008285.g006>

strong purifying selection on target sites across much of the transcriptome (Fig 6, S1 Zipped File), while other ancient miRNAs, including miR-21, appear to only exert purifying selection on a small subset of the transcriptome; in contrast, we also identified evidence of rapidly changing targets for conserved miRNAs such as miR-148. By analyzing the distribution of selection across genes, our approach is able to differentiate strong selection acting on a small number of target mRNAs from weak selection acting broadly across the transcriptome, facilitating a deeper understanding of the types of roles that individual deeply conserved miRNAs are likely to play.

Changes in miRNA target site efficacy are constrained

MicroRNA target sites exist with a range of efficacies. In general, the weakest sites are those with a 6 nucleotide match (6mers), created by pairing of miRNA nucleotides 2–7 with 3'UTR sequence. The most effective sites (8mers) supplement the core 6 nucleotide match with both an additional base pair to position 8 of the miRNA (m8) and an unpaired A (A1). We therefore focused primarily on 8mer sites, reasoning that signatures of selection would be most pronounced for the strongest class of target sites. There are also seven nucleotide target sites, typically intermediate in efficacy between 6mer and 8mer sites, which contain either the m8 or A1 nucleotide; 7mer-m8 sites tend to be more effective than 7mer-A1 sites [19]. Although the existence and relative efficacies of these different classes of miRNA target sites are well established, the relative strength of selection operating to convert site-types from one state to another has not been examined previously.

We used our simulated 3'UTR datasets as a benchmark to search for constraints on evolutionary changes between the four classes of miRNA target sites, examining those corresponding to the deeply conserved miRNAs only. We compared the frequency with which target sites convert between site types to the frequency with which other, background, 8mers convert between site types, in order to obtain estimates of the strength of selection operating on these events. We found that miRNA target site conversion events of almost every type tend to be significantly depleted when compared to equivalent events acting on background 8mers, which we assumed approximate neutral evolution (Fig 7, S2 Fig). Nearly every target site type demonstrates selective pressure against both strengthening existing binding sites (Fig 7A, beneath and to the left of the shaded diagonal) and against weakening existing binding sites (above and to the right of the shaded diagonal). Target sites of miRNAs in group 1 (those with constrained gain and loss rates; Fig 4) show constraint against every type of conversion (Fig 7B), miRNAs in group 2 (constrained against losses but not gains) show evidence of constraint against weakening target sites (above and to the right of the shaded diagonal) and weaker evidence of constraint against strengthening target sites (below and to the left of the shaded diagonal, Fig 7C), while target sites for miRNAs in group 3 (whose 8mer target sites are evolving neutrally) are converting into other site types at rates close to neutrality, as expected (Fig 7D). We note that some miRNA target site conversion events, denoted with an asterisk (Fig 7C and 7D), occurred

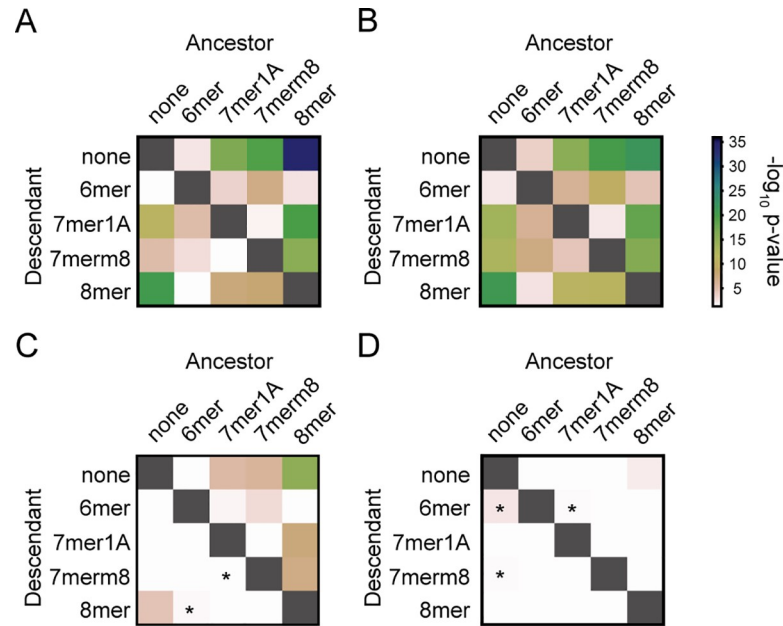


Fig 7. Interconversions that strengthen or weaken miRNA target sites are disfavored. Each panel calculates pairwise comparisons between the foreground and background rates at which the indicated miRNA target sites undergo substitutions (see [methods](#)). The ancestral target site type is represented along the x-axis, and the descendant along the y-axis. Substitutions that increase the average strength of a target site are below and to the left of the diagonal (grey boxes), while substitutions that decrease the average strength are above and to the right. Significance is color-coded (see legend) and plotted as $-\log_{10}(\text{P-value})$, using Wilcoxon rank-sum comparisons between the background (all 8mers) and foreground (miRNA sequences) rates. Insignificant P-values ($P > 0.05$) are color coded white, and significant P-values that favor strengthening or weakening of miRNA target sites (most of which are only nominally significant, and not significant with a multiple comparison correction) are marked with an asterisk. (A) Analysis of target site conversion events for all 75 deeply conserved vertebrate miRNAs (B, C, D) Analysis of target site conversion events for deeply conserved miRNAs with strong selection against both gain and loss events in their target sites (B; referred to as group 1 in the main text), deeply conserved miRNAs with strong selection against loss events but not gain events (C; group 2), and deeply conserved miRNAs with no strong selection against loss or gain events (D, group 3).

<https://doi.org/10.1371/journal.pgen.1008285.g007>

slightly more frequently than the background set of all 8mers, although these events were of marginal significance. Taken together, these results suggest that modifications to the strength of an existing miRNA target site are sufficiently consequential to impact the evolution of 3'UTR sequence, and that weakening of an existing site is generally more detrimental than strengthening a site (blocks are darker in upper right diagonals than lower left, Wilcoxon Rank p-value 5.9×10^{-9}). Because selection appears to act not only against creating novel deleterious sites but also against strengthening existing beneficial sites, we note that these results are most consistent with the hypothesis that many miRNA target sites exist to modulate or fine-tune gene expression rather than to maximally repress gene expression.

In addition to examining patterns of target site conversion, we also examined whether 8mer site loss and gain was impacted by the predicted efficacy of the target site. Multiple parameters impact target site efficacy but local AU content both exerts a relatively large effect and impacts many sites [19]. Thus, we examined whether 8mer target sites corresponding to deeply conserved miRNAs were lost and gained at different rates in regions of high and low AU content, which are predicted to be more effective and less effective sites, respectively [47]. Regions of high AU content appeared more constrained against site loss and gain (S3A Fig and S3B Fig, respectively) when compared against site loss and gain in regions of low content (S3C Fig and S3D Fig, respectively; results for individual miRNAs in S11–S14 Tables),

however, only the difference between high AU content gains and low AU content gains was significant ($P = 0.007$, similar loss comparison $P = 0.227$). In particular, gain of sites occurs at rates closer to the neutral rate for regions of 3'UTRs refractory to miRNA mediated repression. Loss of sites, however, did not significantly differ as a function of 3'UTR sequence context, perhaps reflecting the significant selection against weakening or strengthening existing sites (Fig 7).

Analysis of constrained 3'UTR 8mer motifs

The regulatory information within 3'UTRs extends beyond that mediated by miRNAs [57]. Therefore, we decided to systematically examine all 8mers for selective constraint in mutations that either create or destroy the 8mer. We examined the 1,000 most constrained 8mers (1.5% of all 8mers), ranked independently for losses and gains. As expected, 8mers that correspond to or overlap with miRNA target sites represent a large fraction of the most constrained 8mers. We also examined 8mers with overlap to Pumilio binding sites, cleavage and polyadenylation signals, AU-rich elements and Fox binding sites [58–60]. Cumulatively, these established 3'UTR regulatory elements represent 70% and 50.2%, respectively, of the 1,000 8mers with the most constrained losses and gains (Fig 8A and 8B, respectively). As a background set, we repeated this analysis on the 1,000 8mers with the least constrained losses and gains, which had only 6.9% and 9.1% of elements deriving from known motifs, respectively (Fig 8C and 8D, respectively). Thus, binding sites for established 3'UTR regulatory elements exhibit marked constraints in sequence changes that either create or destroy existing sites. It seems likely, therefore, that our approach could be used to identify novel regulatory 3'UTR motifs by searching for sequence motifs that are both rarely lost and rarely gained during evolution.

Discussion

We set out to undertake a comprehensive analysis of the evolutionary relationships between miRNAs and their targets. Previous work has demonstrated that conservation of a target site can be used to predict functional sites [21, 35, 37, 61, 62]. The existence of 'antitargets' of miRNAs, that is, 3'UTRs under selection to avoid containing miRNA target sites, has also indicated that selection acts to prevent the gain of detrimental target sites [31, 63]. Many additional studies have made valuable contributions to our understanding of the origins and evolution of miRNA target sites [20, 32, 42, 48, 64]. However, a systematic evolutionary analysis of all changes, that is, rates of gains and losses in individual target sites of individual miRNAs,

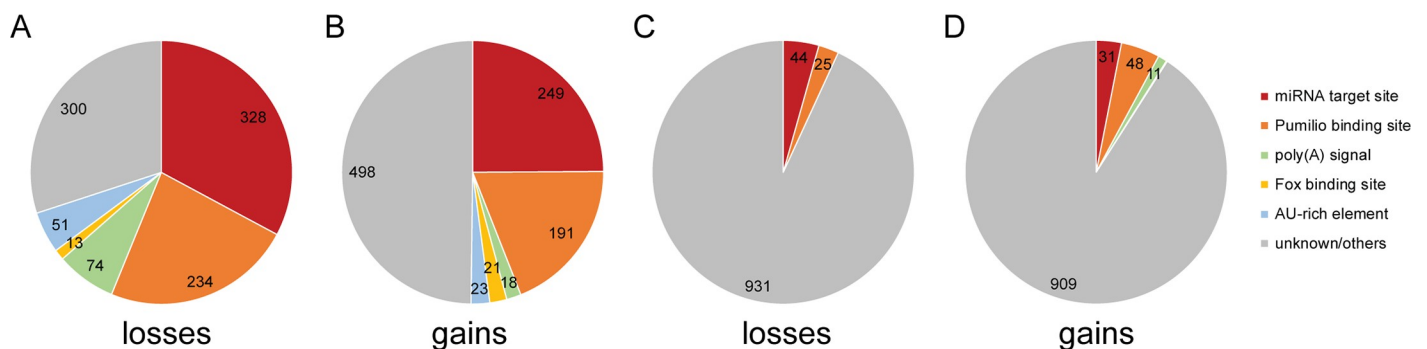


Fig 8. Analysis of rates of loss and gain of functional 3'UTR regulatory elements. (A, B) Fractions of the 1,000 most constrained 8mers corresponding to miRNA target sites, Pumilio response elements, poly(A) signal sequences, Fox protein binding sites, AU rich elements, and motifs of unknown function, for losses (A) and gains (B), analyzed as described in Fig 1. (C, D) Fractions of the 1,000 least constrained 8mers, otherwise as described in panels A and B.

<https://doi.org/10.1371/journal.pgen.1008285.g008>

constitutes an important and largely unexamined component of miRNA biology. The role of a miRNA is defined by the collection of genes that it targets. Thus, it is only by systematically studying changes in targeting that we can gain insights into the ways in which individual miRNAs change roles over evolutionary time. Our goal here was to construct a reliable background model of 3'UTR sequence change, and use this model to infer rates of miRNA target site gain and loss attributable to selection. This approach has allowed us to gain new insights into the changing roles of miRNAs during evolution, including at the level of individual miRNAs.

We find strong selection against loss of the target sites of deeply conserved miRNAs (Fig 2A), consistent with previous studies [37]. We also find strong selection against target site gain (Fig 2B); although this result has been hinted at previously [31], the extent of the signal is surprising, and comparable to the selection against losses. While deeply conserved miRNAs might be expected to have acquired deeply conserved targets, the effects of miRNAs on gene expression are subtle, and the creation of accidental target sites would typically exert modest effects on gene expression levels. The broad extent of the signal against target site gain indicates that for most deeply conserved miRNAs, even subtle reductions in gene expression across potential targets are strongly selected against. Importantly, when we repeated these analyses in *Drosophila* species, we observed strong selection against both gain and loss of miRNA target sites (Fig 5), suggesting that these patterns of selection acting on miRNA target site gain and loss are likely to represent diverse bilaterian lineages.

Patterns of target site loss and gain for individual miRNAs

In general, miRNAs with the slowest turnover rates of target sites are in good agreement with those described as having ancient functions. For example, the let-7 miRNA is known to have at least one target, lin-41, which has a target site that is maintained in organisms from nematodes to humans [32], and our analyses indicate that let-7 is highly constrained, both in terms of target site loss and gain across mammals. Similarly, the target sites for other well-studied and ancient miRNAs, such as miR-124 and miR-9, are also highly constrained. Interestingly, some of the miRNAs that stand out as having the slowest rates of change in targeting, such as miR-20 and miR-30 (S1 and S2 Tables) are not prominently described in the literature as having deeply conserved targeting. Both miR-20 and miR-30 are expressed in a wide range of tissues, and are amongst the most broadly expressed miRNAs (S15 Table). The success of our method in enriching for deeply conserved miRNAs with known deeply conserved functions and broad expression patterns of these miRNAs implies that miR-20 and miR-30 may also have deeply conserved functions. miR-30 has been shown to play a role in thermogenesis, which may form part of a conserved pathway among mammals, and appears to play important roles in adipogenesis and intestinal epithelial cell homeostasis [65–67], while miR-20 has roles in suppressing angiogenesis and is a member of a miRNA family that is essential in mice [68–70]. Thus, both of these miRNAs may represent promising candidates for the discovery of conserved and essential miRNA-mediated regulatory pathways among mammals, and potentially other bilaterian species.

In contrast to let-7 and many other ancient miRNAs, miR-146 is an unusual example of a deeply conserved miRNA whose target set is minimally constrained. These observations indicate that this miRNA is gaining novel targets and losing existing targets relatively rapidly, with neutral or potentially faster than neutral rates. Notably, miR-146 contributes to diverse immune system functions, including fever response, tumor suppression, T-cell homeostasis, and cytokine production in multiple immune cell lineages [71–75]. It is conceivable that miR-146 targets may be rapidly co-evolving with pathogens as part of an 'arms race' scenario that has frequently been seen in host-pathogen interactions [76]. Thus, miR-146 may represent a

promising candidate for the discovery of novel rapidly evolving (non-constrained) regulatory pathways that couple deeply conserved miRNAs to rapidly shifting target sites.

We also considered miRNAs beyond those we identified as deeply conserved, none of which exhibited evidence of strong selection against site loss or gain (Fig 2G and 2H), consistent with previous studies indicating a lack of conservation of target sites for such miRNAs [48]. The existence of so many miRNAs, including many that are broadly conserved across mammals, which exhibit minimal or no evidence of selective pressure on their target sites remains somewhat perplexing. The simplest interpretation of these results is that such miRNAs have a very restricted set of targets that are under selection, but that such targets are greatly outnumbered by those that are evolving at neutral rates. Alternative possibilities, which have already been established in the evolution of transcription factor binding site (TFBS) networks [8–10, 16] and posited for miRNA target site networks [38], include the existence of a set of targets that are evolving rapidly, but that are biologically consequential. In this regard, it is important to note that we detected selection against target site loss and gain for deeply conserved miRNAs even when we restricted our analysis to regions of 3'UTR sequence undergoing more rapid sequence change (Fig 3). This result indicates that even when only weakly conserved target sites are analyzed, deeply conserved miRNAs still exert significant selective pressure against sequence change, while less deeply conserved miRNAs show no detectable selection against losses or gains of target sites, even when more deeply conserved 3'UTR sequences are included. Thus, over a relatively short evolutionary timescale, selection acts on targets corresponding to ancient miRNAs, but does not act on less well-conserved miRNAs, even though such miRNAs pre-date the timescale considered. It is clear that the evolutionary ages of miRNAs plays a larger role in the strength of selection acting across the transcriptome than the age of individual target sites.

Patterns of target site evolution classify miRNAs into distinct groups

We note the existence of three markedly different groups of deeply conserved miRNAs, each with characteristic patterns of target site evolution. The first group contains miRNAs whose target sites are constrained for both losses and gains, and is the largest of the three (Fig 4C, 4E and 4F). MicroRNAs of this type are to be expected, as it is clear that selective pressures exist that maintain existing sites [37] and preclude the evolution of new sites [31], which would lead to reductions in losses and gains, respectively. The second group contains miRNAs which exhibit gains of sites at neutral rates but that are still constrained against loss of existing sites (Fig 4C, 4E and 4F). Such miRNAs are more likely to gain new consequential sites across mammalian evolution than the first group, for which gain of new sites is selected against and comparably rare. The existence of this second, large, group of miRNAs is unexpected, as it indicates that some miRNAs have deeply conserved existing targets, yet gain novel targets at approximately neutral rates. The patterns of target site evolution of the third and final group, which includes many ancient miRNAs, are the most surprising, in that they do not exhibit constraints on target site loss or gain. While this third group has parallels in transcription factor networks that are known to include deeply conserved transcription factors with rapidly shifting targets [16], this group is largely unanticipated within the miRNA literature, and represents miRNAs that appear to be deeply conserved while gaining and losing target sites at approximately neutral rates. It is important to acknowledge the existence of this final class, as it represents miRNAs for which conservation of individual target sites is less likely to represent a meaningful signal. In addition, the existence of this class of deeply conserved miRNAs presents a potential mechanism that would allow deeply conserved miRNAs to drive striking shifts in gene regulatory networks in a small amount of time without any change in miRNA

sequence. Amongst the most deeply conserved miRNAs, only three (of 75) exhibited selection against site gain while losing sites at a neutral rate. Thus, our results indicate that it is almost never the case that a deeply conserved miRNA exhibits strong deleterious effects as novel target sites accumulate when no selection is detectable against loss of existing targets. By classifying miRNAs into these groups, it is now possible to make reasonable conjectures as to which deeply conserved miRNAs are most likely to be involved in deeply conserved processes, which are expanding into novel functional niches, and which might potentially be associated with rapidly evolving species-specific functions.

The number of consequential targets of individual miRNAs has been a subject of some debate. While preferential conservation of miRNA target sites within 3'UTRs demonstrates that hundreds of genes are evolving in response to individual miRNAs [37, 48], others have shown that the phenotypes resulting from loss of function of miRNAs can often be recapitulated or ameliorated by altering the expression of a very small subset of target genes [77, 78]. Moreover, the narrow expression patterns observed for some miRNAs suggests highly tissue-specific functions [79]. Still other studies posit that miRNAs function partly to repress "leaky" or "noisy" gene expression of a large number of genes whose expression is undesirable in specific cells or a stages of development [80, 81]. Under this model, even highly tissue-specific miRNAs may still consequentially regulate large cohorts of targets as a consequence of selection; it is important to note that such targeting may be difficult to robustly test or detect using experimental methods [54]. Our method goes beyond traditional metrics of sequence conservation by examining individual targets of miRNAs for evidence of selective constraint against target site turnover (Fig 6). We have observed that transcriptome-wide or gene specific selection on target sites is not a binary trait that can be applied homogeneously to all deeply conserved miRNAs. While changes in target sites corresponding to some miRNAs exhibit selective pressure across much of the transcriptome, others exhibit strong selective pressure on a very small number of targets. Our approach may therefore be applied in novel ways to distinguish between miRNAs that serve as broad-spectrum regulators of the genome and those with very specific, restricted functions. Importantly, our results imply that different deeply conserved miRNAs have evolved under markedly different modes of selection, and suggest that no single theory of miRNA targeting is likely to be broadly applicable.

Identifying and characterizing selective pressures acting on novel motifs

Recent studies by multiple groups, using a variety of approaches, indicate that regulation by miRNAs is only a modest fraction of the regulatory information found in mammalian 3'UTRs [82–86]. While our method has been applied primarily to the characterization of selective pressures acting on miRNAs, we also analyzed selective pressures operating on all possible 8mer motifs, and successfully recovered motifs corresponding to RNA binding proteins with regulatory functions. We identified AU-rich elements, Fox and Pumilio binding sites, and polyadenylation signal sequences (Fig 8) as motifs with slow rates of loss and gain. Our approach is therefore capable of characterizing selective pressures acting on a wide range of nucleotide sequence motifs. Importantly, we also found a large number of sequences with slow rates of loss and gain in 3'UTRs that do not correspond to previously recognized regulatory elements, such motifs may represent novel 3'UTR functional elements.

Fine-tuning by MicroRNA target sites

A diversity of functional roles have been proposed for miRNA target sites. Broadly speaking, these roles constitute three classes: subtle quantitative control of gene expression, often referred to as fine-tuning [55], miRNA-mediated switches that maximally repress a gene, or

contribute to this type of regulation in concert with other mechanisms [55], and sites that titrate miRNAs to a level optimal for regulation of a very small subset of true targets, an idea encapsulated by the competing endogenous RNA (ceRNA) hypothesis, [87, 88]. Undoubtedly all three classes exist, but their relative prevalence and importance is less certain. Our data indicate that substitutions that either strengthen or weaken existing target sites are disfavored (Fig 7); the simplest interpretation of this result is that target sites exist predominantly to optimize levels of the targeted transcript. In contrast, if miRNA target sites typically acted to maximize repression of gene expression, we would expect mutations that increase site efficacy to be selected for; aggregated across miRNAs, we observe the opposite: selection against sites increasing in strength. The ceRNA hypothesis implies selection maintaining the total number and strength of sites with minimal selection acting on their location across the transcriptome: our results are directly opposed to this model in that we observe selective pressure acting against the creation and destruction of most target sites, consistent with biochemical studies that imply ceRNAs are exceedingly rare [89, 90]. Thus, our results are most consistent with the model that miRNAs often serve to precisely modulate, or fine-tune, the levels of their target transcripts, as evidenced by selection not only maintaining the locations and numbers of sites, but the potency of the sites themselves.

A model of miRNA aging

We observe that many miRNAs known to have deeply conserved targeting, such as miR-9 and let-7, and additional miRNAs that we identified, such as miR-20 and miR-30, are also broadly expressed [53]. In general, we find that miRNAs with the strongest constraint against gain and loss of target sites (those in group 1) tend to have broader expression patterns than miRNAs with weaker constraint against gain of targeting (those in group 2, S15 Table) and much broader expression patterns than miRNAs with weak constraint against gain and loss of targeting (those in group 3, S15 Table). However, it is important to note that there are also exceptions. Among miRNAs that defy expected relationships between expression and conservation of targeting, miR-137 is among the most narrowly expressed in the dataset (S15 Table), and yet experiences stronger combined genome-wide selection against the acquisition of novel target sites and the destruction of existing sites than all but 4 miRNAs (miR-20, miR-30, miR-124, and miR-25; S1 and S2 Tables). Furthermore, when we examined gene-specific selection for miR-137, we observed broad selection across the transcriptome against both loss and gain of targeting (see S1 Zipped File for the miR-137 seed 'AGCAATAA'). At the other extreme, miR-146, which gains and loses miRNA targets faster than almost any other deeply conserved miRNA, is broadly expressed, being detected in most tissues. These results illustrate that while there is a trend for broadly expressed miRNAs to show more constrained evolution of targeting than narrowly expressed miRNAs, it is quite possible for a narrowly expressed miRNA to have highly constrained evolution of targeting, and, for a broadly expressed miRNA to gain and lose target sites rapidly.

In addition to a correlation between conservation of miRNA targeting (as assessed by miRNA group) and miRNA expression (S15 Table), we also observe that miRNAs in group 1 are conserved more deeply than miRNAs in group 2 (Wilcoxon Rank $P = 0.03$), and that miRNAs in group 1 are conserved more deeply than miRNAs in group 3 (S15 Table, Wilcoxon Rank $P = 0.002$). Combining these observations with the broad expression patterns of miRNAs [53] having the slowest gain rates, it seems feasible that in general, as miRNAs age and are expressed more broadly, this process allows novel targets of these miRNAs to become increasingly likely to be exposed to negative selection in at least one tissue, effectively 'trapping' miRNAs with the strongest deleterious effects as novel target sites arise into regulating the targets

they had already established, and hindering them from accumulating more. Under this model, a miRNA that originates with strong selection against novel targets and weak conservation of existing targets is likely to lose its existing targets without acquiring new beneficial ones. Such a miRNA would quickly lose its function and be unlikely to become deeply conserved. This model provides a ready explanation for the dearth of miRNAs that have strong selection against gaining new targets and weak selection against losing existing targets (Fig 4C). Among miRNAs with strongly deleterious effects for novel targets, only those with highly beneficial existing targets (and therefore strong selection against losing targets) are likely to be preserved. Meanwhile, miRNAs with narrow expression profiles, or other causes of weak selection against novel targets [91], are capable of exploring novel targeting interactions, and can be conserved so long as a meaningful fraction of existing targets are beneficial. Conceivably, a single beneficial target site would be sufficient to maintain a miRNA. Our results also highlight notable exceptions to this trend that may represent fruitful topics for further study. For example, miR-137 has target sites that are gained and lost very rarely despite being younger than almost all deeply conserved miRNAs, while miR-338 and miR-203a have target sites that are gained and lost relatively rapidly despite being older than most miRNAs. However, the overall trend toward miRNAs with more constrained target site evolution tending to be older (S15 Table), suggests a model of miRNA aging that is consistent with earlier work indicating that miRNAs generally acquire targets as they age [42]. This classification may guide the search for essential, deeply conserved functions of relatively understudied miRNAs of the first class, like miR-30 and miR-20, while suggesting clade or even species-specific roles for deeply conserved miRNAs of the third class, like miR-146.

Overall significance

We have demonstrated the utility of our technique for examining a wide range of evolutionary questions. We have examined the relative selective pressures operating on both deeply conserved and more recently evolved miRNAs. We have shown that for deeply conserved miRNAs, even target sites that have been completely lost in at least a subset of mammalian species still undergo substitutions at a rate that is lower than other comparable 8mers. We have presented evidence that substitutions that strengthen or weaken existing miRNA targeting are disfavored. A systematic examination of the rates at which miRNA target sites undergo substitutions makes these questions tractable. In addition, by quantifying rates of sequence motif gain and loss, our approach holds promise as a general method for studying regulatory sequence elements (Fig 8).

A thorough examination of the dynamics that have existed between miRNAs and their targets over time gives important information about the coevolution of miRNAs and their target genes, a subject that we are only beginning to understand. Our approach successfully recovers a signal for the binding sites of deeply conserved miRNAs, and has direct applications for improved miRNA functional analysis. Taken together, our approach has the ability to identify the overall strength of selection acting on both the loss and gain of miRNA binding sites, as well as alterations in strength of targeting and the distribution of natural selection across the genome. By correlating these metrics with the age and expression pattern of a given miRNA, we hope to shed light on the varied and changing roles of individual deeply conserved miRNAs.

Materials and methods

3'UTR sequences

Transcript coordinates were downloaded from the UCSC genome browser [5]. Transcripts with overlapping 3'UTRs were consolidated, retaining the longest 3'UTR isoform. Insertions

and deletions of multiple nucleotides were collapsed to a single 'N' in all species. 3'UTRs with a collapsed length of less than 50 nucleotides were removed from analysis, as were 3'UTRs in which any species was completely missing from the alignment. For analysis using mammalian sequences, we used the hg38 Refseq track and multiz 100 way alignments [92] to the hg38 reference. We cross-referenced these 100 species against miRBase version 21 [22], and found 60 species that overlapped with miRBase entries. Of these, 25 species had over 80% of their sequenced genome material in chromosomes, and of these, we chose to examine the 20 mammals (to maximize available alignable sequence to the human 'anchor' of the hg38 100way alignment). Our phylogeny of these 20 mammals was pruned from the publicly available UCSC molecular phylogeny of the 100way aligned species. From this pruned phylogeny, we chose ten mammalian species to maximize representation of diverse clades, choosing the highest quality genome available within each subgroup (e.g. one ruminant, the three deepest diverged clades of primates, one rodent, one lagomorph, one carnivore, etc.) while also ensuring a clear outgroup (one marsupial, opossum) (Fig 1A). For analysis of *Drosophila* sequences, we used UCSC multiz 27 way alignments [92] to the dm6 reference, examining the seven *Drosophila* species *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. biarmipes*, and *D. suzukii*. The complete datasets comprised 11,378 and 8,755 aligned 3'UTRs for the mammalian and *Drosophila* phylogenies, respectively.

Ancestral reconstruction and simulation

We used the aligned 3'UTRs in concert with the phylogeny provided with the UCSC multiz alignments of the dm6 and hg38 genomes as input for the program DNAML, which can be obtained as part of the phylip package [46]. The program outputs inferred ancestral sequences at every internal node of the phylogeny, for every alignment. Every ancestral sequence was paired with each of its descendants, such that every branch of the phylogeny was represented once per 3'UTR, and for every ancestor/descendant pair, we searched all trinucleotides within the alignments from all 3'UTR transcripts, in order to capture the effects of immediately adjacent nucleotides on the probability of a substitution. For each trinucleotide, we compared the ancestral trinucleotide to that of the aligned descendant. Because we were interested in the mutation probability of a nucleotide given its two flanking nucleotides, and to avoid double counting mutations, we ignored all trinucleotides in which the first or last nucleotide differed between the ancestor and descendant. Each ancestor/descendant pair thus produced a table of nucleotide substitution frequencies given fixed flanking nucleotides on either side. Because trinucleotides without a mutated central nucleotide were also recorded in the table, both the probability of incurring a substitution and the frequency of each type of substitution were accurately reproduced in the table. We generated one table (aggregated across all aligned 3'UTRs) per branch of the phylogeny. We also experimented with two alternative models similar to the one above, a model that examined two flanking nucleotides on either side of the central nucleotide (a 'pentamer' model, S4A Fig and S4B Fig for losses and gains, respectively), and a second model that examined individual nucleotides without regard to flanking nucleotides (a 'monomer' model, S4C Fig and S4D Fig for losses and gains, respectively). All three models had power to differentiate selection acting on miRNAs from the background set of all 65,536 8mers, (Pentamer model: $P = 5.8 \times 10^{-28}$ and $P = 1.4 \times 10^{-20}$ for losses and gains, respectively. Monomer model: $P = 5.8 \times 10^{-31}$ and $P = 3.0 \times 10^{-11}$ for losses and gains respectively), and the miRNAs identified with the standard model (termed the 'trimer' model) were well correlated with the pentamer model (S4E Fig and S4F Fig for losses and gains, respectively, $R^2 = 0.76$, $P = 4.8 \times 10^{-24}$, and $R^2 = 0.80$, $P = 1.4 \times 10^{-27}$ for losses and gains, respectively) and weakly correlated with the monomer model (S4G Fig and S4H Fig for losses and gains, respectively,

$R^2 = 0.03$, $P = 0.11$, and $R^2 = 0.50$, $P = 1.4 \times 10^{-12}$ for losses and gains, respectively). We selected the trimer model, as this model offered a balance between the effects of nucleotide context, together with a sequence length much shorter than that of the miRNA target sites themselves. We note that the trimer model produced the tightest distribution of standard deviation scores for our background set of all 8mers (blue line, Fig 2A and 2B versus S4A through S4D Fig).

To generate our simulated dataset, we scanned all ancestral trinucleotides a second time. We used the substitution probability table generated in the first pass to generate a weighted random process of choosing a descendant trinucleotide for each ancestor, and appended the central nucleotide of this randomly generated trinucleotide to a growing simulated descendant sequence. Each ancestor-descendant pair was simulated independently, such that the mutations that occurred in a descendant along an internal node do not propagate to its descendant. This is done to match the empirical probabilities of site conversions, which were also measured independently on each branch. This process was used to generate simulated descendant sequences for all 3'UTRs in all ancestor/descendant pairs, resulting in a complete phylogeny of simulated descendants for all sequences. This process was repeated 100 times to generate 100 simulated descendant 3'UTR datasets in mammals, each containing 11,378 genes in 17 ancestor/descendant pairs. In insects, this process was repeated 100 times to generate 100 simulated descendant 3'UTR datasets, each containing 8,755 genes, in eleven ancestor/descendant pairs.

We created shuffled 3'UTRs by breaking 3'UTRs in half and then interspersing 5mers from the first half of each 3'UTR with 5mers from the second half. The entire alignment for each 3'UTR was shuffled the same way (to avoid differential effects across different parts of the phylogeny), and miRNA target site turnover rates analyzed as described below.

Statistical analysis and ranking of expected and observed target site turnover rates

We examined every 8mer in both the simulated and observed datasets, and recorded instances in which the ancestral 8mer differed from its aligned descendant. These events were considered as a 'gain' of the descendant 8mer and as a 'loss' of the ancestral 8mer. MicroRNA target sites (8mers, 7mer-m8, 7mer-A1 and 6mers) were defined as matches to the 'seed' region of a known miRNA 8mer, based on established target site sequence preferences [19]. More specifically, we used python scripts to examine every 8mer, and determine whether the ancestor and/or descendant contained a match to the core 6mer target site (nucleotides 2–7), a core 7mer-A1 site (nucleotides 2–8 of the target site 8mer), a core 7mer-m8 site (nucleotides 1–7 of the target 8mer), or a perfect 8mer (nucleotides 1–8 of the 8mer). 8mer gain/loss events were defined as a discrepancy between the existence or nonexistence of a perfect 8mer match in the ancestor and descendant 8mers, while site conversions were defined as transitions from one type of match to a given 8mer to another type of match to the same 8mer. After scanning every ancestor/descendant paired alignment in this way, the result was tallied for each 8mer as a summed count of site gain events within the entire dataset (across all branches and genes) and a summed count of site loss events within the entire dataset. Because there were 100 simulated datasets and one observed dataset, our final summary recorded one observed 'gain' count for each 8mer and 100 simulated 'gain' counts, as well as similarly calculated observed and simulated 'loss' counts. We calculated the mean and standard deviation of the 100 simulated counts for each 8mer, and calculated the number of standard deviations each observed 8mer count fell above or below the mean of the simulated counts. Finally, the standard deviation calculations for each 8mer were ranked relative to each other in the 'gain' and 'loss' categories. The strength of selection acting on the mutation rates of an 8mer relative to other 8mers can be approximated from its rank, while a conservative estimate of the probability of observing a

similar or more extreme standard deviation can be approximated using Chebyshev's inequality [93], which states that for any distribution of values, the upper bound on the probability of observing a value that is k standard deviations from the mean is $P < 1/k^2$.

Because our aligned intronic sequence (selecting only regions with no insertions or deletions in any species) was roughly 10 times the size of our 3'UTR alignment, and in order to avoid altering our power to detect selection and avoid intronic splice sites that occur at the outer edges of introns, we examined the central 1/10th of every intron. Analysis was then performed as above.

Gene-specific analysis

We began by re-analyzing our dataset using the same techniques as before, but, instead of recording the total number of substitutions that occur across all 3'UTRs, we recorded the total number of substitutions that occurred for each 8mer within each gene. As previously, we compared the observed number of substitutions to the mean and standard deviation of 100 simulated substitutions, in order to count the number of standard deviations that the observed value fell above or below the mean expected value. We plotted this number of standard deviations for each gene on the y-axis, and ordered genes from those that were changing many standard deviations below expected rates to those changing many standard deviations above expected rates. Stochastic processes followed patterns that were not always intuitive. For example, an 8mer that almost never undergoes substitutions will appear to be changing at an extremely rapid rate within the small subset of genes that do undergo substitutions, even if no selection is present. To establish expected rates of substitution, we compared our observed dataset to the mean of 99 simulations, and additionally compared a random simulated dataset to the mean of the other 99 simulations. We repeated this process using every neutral simulation as foreground to generate 100 'neutral foreground' datasets and one 'selected foreground' dataset, and plotted the results ([S1 Zipped File](#)).

Removal of miRNA target sites perfectly conserved across analyzed 3'UTRs

We first located all positions in our observed alignments that have undergone substitutions in at least one descendant species, creating a set of mutated sites. We next counted the number of transition events (gains or losses) that occurred for 8mers overlapping by at least one nucleotide with these mutated sites. This resulted in an elimination of all perfectly conserved 8mer sites from the data. To ensure equally sized and identical pools of simulated ancestral 8mers to the observed dataset, we also examined the same coordinates within simulated alignments (regardless of whether a substitution occurred at these locations). Thus, our observed dataset contained only ancestral 8mers that have undergone at least one substitution in a descendant species, while our simulated dataset contained the same ancestral 8mers, regardless of whether they have undergone a substitution. The resulting observed and simulated datasets were used to examine whether miRNA binding sites that have undergone at least one substitution are constrained relative to other 8mers that have undergone at least one substitution.

miRNA site-conversion rates

miRNA 8mer target sites can be conceived as a 6mer 'core' site, which base pairs to nucleotides 2–7 of the miRNA, with two specific additional flanking nucleotides. Each flanking nucleotide alone converts a 6mer site into a different type of 7mer site. An 8mer→7mer_{1A} substitution is caused when the target site undergoes a substitution at its first nucleotide, an 8mer→7mer_{m8} substitution is a substitution at the last nucleotide, an 8mer→no site substitution is any substitution in the core 6 nucleotides, with equivalent conversions calculated between other classes of

target sites. These principles were generalized for all 8mers. By examining rates at which all 8mers undergo mutations at their first nucleotides, we can compare rates of 7mer_{m8}→8mer and 8mer→7mer_{m8} conversions for both miRNAs and other 8mers, which were used as controls. We repeated this analysis for all site conversion types (6mer→7mer_{m8}, 6mer→7mer_{1A}, 6mer→8mer, none→8mer, 7mer_{m8}→8mer, 7mer_{1A}→8mer, none→6mer, none→7mer_{1A}, and none→7mer_{m8} constituted our ‘gain’ event types, while each of these in reverse constituted site ‘loss’ events). Each 8mer conversion rate was ranked relative to the same conversion rate for all other 8mers as in our initial analysis, and as before, we performed a Wilcoxon rank-sum test to evaluate whether the set of target sites of conserved miRNAs have ranks that tend to fall significantly above or below the ranks of the complete set of all 8mers.

AU enriched/depleted sites

For each ancestral sequence in the phylogeny and each gene in our dataset, we recorded the coordinates of 8mer sites whose 30 preceding nucleotides and 30 following nucleotides (60 nucleotides total) had fewer than 30 A or T nucleotides (low AU sites). We also recorded the coordinates of 8mer sites whose 30 nucleotides on either side had greater than 30 A or T nucleotides (high AU sites). Importantly, we allowed different parts of the phylogeny even within the same gene to have different coordinates that showed as ‘high AU’ or ‘low AU’. We then analyzed how frequently every 8mer changed at these coordinates, in both our simulated and observed datasets.

miRNA expression analysis

We used published data [53] to estimate miRNA expression values in 172 distinct tissues, which were presented in units of sequenced clones containing the miRNA of interest. We counted the total number of clones recovered in the study (summed across all miRNAs and all tissues), and divided the number of clones recovered in a given tissue for a given miRNA by this total number of clones to derive the fraction of overall miRNA expression represented by a given miRNA in a given tissue. We counted the number of tissues in which a given miRNA’s expression represented greater than 1% of total recovered clones (column G of [S15 Table](#)), and we also counted the number of tissues in which there were any recovered clones for a given miRNA (column F of [S15 Table](#)).

Enrichment of known motifs

We examined the 1,000 8mer motifs with the most constrained rates of turnover (most standard deviations below the mean of simulations) in 3’UTRs, to establish whether known functional motifs were recovered by our method. We considered motifs with at least a 6mer match to known conserved miRNAs to be miRNA motifs. Among the remaining motifs, we searched for potential pumilio motifs (those containing ‘TGT’), and among the remaining motifs, we searched for potential Polyadenylation signals (those containing ‘AATAAA’, ‘ATTA AA’, ‘AGTAAA’, ‘TATAAA’, ‘CATAAA’, ‘GATAAA’, ‘AATATA’, ‘AATACA’, ‘AATAGA’, ‘ACTAAA’, ‘AAGAAA’, and ‘AATGAA’), and among the remaining motifs, we searched for potential AU-elements (those containing ‘ATTTA’). For comparison, we also examined the 1,000 8mer motifs with the most rapid rates of turnover (most standard deviations above the mean of simulations) in 3’UTRs.

All scripts may be found at ftp://www.alfredsimkin.net/neutral_motif_simulator.

Supporting information

S1 Fig. Power to detect selection as a function of branch length. We carried out the analyses of the main text for individual branches along the mammalian phylogeny. When analyzing the 15 shortest branches, in green (as measured by depletion of observed miRNA target site turn-over events relative to expected events, see bottom portion of [S1](#) and [S2](#) Tables), we observed R^2 values of 0.883 ($P = 2.0 \times 10^{-7}$) and 0.865 ($P = 5.1 \times 10^{-7}$) for losses (**A**) and gains (**B**) respectively. When two very long branches were included (black lines) we observed R^2 values of 0.64 ($P = 1.1 \times 10^{-4}$) and 0.022 ($P = 0.567$) for losses (**A**) and gains (**B**) respectively.
(TIF)

S2 Fig. Interconversions that strengthen or weaken miRNA target sites are disfavored. (A-T) Analysis of standard deviation metrics, calculated as in [Fig 2](#), for target site interconversion events between the site types none, 6mer, 7merA1, 7merm8, and 8mer. Wilcoxon rank sum statistics for miRNAs (red line) relative to all 8mers (blue line) are captured as p-values and shown in [Fig 7](#).
(TIF)

S3 Fig. Skew of miRNA binding sites in AU-rich and AU-poor regions. (A, B) AU-rich regions of 3'UTRs analyzed with respect to miRNA target site losses (A) and miRNA target site gains (B). **(C, D)** AU-poor regions of 3'UTRs analyzed with respect to miRNA target site losses (C) and miRNA target site gains (D).
(TIF)

S4 Fig. Comparison of a flanking single nucleotide model to other models of sequence evolution. The flanking single nucleotide model described in the text ('trimer' model) was compared to a model that utilized two flanking nucleotides on each side (**A, B**), a 'pentamer' model, and to **(C, D)**, a model that measured frequencies of nucleotide substitution without considering flanking nucleotides on either side (a 'monomer' model). Analysis performed as described in [Fig 2A and 2B](#) (losses on the left, gains on the right). **(E, F)** Correlations between miRNA target site loss (E) and gain (F) ranks between the 'trimer' (x-axis) and 'pentamer' model (y-axis). **(G, H)** Correlations between the 'trimer' (x-axis) and 'monomer' (y-axis) models, for losses (G) and gains (H), respectively.
(TIF)

S1 Table. Losses in 3'UTR target sites of deeply conserved miRNAs. Target sites are listed by 8mer target site sequence (8mer), recording loss rank (rank), number of standard deviations observed loss counts diverged from the mean of simulated counts (column D), the observed loss count, the mean of the simulated counts and list of miRNAs in the family corresponding to the 8mer target site.
(XLSX)

S2 Table. Gains in 3'UTR target sites of deeply conserved miRNAs. As described in [S1 Table](#), but for site gains.
(XLSX)

S3 Table. Losses in target sites of deeply conserved miRNAs, analyzed in intron sequence. As described in [S1 Table](#).
(XLSX)

S4 Table. Gains in target sites of deeply conserved miRNAs, analyzed in intron sequence. As described in [S2 Table](#).
(XLSX)

S5 Table. Losses in 3'UTR target sites of moderately conserved miRNAs. As described in [S1 Table](#).

(XLSX)

S6 Table. Gains in 3'UTR target sites of moderately conserved miRNAs. As described in [S2 Table](#).

(XLSX)

S7 Table. Losses in 3'UTR target sites of deeply conserved miRNAs, analyzed after removing highly conserved 3'UTR sequence. As described in [S1 Table](#).

(XLSX)

S8 Table. Gains in 3'UTR target sites of deeply conserved miRNAs, analyzed after removing highly conserved 3'UTR sequence. As described in [S2 Table](#).

(XLSX)

S9 Table. Losses in 3'UTR target sites of deeply conserved insect miRNAs. As described in [S1 Table](#).

(XLSX)

S10 Table. Gains in 3'UTR target sites of deeply conserved insect miRNAs. As described in [S2 Table](#).

(XLSX)

S11 Table. Losses in 3'UTR target sites of deeply conserved miRNAs, analyzed in high AU content contexts. As described in [S1 Table](#).

(XLSX)

S12 Table. Gains in 3'UTR target sites of deeply conserved miRNAs, analyzed in high AU content contexts. As described in [S2 Table](#).

(XLSX)

S13 Table. Losses in 3'UTR target sites of deeply conserved miRNAs, analyzed in low AU content contexts. As described in [S1 Table](#).

(XLSX)

S14 Table. Gains in 3'UTR target sites of deeply conserved miRNAs, analyzed in low AU content contexts. As described in [S2 Table](#).

(XLSX)

S15 Table. Expression levels of 75 deeply conserved vertebrate miRNAs across 172 tissues, and correlations with miRNA age and group number.

(XLSX)

S1 Zipped File. Analysis of target site gains and losses for individual miRNAs, as described in [Fig 6](#).

(7Z)

Acknowledgments

We thank Emily Davenport and Giles Hooker for helpful discussions.

Author Contributions

Conceptualization: Alfred Simkin, Andrew Grimson.

Formal analysis: Alfred Simkin, Rene Geissler, Alexa B. R. McIntyre.

Investigation: Alfred Simkin, Andrew Grimson.

Methodology: Alfred Simkin.

Supervision: Andrew Grimson.

Visualization: Rene Geissler.

Writing – original draft: Alfred Simkin, Andrew Grimson.

Writing – review & editing: Alfred Simkin, Andrew Grimson.

References

- Schmitz JF, Zimmer F, Bornberg-Bauer E. Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Res.* 2016; 44(13):6287–97. <https://doi.org/10.1093/nar/gkw492> PMID: 27288445; PubMed Central PMCID: PMC5291267.
- Schwaiger M, Schonauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, et al. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* 2014; 24(4):639–50. <https://doi.org/10.1101/gr.162529.113> PMID: 24642862; PubMed Central PMCID: PMC3975063.
- Vinogradov AE, Anatskaya OV. Organismal complexity, cell differentiation and gene expression: human over mouse. *Nucleic Acids Res.* 2007; 35(19):6350–6. <https://doi.org/10.1093/nar/gkm723> PMID: 17881362; PubMed Central PMCID: PMC2095826.
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature.* 2005; 437(7055):69–87. <https://doi.org/10.1038/nature04072> PMID: 16136131.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002; 12(6):996–1006. <https://doi.org/10.1101/gr.229102> PMID: 12045153; PubMed Central PMCID: PMC186604.
- Won YJ, Sivasundar A, Wang Y, Hey J. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. *Proc Natl Acad Sci U S A.* 2005; 102 Suppl 1:6581–6. <https://doi.org/10.1073/pnas.0502127102> PMID: 15851665; PubMed Central PMCID: PMC1131877.
- Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife.* 2015; 4. <https://doi.org/10.7554/eLife.04837> PMID: 25779349; PubMed Central PMCID: PMC4362205.
- Anderson DW, McKeown AN, Thornton JW. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *Elife.* 2015; 4:e07864. <https://doi.org/10.7554/eLife.07864> PMID: 26076233; PubMed Central PMCID: PMC4500092.
- Nakagawa S, Gisselbrecht SS, Rogers JM, Hartl DL, Bulyk ML. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc Natl Acad Sci U S A.* 2013; 110(30):12349–54. <https://doi.org/10.1073/pnas.1310430110> PMID: 23836653; PubMed Central PMCID: PMC3725104.
- Presnell JS, Schnitzler CE, Browne WE. KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes. *Genome Biol Evol.* 2015; 7(8):2289–309. <https://doi.org/10.1093/gbe/evv141> PMID: 26232396; PubMed Central PMCID: PMC4558859.
- Teichmann SA, Babu MM. Gene regulatory network growth by duplication. *Nat Genet.* 2004; 36(5):492–6. <https://doi.org/10.1038/ng1340> PMID: 15107850.
- Thompson D, Regev A, Roy S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu Rev Cell Dev Biol.* 2015; 31:399–428. <https://doi.org/10.1146/annurev-cellbio-100913-012908> PMID: 26355593.
- Duque T, Samee MA, Kazemian M, Pham HN, Brodsky MH, Sinha S. Simulations of enhancer evolution provide mechanistic insights into gene regulation. *Mol Biol Evol.* 2014; 31(1):184–200. <https://doi.org/10.1093/molbev/mst170> PMID: 24097306; PubMed Central PMCID: PMC3879441.
- Otto W, Stadler PF, Lopez-Giraldez F, Townsend JP, Lynch VJ, Wagner GP. Measuring transcription factor-binding site turnover: a maximum likelihood approach using phylogenies. *Genome Biol Evol.* 2009; 1:85–98. <https://doi.org/10.1093/gbe/evp010> PMID: 20333180; PubMed Central PMCID: PMC2817405.
- Tanay A, Gat-Viks I, Shamir R. A global view of the selection forces in the evolution of yeast cis-regulation. *Genome Res.* 2004; 14(5):829–34. <https://doi.org/10.1101/gr.2064404> PMID: 15123582; PubMed Central PMCID: PMC479109.

16. Habib N, Wapinski I, Margalit H, Regev A, Friedman N. A functional selection model explains evolutionary robustness despite plasticity in regulatory networks. *Mol Syst Biol.* 2012; 8:619. <https://doi.org/10.1038/msb.2012.50> PMID: 23089682; PubMed Central PMCID: PMC3501536.
17. Bartel DP. Metazoan MicroRNAs. *Cell.* 2018; 173(1):20–51. <https://doi.org/10.1016/j.cell.2018.03.006> PMID: 29570994; PubMed Central PMCID: PMC6091663.
18. Moran Y, Agron M, Praher D, Technau U. The evolutionary origin of plant and animal microRNAs. *Nat Ecol Evol.* 2017; 1(3):27. <https://doi.org/10.1038/s41559-016-0027> PMID: 28529980; PubMed Central PMCID: PMC5435108.
19. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009; 136(2):215–33. <https://doi.org/10.1016/j.cell.2009.01.002> PMID: 19167326; PubMed Central PMCID: PMC3794896.
20. Chen K, Rajewsky N. Deep conservation of microRNA-target relationships and 3'UTR motifs in vertebrates, flies, and nematodes. *Cold Spring Harb Symp Quant Biol.* 2006; 71:149–56. <https://doi.org/10.1101/sqb.2006.71.039> PMID: 17381291.
21. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell.* 2003; 115(7):787–98. [https://doi.org/10.1016/s0092-8674\(03\)01018-3](https://doi.org/10.1016/s0092-8674(03)01018-3) PMID: 14697198.
22. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014; 42(Database issue):D68–73. <https://doi.org/10.1093/nar/gkt1181> PMID: 24275495; PubMed Central PMCID: PMC3965103.
23. Xu J, Zhang R, Shen Y, Liu G, Lu X, Wu CI. The evolution of evolvability in microRNA target sites in vertebrates. *Genome Res.* 2013; 23(11):1810–6. <https://doi.org/10.1101/gr.148916.112> PMID: 24077390; PubMed Central PMCID: PMC3814881.
24. Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, et al. The birth and death of microRNA genes in *Drosophila*. *Nat Genet.* 2008; 40(3):351–5. <https://doi.org/10.1038/ng.73> PMID: 18278047.
25. Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, et al. Birth and expression evolution of mammalian microRNA genes. *Genome Res.* 2013; 23(1):34–45. <https://doi.org/10.1101/gr.140269.112> PMID: 23034410; PubMed Central PMCID: PMC3530682.
26. Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in *Drosophila* species. *Genome Biol Evol.* 2010; 2:180–9. <https://doi.org/10.1093/gbe/evq009> PMID: 20624724; PubMed Central PMCID: PMC2942034.
27. Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in plant species. *Genome Biol Evol.* 2012; 4(3):230–9. <https://doi.org/10.1093/gbe/evs002> PMID: 22223755; PubMed Central PMCID: PMC3318440.
28. Lyu Y, Shen Y, Li H, Chen Y, Guo L, Zhao Y, et al. New microRNAs in *Drosophila*—birth, death and cycles of adaptive evolution. *PLoS Genet.* 2014; 10(1):e1004096. <https://doi.org/10.1371/journal.pgen.1004096> PMID: 24465220; PubMed Central PMCID: PMC3900394.
29. Mohammed J, Flynt AS, Panzarino AM, Mondal MMH, DeCruz M, Siepel A, et al. Deep experimental profiling of microRNA diversity, deployment, and evolution across the *Drosophila* genus. *Genome Res.* 2018; 28(1):52–65. <https://doi.org/10.1101/gr.226068.117> PMID: 29233922; PubMed Central PMCID: PMC5749182.
30. Penso-Dolfín L, Moxon S, Haerty W, Di Palma F. The evolutionary dynamics of microRNAs in domestic mammals. *Sci Rep.* 2018; 8(1):17050. <https://doi.org/10.1038/s41598-018-34243-8> PMID: 30451897; PubMed Central PMCID: PMC6242877.
31. Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, et al. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science.* 2005; 310(5755):1817–21. <https://doi.org/10.1126/science.1121158> PMID: 16308420.
32. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature.* 2000; 408(6808):86–9. <https://doi.org/10.1038/35040556> PMID: 11081512.
33. Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell.* 2005; 123(6):1133–46. <https://doi.org/10.1016/j.cell.2005.11.023> PMID: 16337999.
34. Wolter JM, Le HH, Linse A, Godlove VA, Nguyen TD, Kotagama K, et al. Evolutionary patterns of metazoan microRNAs reveal targeting principles in the let-7 and miR-10 families. *Genome Res.* 2017; 27(1):53–63. <https://doi.org/10.1101/gr.209361.116> PMID: 27927717; PubMed Central PMCID: PMC5204344.
35. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature.* 2005; 434(7031):338–45. <https://doi.org/10.1038/nature03441> PubMed Central PMCID: PMC2923337. PMID: 15735639

36. Luo J, Wang Y, Yuan J, Zhao Z, Lu J. MicroRNA duplication accelerates the recruitment of new targets during vertebrate evolution. *RNA*. 2018; 24(6):787–802. <https://doi.org/10.1261/rna.062752.117> PMID: 29511046; PubMed Central PMCID: PMC5959248.
37. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120(1):15–20. <https://doi.org/10.1016/j.cell.2004.12.035> PMID: 15652477.
38. Loh YH, Yi SV, Streelman JT. Evolution of microRNAs and the diversification of species. *Genome Biol Evol*. 2011; 3:55–65. <https://doi.org/10.1093/gbe/evq085> PMID: 21169229; PubMed Central PMCID: PMC3017390.
39. Franchini P, Xiong P, Fruciano C, Meyer A. The Role of microRNAs in the Repeated Parallel Diversification of Lineages of Midas Cichlid Fish from Nicaragua. *Genome Biol Evol*. 2016; 8(5):1543–55. <https://doi.org/10.1093/gbe/evw097> PMID: 27189980; PubMed Central PMCID: PMC4898811.
40. Hatlen AM, A. Pervasive selection against microRNA target sites in human populations. *bioRxiv*. 2018.
41. Leclercq M, Diallo AB, Blanchette M. Prediction of human miRNA target genes using computationally reconstructed ancestral mammalian sequences. *Nucleic Acids Res*. 2017; 45(2):556–66. <https://doi.org/10.1093/nar/gkw1085> PMID: 27899600; PubMed Central PMCID: PMC5314757.
42. Nozawa M, Fujimi M, Iwamoto C, Onizuka K, Fukuda N, Ikeo K, et al. Evolutionary Transitions of MicroRNA-Target Pairs. *Genome Biol Evol*. 2016; 8(5):1621–33. <https://doi.org/10.1093/gbe/evw092> PMID: 27189995; PubMed Central PMCID: PMC4898806.
43. Shou C, Bhardwaj N, Lam HY, Yan KK, Kim PM, Snyder M, et al. Measuring the evolutionary rewiring of biological networks. *PLoS Comput Biol*. 2011; 7(1):e1001050. <https://doi.org/10.1371/journal.pcbi.1001050> PMID: 21253555; PubMed Central PMCID: PMC3017101.
44. Simkin AT, Bailey JA, Gao FB, Jensen JD. Inferring the evolutionary history of primate microRNA binding sites: overcoming motif counting biases. *Mol Biol Evol*. 2014; 31(7):1894–901. <https://doi.org/10.1093/molbev/msu129> PMID: 24723422; PubMed Central PMCID: PMC4069616.
45. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res*. 2018; 46(D1):D762–D9. <https://doi.org/10.1093/nar/gkx1020> PMID: 29106570; PubMed Central PMCID: PMC5753355.
46. Felsenstein J. PHYLIP-phylogeny inference package (Version 3.2). *Cladistics*. 2002; 5:164–6.
47. Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*. 2007; 27(1):91–105. <https://doi.org/10.1016/j.molcel.2007.06.017> PMID: 17612493; PubMed Central PMCID: PMC3800283.
48. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009; 19(1):92–105. <https://doi.org/10.1101/gr.082701.108> PMID: 18955434; PubMed Central PMCID: PMC2612969.
49. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, et al. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature*. 2007; 450(7167):219–32. <https://doi.org/10.1038/nature06340> PMID: 17994088; PubMed Central PMCID: PMC2474711.
50. Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. *Cell*. 2012; 149(3):515–24. <https://doi.org/10.1016/j.cell.2012.04.005> PMID: 22541426; PubMed Central PMCID: PMC3351105.
51. Halligan DL. Patterns of Evolutionary Constraints in Intronic and Intergenic DNA of *Drosophila*. *Genome Research*. 2004; 14(2):273–9. <https://doi.org/10.1101/gr.1329204> PMID: 14762063
52. Blanchette M. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Research*. 2004; 14(4):708–15. <https://doi.org/10.1101/gr.1933104> PMID: 15060014
53. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*. 2007; 129(7):1401–14. <https://doi.org/10.1016/j.cell.2007.04.040> PMID: 17604727; PubMed Central PMCID: PMC2681231.
54. Pinzon N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, et al. microRNA target prediction programs predict many false positives. *Genome Res*. 2017; 27(2):234–45. <https://doi.org/10.1101/gr.205146.116> PMID: 28148562; PubMed Central PMCID: PMC5287229.
55. Fridrich A, Hazan Y, Moran Y. Too Many False Targets for MicroRNAs: Challenges and Pitfalls in Prediction of miRNA Targets and Their Gene Ontology in Model and Non-model Organisms. *Bioessays*. 2019; 41(4):e1800169. <https://doi.org/10.1002/bies.201800169> PMID: 30919506; PubMed Central PMCID: PMC6701991.
56. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of meta-zoan microRNAs. *Nat Rev Genet*. 2004; 5(5):396–400. <https://doi.org/10.1038/nrg1328> PMID: 15143321.

57. Mayr C. Regulation by 3'-Untranslated Regions. *Annu Rev Genet.* 2017; 51:171–94. <https://doi.org/10.1146/annurev-genet-120116-024704> PMID: 28853924.
58. Beaulieu E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 2000; 10(7):1001–10. <https://doi.org/10.1101/gr.10.7.1001> PMID: 10899149; PubMed Central PMCID: PMC310884.
59. Lagnado CA, Brown CY, Goodall GJ. AUUUA is not sufficient to promote poly(A) shortening and degradation of an mRNA: the functional sequence within AU-rich elements may be UUAUUUA(U/A)(U/A). *Mol Cell Biol.* 1994; 14(12):7984–95. <https://doi.org/10.1128/mcb.14.12.7984> PMID: 7969138; PubMed Central PMCID: PMC359337.
60. Li X, Quon G, Lipshitz HD, Morris Q. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA.* 2010; 16(6):1096–107. <https://doi.org/10.1261/ma.2017210> PMID: 20418358; PubMed Central PMCID: PMC2874161.
61. Kheradpour P, Stark A, Roy S, Kellis M. Reliable prediction of regulator targets using 12 Drosophila genomes. *Genome Res.* 2007; 17(12):1919–31. <https://doi.org/10.1101/gr.7090407> PMID: 17989251; PubMed Central PMCID: PMC2099599.
62. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. *Nat Genet.* 2005; 37(5):495–500. <https://doi.org/10.1038/ng1536> PMID: 15806104.
63. Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci U S A.* 2006; 103(8):2746–51. <https://doi.org/10.1073/pnas.0511045103> PMID: 16477010; PubMed Central PMCID: PMC1413820.
64. Wang Y, Luo J, Zhang H, Lu J. microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. *Mol Biol Evol.* 2016; 33(9):2232–47. <https://doi.org/10.1093/molbev/msw089> PMID: 27189568; PubMed Central PMCID: PMC4989102.
65. Hu F, Wang M, Xiao T, Yin B, He L, Meng W, et al. miR-30 promotes thermogenesis and the development of beige fat by targeting RIP140. *Diabetes.* 2015; 64(6):2056–68. <https://doi.org/10.2337/db14-1117> PMID: 25576051; PubMed Central PMCID: PMC4876748.
66. Peck BC, Sincavage J, Feinstein S, Mah AT, Simmons JG, Lund PK, et al. miR-30 Family Controls Proliferation and Differentiation of Intestinal Epithelial Cell Models by Directing a Broad Gene Expression Program That Includes SOX9 and the Ubiquitin Ligase Pathway. *J Biol Chem.* 2016; 291(31):15975–84. <https://doi.org/10.1074/jbc.M116.733733> PMID: 27261459; PubMed Central PMCID: PMC4965549.
67. Zaragosi LE, Wdziekonski B, Brigand KL, Villageois P, Mari B, Waldmann R, et al. Small RNA sequencing reveals miR-642a-3p as a novel adipocyte-specific microRNA and miR-30 as a key regulator of human adipogenesis. *Genome Biol.* 2011; 12(7):R64. <https://doi.org/10.1186/gb-2011-12-7-r64> PMID: 21767385; PubMed Central PMCID: PMC3218826.
68. Hua Z, Lv Q, Ye W, Wong CK, Cai G, Gu D, et al. MiRNA-directed regulation of VEGF and other angiogenic factors under hypoxia. *PLoS One.* 2006; 1:e116. <https://doi.org/10.1371/journal.pone.0000116> PMID: 17205120; PubMed Central PMCID: PMC1762435.
69. Shehadeh LA, Sharma S, Pessanha M, Wei JQ, Liu J, Yuan H, et al. MicroRNA-20a constrains p300-driven myocardial angiogenic transcription by direct targeting of p300. *PLoS One.* 2013; 8(11):e79133. <https://doi.org/10.1371/journal.pone.0079133> PMID: 24236097; PubMed Central PMCID: PMC3827282.
70. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkland SJ, et al. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell.* 2008; 132(5):875–86. <https://doi.org/10.1016/j.cell.2008.02.019> PMID: 18329372; PubMed Central PMCID: PMC2323338.
71. Lee HM, Kim TS, Jo EK. MiR-146 and miR-125 in the regulation of innate immunity and inflammation. *BMB Rep.* 2016; 49(6):311–8. <https://doi.org/10.5483/BMBRep.2016.49.6.056> PMID: 26996343; PubMed Central PMCID: PMC5070718.
72. Testa U, Pelosi E, Castelli G, Labbaye C. miR-146 and miR-155: Two Key Modulators of Immune Response and Tumor Development. *Noncoding RNA.* 2017; 3(3). <https://doi.org/10.3390/ncrna3030022> PMID: 29657293; PubMed Central PMCID: PMC5831915.
73. Wang X, Cao J, Yu Y, Ma B, Gao C, Lu J, et al. Role of MicroRNA 146a in Regulating Regulatory T Cell Function to Ameliorate Acute Cardiac Rejection in Mice. *Transplant Proc.* 2019; 51(3):901–12. <https://doi.org/10.1016/j.transproceed.2019.01.026> PMID: 30979483.
74. Xiao Q, Zhu X, Yang S, Wang J, Yin R, Song J, et al. LPS induces CXCL16 expression in HUVECs through the miR-146a-mediated TLR4 pathway. *Int Immunopharmacol.* 2019; 69:143–9. <https://doi.org/10.1016/j.intimp.2019.01.011> PMID: 30710793.
75. Zhang L, Fu Y, Wang H, Guan Y, Zhu W, Guo M, et al. Severe Fever With Thrombocytopenia Syndrome Virus-Induced Macrophage Differentiation Is Regulated by miR-146. *Front Immunol.* 2019;

- 10:1095. <https://doi.org/10.3389/fimmu.2019.01095> PMID: 31156641; PubMed Central PMCID: PMC6529556.
76. Simkin A, Wong A, Poh YP, Theurkauf WE, Jensen JD. Recurrent and recent selective sweeps in the piRNA pathway. *Evolution*. 2013; 67(4):1081–90. <https://doi.org/10.1111/evo.12011> PMID: 23550757; PubMed Central PMCID: PMC3992950.
 77. Bassett AR, Azzam G, Wheatley L, Tibbit C, Rajakumar T, McGowan S, et al. Understanding functional miRNA-target interactions in vivo by site-specific genome engineering. *Nat Commun*. 2014; 5:4640. <https://doi.org/10.1038/ncomms5640> PMID: 25135198; PubMed Central PMCID: PMC4143950.
 78. Lai EC. Two decades of miRNA biology: lessons and challenges. *RNA*. 2015; 21(4):675–7. <https://doi.org/10.1261/rna.051193.115> PMID: 25780186; PubMed Central PMCID: PMC4371328.
 79. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. *Curr Biol*. 2002; 12(9):735–9. [https://doi.org/10.1016/s0960-9822\(02\)00809-6](https://doi.org/10.1016/s0960-9822(02)00809-6) PMID: 12007417.
 80. Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW. A microRNA imparts robustness against environmental fluctuation during development. *Cell*. 2009; 137(2):273–82. <https://doi.org/10.1016/j.cell.2009.01.058> PMID: 19379693; PubMed Central PMCID: PMC2674871.
 81. Wu CI, Shen Y, Tang T. Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res*. 2009; 19(5):734–43. <https://doi.org/10.1101/gr.084640.108> PMID: 19411598; PubMed Central PMCID: PMC3647535.
 82. Cottrell KA, Chaudhari HG, Cohen BA, Djuranovic S. PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. *Nat Commun*. 2018; 9(1):301. <https://doi.org/10.1038/s41467-017-02745-0> PMID: 29352242; PubMed Central PMCID: PMC5775260.
 83. Geissler R, Simkin A, Floss D, Patel R, Fogarty EA, Scheller J, et al. A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes Dev*. 2016; 30(9):1070–85. <https://doi.org/10.1101/gad.277392.116> PMID: 27151978; PubMed Central PMCID: PMC4863738.
 84. Rabani M, Pieper L, Chew GL, Schier AF. A Massively Parallel Reporter Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation. *Mol Cell*. 2017; 68(6):1083–94 e5. <https://doi.org/10.1016/j.molcel.2017.11.014> PMID: 29225039; PubMed Central PMCID: PMC5994907.
 85. Wissink EM, Fogarty EA, Grimson A. High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics*. 2016; 17:177. <https://doi.org/10.1186/s12864-016-2479-7> PMID: 26941072; PubMed Central PMCID: PMC4778349.
 86. Zhao W, Pollack JL, Blagev DP, Zaitlen N, McManus MT, Erle DJ. Massively parallel functional annotation of 3' untranslated regions. *Nat Biotechnol*. 2014; 32(4):387–91. <https://doi.org/10.1038/nbt.2851> PMID: 24633241; PubMed Central PMCID: PMC3981918.
 87. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146(3):353–8. <https://doi.org/10.1016/j.cell.2011.07.014> PMID: 21802130; PubMed Central PMCID: PMC3235919.
 88. Seitz H. Redefining microRNA targets. *Curr Biol*. 2009; 19(10):870–3. <https://doi.org/10.1016/j.cub.2009.03.059> PMID: 19375315.
 89. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell*. 2014; 56(3):347–59. <https://doi.org/10.1016/j.molcel.2014.09.018> PMID: 25449132; PubMed Central PMCID: PMC5048918.
 90. Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol Cell*. 2014; 54(5):766–76. <https://doi.org/10.1016/j.molcel.2014.03.045> PMID: 24793693; PubMed Central PMCID: PMC4267251.
 91. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol*. 2011; 18(10):1139–46. <https://doi.org/10.1038/nsmb.2115> PMID: 21909094; PubMed Central PMCID: PMC3190056.
 92. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004; 14(4):708–15. <https://doi.org/10.1101/gr.1933104> PMID: 15060014; PubMed Central PMCID: PMC383317.
 93. Tchebichef P. Des valeurs moyennes. *Journal de Mathématiques Pures et Appliquées*. 1867; 2(12):177–84.