Supporting Information

Schnall-Levin et al. 10.1073/pnas.1006172107

SI Text

Cell Transfections. S2R$^+$ cells were maintained in Schneider's medium (Invitrogen), supplemented with 10% FBS and 1% pen-strep. Cells were transfected in 12-well plates using the Effectene Transfection Kit (Qiagen) according to the manufacturer's instructions. For each ORF, cells were cotransfected with three plasmids: (*i*) ORF-WT fused to a Myc tag (or FLAG tag in tag-swap), (*ii*) ORF-MUT fused to a FLAG tag (or Myc tag in tag-swap), and (*iii*) either mCherry (Control) or mCherry-microRNA, all under the Actin promoter. Cells were cultured for three days, lysed, and analyzed by Western blot using LI-COR reagents. Imaging and quantification were performed using the LI-COR Aerius Infrared Imaging System.

Microarrays. S2R$^+$ cells were transfected with either mCherry (Control) or mCherry-miR-1, both under the Actin promoter, as described above. Cells were cultured for two days before harvest and total RNA was extracted using TriZol reagent (Invitrogen) and further purified using Qiagen RNeasy column. Recovered RNA was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technology). RNA integrity was assessed using an Agilent 2100 bioanalyzer. Samples were labeled following Agilent's two-color microarray-based gene expression analysis (Quick Amp labeling) protocol. Gene expression profiles were generated using a customized Agilent 8 × 15k whole *Drosophila* Genome Microarray, processed in duplicate, and expression levels were extracted using Agilent Feature Extraction software. Log ratios were averaged over multiple probes and over technical replicates. Only probes with signal above the median were included in the analysis.

Mutagenesis. Mutagenesis was carried out with the QuikChange II Site-Directed Mutagenesis Kit (Stratagene). All miRNA seed sites were disrupted with two synonymous point mutations.

Below we show the microRNA seed sites in their context and the mutagenesis primers used. Seed matches are shown in bold and amino acid sequences are shown below the nucleotide sequences. Mutated positions are highlighted in red. All nucleotide sequences displayed begin in-frame.

1. Jaguar (FBgn0011225). (i) K Box 8-mer site:

WT: TC**CTGTGATA**TT
AA: Ser Cys Asp Ile
Mut: TC**CTGCGACA**TT

Forward primer: GATGCTATCAACACGTCCTGCGACATTGAGCTGCTGGAGGCCTG
Reverse primer: CAGGCCTCCAGCAGCTCAATGTCGCAGGACGTGTTGATAGCATC

2. CG11178 (FBgn0030499). (i) Mir-1 8-mer site:

WT: **ACATTCCA**G
AA: Thr Phe Gln
Mut: **ACTTTTCA**G

Forward primer: CAGCCCAAGAGATCTCAGTTACTTTTCAGAATCATAAGGACGTCGAAG
Reverse primer: CTTCGACGTCCTTATGATTCTGAAAAGTAACTGAGATCTCTTGGGCTG

3. CG8494 (FBgn0033916). (i) Mir-1 8-mer site:

WT: **ACATTCCA**G
AA: Thr Phe Gln
Mut: **ACTTTTCA**G

Forward primer: GTCCACACGCGAGGAAACTTTTCAGGATCTCTCGCTGCCC
Reverse primer: GGGCAGCGAGAGATCCTGAAAAGTTTCCTCGCGTGTGGAC
(ii) Mir-1 7-mer site #1:

WT: **CATTCCA**AG
AA: His Ser Lys
Mut: **CACTCGA**AG

Forward primer: CCATGGTGGGATTTTGCACTCGAAGGCGGACGTAATCAGC
Reverse primer: CTGATTACGTCCGCCTTCGAGTGCAAAATCCCACCATGGG
(iii) Mir-1 7-mer site #2:

WT: C**CATTCCA**A
AA: Pro Phe Gln
Mut: C**CCTTTCA**A

Forward primer: AGCTTACCCATTTCGATACCCTTTCAAAGCGACAATTTCCAGGTG
Reverse primer: ACCTGGAAATTGTCGCTTTGAAAGGGTATCGAAATGGGTAAGCTG

4. Smaug (FBgn0016070). (i) K Box 8-mer site:

WT: CT**CTGTGATA**AT
AA: Leu Cys Asp Asn
Mut: CT**CTGCGACA**AT

Forward primer: GGTCGATCAATCCACTCTGCGACAATCTTAATGGTATTACCC
Reverse primer: GGTAATACCATTAAGATTGTCGCAGAGTGGATTGATCGACCC

5. Arp87C (FBgn0011745). (i) Mir-1 8-mer site:

WT: C**ACATTCCA**
AA: His Ile Pro
Mut: C**ATATCCCA**

Forward primer: CCGGCTTTGCTGGTGAGCATATCCCAAAATGCAGGTTTCCC
Reverse primer: GGGAAACCTGCATTTTGGGATATGCTCACCAGCAAAGCCGG
(ii) Mir-8 8-mer site:

WT: CA**CAGTATTA**TG
AA: His Ser Ile Met
Mut: CA**CAGCATCA**TG

Forward primer: GATTCGCCATGCCTCACAGCATCATGCGCGTGGACATCGCC
Reverse primer: GGCGATGTCCACGCGCATGATGCTGTGAGGCATGGCGAATC

6. Act88F (FBgn0000047). (i) Mir-8 8-mer site:

WT: C**CAGTATTA**
AA: Pro Val Leu
Mut: C**CCGTTTTA**

Forward primer: GTGGCCCCCGAGGAGCATCCCGTTTTATTGACCGAGGCTCCACTG

Reverse primer: CAGTGGAGCCTCGGTCAATAAAACGGGATGCTCCTCGGGGGCCAC

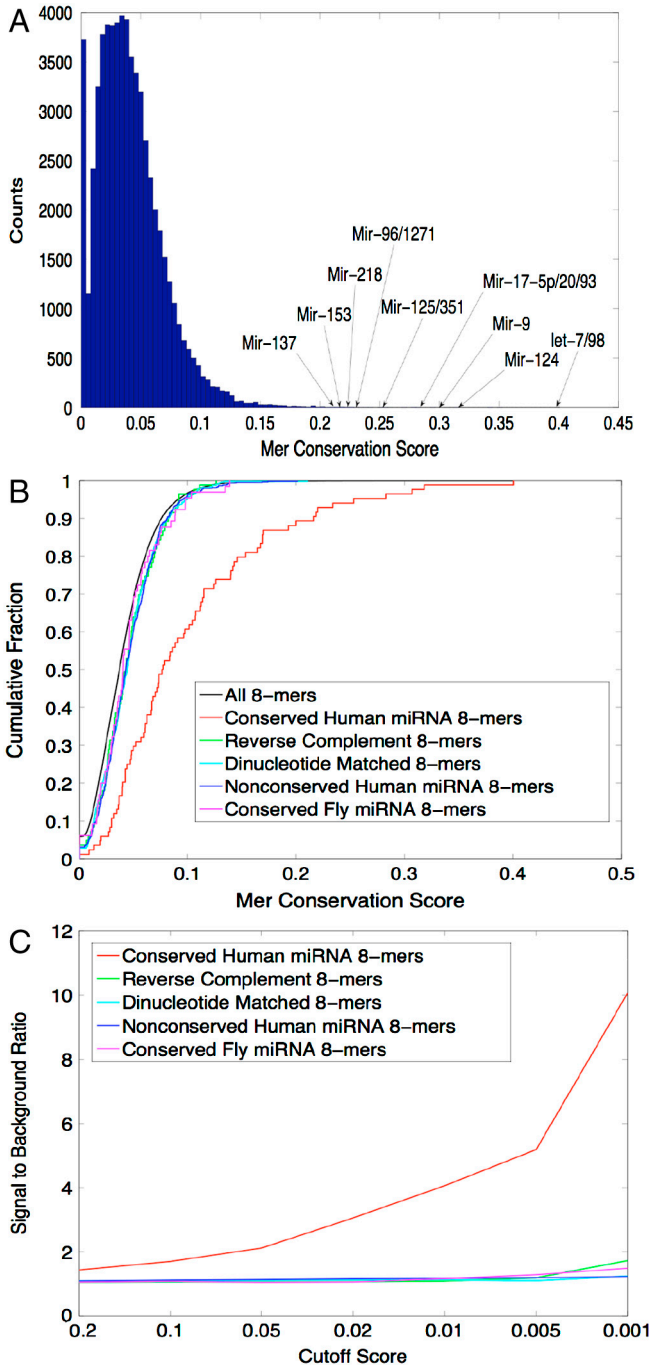

Fig. S1. MicroRNA seed sites are among the most highly conserved motifs in coding regions in humans. (*A*) Histogram of conservation scores for all 65,536 8-mers. A majority of the top conserved 8-mers correspond to miRNA seed sites. (*B*) Cumulative plot of scores for different sets of 8-mers. Shown are all 8-mers (black), conserved human miRNA seeds (red), reverse complements of these seeds (green), 8-mers with identical dinucleotide content to these seeds (cyan), nonconserved human miRNA seeds (blue), and conserved *Drosophila* miRNA seeds (magenta). (*C*) Imposing increasingly stringent conservation cutoff results in higher signal-to-noise ratios for the set of human conserved miRNA seeds, whereas control sets behave as background at all cutoffs.

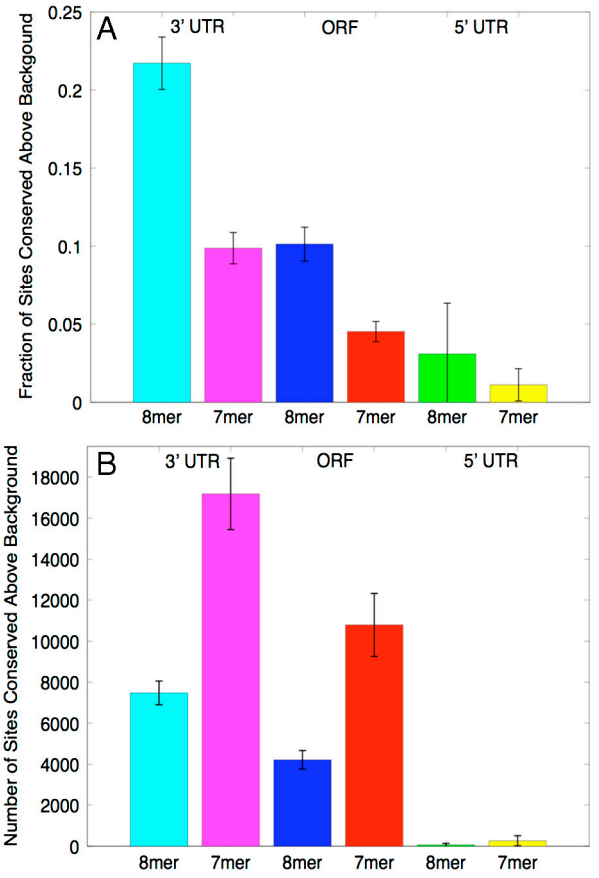

Fig. S2. The scale of conserved miRNA targeting in 3′UTRs, ORFs, and 5′UTRs in humans. (*A*) Fraction of sites conserved above background for both 8-mers and 7-mers in 3′UTRs, ORFs, and 5′UTRs. (*B*) Number of predicted sites above background for 8-mers and 7-mers in 3′UTRs, ORFs, and 5′UTRs. Error bars show standard deviation in the estimates obtained from sampling of background sets (see *Methods*).

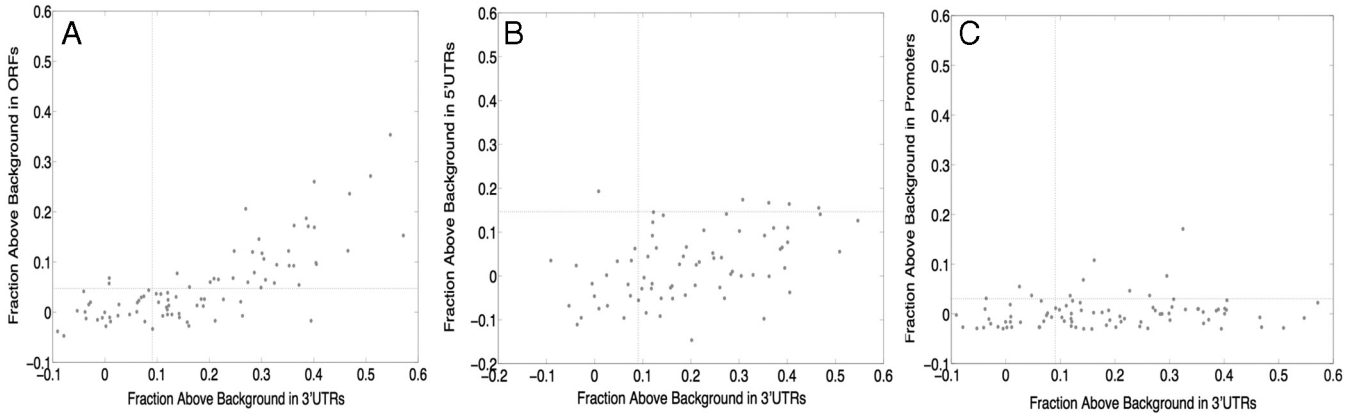

Fig. S3. MicroRNA seeds showing the highest level of conservation in human 3′UTRs tend also to be the most conserved in ORFs and to a very small extent in 5′ UTRs, but not in promoter regions. Shown are the fractions of sites conserved above background at 60% confidence cutoff (for 5′UTRs and promoters the cutoff was chosen to be the same as for 3′UTRs) between (*A*) 3′UTRs and ORFs, (*B*) 3′UTRs and 5′UTRs, and (*C*) 3′UTRs and promoters. Dotted vertical and horizontal lines show the cutoff for conservation above background equal to the maximal amount by which any miRNA was conserved below background.

Table S1. The most conserved 8-mer motifs in *Drosophila* coding regions

Highest scoring motifs	Annotation
AAGACTGA AGACTGAA	mir-14
CTGTGATA TGTGATAC ACTGTGAT TCTGTGAT TGTGATAT TGTGATAA TTGTGATA	K-box (mir-2a/6/11/13/308)
ACATTCCA CATTCCAA AACATTCC	mir-1
ATGAACAA ATGGACAA ATGTACAA	Unknown
TCTAGTCA CTCTAGTC TCTAGTCT	mir-279/286/996
ACATATCA	mir-190
ACCAAAGA	mir-9
TGCATTTA GCATTTAG	mir-277
GTCAATTA	Unknown
CAGTATTA AGTATTAA AAGTATTA	mir-8

The top 26 most conserved 8-mers form 10 motifs, 8 of which correspond to miRNA seeds.

Table S2. The most conserved 8-mer motifs in human coding regions

Highest scoring motifs	Annotation
CTACCTCA TTACCTCA CTACCTCC ACTACCTC CCTACCTC GCTACCTC TCTACCTC CTACCTCG TACCTCAG TACCTCAT	let-7/98
ATGGCGGC ATGGCGGA TGGCGGCG GGCGGCGG GGCGGCGC AGGCGGCG CGGCGGCG	Unknown
GTGCCTTA ATGCCTTA GTGCCTTG AGTGCCTT AAGTGCCT TGCCTTAA	mir-124
ACCAAAGA AACCAAAG TACCAAAG	mir-9
GCACTTTA	mir-17-5p/20/93
CGCCGCCG CCGCCGCC GCCGCCGC GCGCCGCT GCCGTCGG	Unknown
TTAGCTCG	Unknown
CTCAGGGA	mir-125/351
GCGCGCTT	Unknown
TTTGATGA	Unknown
CGCACGCG CGCACTCG	Unknown
GTGCCAAA	mir-96/1271
TGTAAATA	Unknown
AAGCACAA	mir-218
CTATGCAA	mir-153
GAGGTAGG	Unknown
TCGCGCCG	Unknown
AGCAATAA	mir-137

A list of the top 18 highest scoring motifs listed by descending score. miRNA seeds account for 4 of the top 5 and 10 of the top 18 motifs.