EXTENDED EXPERIMENTAL PROCEDURES

Ectopic Pri-miRNA Expression in HEK293 Cells and S2 Cells

A genomic fragment corresponding to the human *mir-1-1* hairpin and flanking sequences was amplified and cloned into both pcDNA3.2/V5-DEST (Invitrogen) and pMT-DEST (Invitrogen) expression plasmids downstream of the attR recombination sites. Query pri-miRNA sequences were recombined into these plasmids at the attR sites using the Gateway system (Invitrogen). Expression plasmids and pMAX-GFP were co-transfected into HEK293 cells using Lipofectamine 2000 (Invitrogen) and co-transfected into S2 cells using Cellfectin (Invitrogen) according to manufacturer's instructions. After 36–48 h, total RNA was collected by addition of Tri-Reagent (Ambion) according to manufacturer's instructions. RNA blots for detecting mature and pre-miRNAs were as described. Ribonuclease protection assays were performed with the RPA III kit (Invitrogen) according to manufacturer's instructions.

For detection of expression by sequencing, total RNA from individual transfections was combined and libraries for small-RNA sequencing prepared as described (Chiang et al., 2010). Sequencing reads were mapped to a miRNA hairpin collection composed of the miRBase-annotated hairpins of miRNAs endogenously expressed in the cell line and the miRBase-annotated hairpins of the transfected miRNAs. Reads were included if they perfectly matched a hairpin in this library and excluded they matched more than one hairpin corresponding to a transfected miRNA. Read counts were normalized to the total reads matching a set of endogenous hairpins that had no transfected counterparts. For each expressed pri-miRNA hairpin, number of reads reported is the number obtained after subtracting the number observed in a normalized, mock-transfected control library.

Microprocessor lysate

Microprocessor lysate was prepared as described (Lee and Kim, 2007), with minor modifications. HEK293T cells were transfected with a mixture of pCK-Drosha-FLAG (Lee and Kim, 2007) pFLAG-HA-DGCR8 (Landthaler et al., 2004), and a transfection-control plasmid pMAX-GFP (Amaxa) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. After 72 h, cells were harvested by rinsing the monolayer in phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HP0_4, [pH 7.4]). Cells were pelleted, resuspended in sonication buffer (100 mM KCl, 0.2 mM EDTA, 20 mM Tris-Cl pH 8.0, and 0.7 µl/ml 2-mercaptoethanol) supplemented with mini-EDTA Free Protease Inhibitor tablets (Roche), and sonicated. After clearing by centrifugation, cell lysis was confirmed by the liberation of GFP into the supernatant. The supernatant was distributed into single-use aliquots, and stored in liquid-nitrogen vapor phase. Pri-miRNA cleavage assays were carried out as described (Lee and Kim, 2007) unless otherwise noted.

Competitive binding and cleavage assays

The competitive binding assay was based on that of Bartel, et al. (Bartel et al., 1991). T7-transcribed ~200 nt pri-miRNA substrates were gel-purified, treated with calf intestinal phosphatase (NEB), extracted in Tri-Reagent (Invitrogen), and 5′ end-labeled using T4 Polynucleotide Kinase (NEB) and γ-[^{32}P]-ATP. The *mir-125a* reference substrate was prepared in the same way, except it was 10–25 nt shorter to enable separation on denaturing gels. Complexes containing Drosha-TN and DGCR8 were immunopurified from Microprocessor lysate in which DroshaTN-FLAG replaced the wild-type Drosha plasmid as described (Lee and Kim, 2007; Han et al., 2009). Competitor and reference RNAs were mixed and incubated with Drosha-TN–DGCR8 for 15-30 min [final concentrations, 250 nM each RNA, 100 mM KCl, 1 mM $MgCl_2$, 0.2 mM EDTA, 20 mM Tris-Cl (pH 8.0), 0.7 µl/ml 2-mercaptoethanol and 300 ng/µl yeast total RNA (Ambion)]. RNA-protein complexes were filtered on Immobilon-NC nitrocellulose discs (Millipore) and washed with at least 10 reaction volumes of sonication buffer. RNA was eluted from the membrane by incubating in elution buffer (300 mM NaCl, 8M urea, and 25 mM EDTA) for 10 min at 85ºC, ethanol precipitated and resolved on denaturing 5% polyacrylamide gels.

For competitive cleavage, 5′ end-labeled query and reference pri-miRNA substrates were mixed and incubated with Microprocessor lysate [final concentrations, 50 nM each RNA, 100 mM KCl, 1 mM $MgCl_2$, 0.2 mM EDTA, 20 mM Tris-Cl (pH 8.0), 0.7 µl/ml 2-mercaptoethanol, 300 ng/µl yeast total RNA, 10 nM Microprocessor complex (concentration estimated exploiting the single-turnover behavior of the Microprocessor when cleaving linear *pri-miR-125a*)]. After incubation for 30 seconds at 37ºC the reaction was stopped by addition of Tri-Reagent (Ambion) with mixing. Extracted RNA was precipitated with isopropanol, then resuspended and resolved on a denaturing 5% polyacrylamide gel.

Synthesis and selection of pri-miRNA variants

Templates for T7 RNA polymerase transcription were assembled from oligonucleotides (IDT) that were synthesized using nucleoside phosphoramidite mixtures to introduce variability at specified positions (Table S1). For body labeling, transcription reactions included α-[^{32}P]-UTP.

Transcripts for producing circular pri-miRNA variants ended with a minimal HDV ribozyme (Schurer et al., 2002) that co-transcriptionally self-cleaved at a defined position to produce homogenous 3′ ends. After treatment with TurboDNAse (Ambion), these transcripts were gel-purified, treated with calf intestinal phosphatase (NEB) to remove the 5′ triphosphate, extracted with Tri-Reagent, precipitated with isopropanol, and treated with T4 polynucleotide kinase (NEB) to remove the 2′-3′ cyclic phosphate as described (Guo et al., 2010). After ethanol precipitation, they were 5′ phosphorylated with T4 polynucleotide kinase, diluted, and circularized using T4 RNA ligase 1 (NEB). Circular pri-miRNAs were purified from linear species on denaturing polyacrylamide gels.

Pools of variants were incubated in Microprocessor lysate, and at one or two time points (for circularized pri-miRNA variants, 1 minute for *mir-125a*, 1 and 4 minutes for *mir-16-1*, 1 and 5 minutes for *mir-30a*, and 3 and 15 minutes for mir-223; for apical stem and loop variants, 5 seconds and 15 seconds for *mir-125a*, 15 seconds and 2 minutes for *mir-16-1*, 30 seconds and 2 minutes for *mir-30a*, and 30 seconds and 2 minutes for *mir-223*) reactions were stopped by addition of Tri-Reagent (Ambion) with mixing, and cleaved products were purified on denaturing gels. Cleavage products of circularized pri-miRNA variants were ligated to oligonucleotide adaptors containing barcode sequences using T4 DNA ligase (NEB) and DNA splints (Table S1), reverse transcribed, and amplified. To represent the initial pools, a sample of phosphorylated, uncircularized RNA was reverse transcribed and amplified. For the apical stem-loop variants, pre-miRNA cleavage products of linear pri-miRNA variants were reverse transcribed, amplified and sequenced. To represent the initial pools of these variants, a sample of each pool was taken and reverse transcribed and amplified.

High-throughput sequencing and analysis

Amplicons from the initial pools and the cleaved products were pooled for Illumina paired-end sequencing (75 nt reads per end) for circularized substrate selections, and Illumina single-read sequencing (54 nt reads) for apical stem-loop selections. Sequencing reads were divided into experimental groups according to constant sequences specific to each pri-miRNA and barcodes indicating time points. After filtering for sequencing quality, discarding any sequences that had an error rate ≥0.1 (phred score ≤10) at any variant position, the sequencing error averaged <0.001 per variant position (average phred score >30). Sequences in which the length of a partially randomized region differed from that of the wildtype were also discarded, thereby eliminating many sequences with insertions or deletions. Libraries were collapsed so that sequences that appeared multiple times with the same bar code were considered just once in the analysis (although in retrospect this precaution was not required because there was no group of dominant, multi-copy sequences that would have biased the analyses). Analyses were also restricted to products cleaved at the wild-type processing sites, which were inferred from the dominant reads in small-RNA sequencing data (Landgraf et al., 2007; Bar et al., 2008; Chiang et al., 2010; Witten et al., 2010), except for miR-16-1* and miR-223, which appear to undergo post-cleavage 3′-end trimming (Han et al., 2011).

To calculate the information content at each position, we used the data from the initial sequences and the product sequences to calculate the relative cleavage of each base versus that of the other three bases. For example, for the A residue, the three relative cleavage values are given below, where $P(N)$ is estimated by the frequency of a base in the initial pool, and $P(N|\text{cleavage})$ is estimated by the frequency of that base in the product sequences.

$$\frac{P(\text{cleavage}|C)}{P(\text{cleavage}|A)} = \frac{P(C|\text{cleavage})}{P(A|\text{cleavage})} \Big/ \frac{P(C)}{P(A)}$$

$$\frac{P(\text{cleavage}|G)}{P(\text{cleavage}|A)} = \frac{P(G|\text{cleavage})}{P(A|\text{cleavage})} \Big/ \frac{P(G)}{P(A)}$$

$$\frac{P(\text{cleavage}|U)}{P(\text{cleavage}|A)} = \frac{P(U|\text{cleavage})}{P(A|\text{cleavage})} \Big/ \frac{P(U)}{P(A)}$$

We then used Bayes' Theorem (Pitman, 1993) to infer the nucleotide composition that would have resulted after selection from a pool of variants in which there was an equal probability of an A, C, G, or U at this position. For example, the formula to infer the frequency of A at a particular position after selection from such a pool was

$$P_{\text{inferred A}} = P(A|\text{cleavage}) = \left[1 + \frac{P(\text{cleavage}|C)}{P(\text{cleavage}|A)} + \frac{P(\text{cleavage}|G)}{P(\text{cleavage}|A)} + \frac{P(\text{cleavage}|U)}{P(\text{cleavage}|A)}\right]^{-1}$$

The inferred post-selection distribution was then used to calculate information content scores for each nucleotide at each position. For example, the information content for A at a particular position was calculated as

$$I_A = P_{\text{inferred A}} \times [log_2(P_{\text{inferred A}}) + 2]$$

If results from two time points were available, information content values were averaged.

For evaluating motifs, we calculated a relative cleavage value based on the frequencies of the motif in the reference and selected pools [P(motif$_i$) and P(motif$_i$)|cleavage), respectively], and the frequencies of a reference motif in the reference and selected pools [P(motif$_{\text{ref}}$) and P(motif$_{\text{ref}}$)|cleavage), respectively].

$$\text{Relative cleavage} = \frac{P(\text{motif}_i|\text{cleavage})}{P(\text{motif}_{\text{ref}}|\text{cleavage})} \Big/ \frac{P(\text{motif}_i)}{P(\text{motif}_{\text{ref}})}$$

We also used an odds ratio score to calculate the enrichment for particular motifs by using the frequency of the motif in the reference and selected pools [*P*(motif$_i$) and *P*(motif$_i$)|cleavage), respectively].

$$\text{Odds ratio} = \frac{P(\text{motif}_i\,|\text{cleavage})}{1 - P(\text{motif}_i\,|\text{cleavage})} \bigg/ \frac{P(\text{motif}_i)}{1 - P(\text{motif}_i)}$$

If two timepoints were available, the geometric mean of the ratios was reported, unless noted otherwise.

To screen for specifically for Watson–Crick pairing between all possible combinations of randomized positions, we used a scoring metric to compare the geometric average of odds ratios for Watson–Crick pairing to that of odds ratios for non-Watson–Crick pairs.

$$\text{Pairing score} = \left(\prod_{\text{Watson-Crick}} \text{Odds ratio}\right)^{1/4} - \left(\prod_{\text{non-Watson-Crick}} \text{Odds ratio}\right)^{1/12}$$

Pri-miRNA collections and positional enrichments of sequence motifs

A list of representative pri-miRNAs used for analyses is provided (Table S2). Because of the large number of questionable annotations in miRBase (Chiang et al., 2010), analysis of human pri-miRNAs was restricted to those of miRNAs conserved in mouse. Coordinates of miRNA loci in miRBase version 17 (Kozomara and Griffiths-Jones, 2011) were used to extract the sequences of each annotated hairpin and 200 genomic bases flanking each side. miRBase hairpin sequences and flanking genomic sequences (20 nt on each side) were folded using RNAFold (Hofacker and Stadler, 2006). The Microprocessor cleavage site was inferred using the predicted structures and the mature sequences annotated in miRBase. In cases in which the 3′ overhang was shorter than 2 nt, the 3′ product was extended to generate a 2 nt overhang. Only hairpins for which the predicted folding and the annotated mature sequences could be reconciled or extended to form a 2 nt 3′ overhang were carried forward for analysis. For hairpins in miRBase-annotated miRNA families, a single representative was chosen to represent the family in each species. For human, *D. melanogaster*, and *C. elegans*, the family member with the most conserved pre-miRNA sequence (as determined by average branch-length score of pre-miRNA nucleotides) was chosen. For other species, the representative was chosen at random.

Whole-genome alignments and phylogenetic trees were obtained from the UCSC genome browser (Fujita et al., 2011). Conservation of a base was evaluated by its branch-length score, defined as the ratio between the total branch length of the species that contained the same base as the reference sequence and the total branch length of the species that had an aligned base at that position.

Enrichment of a motif at a set of positions relative to the cleavage site was computed by generating 100,000 cohorts of pri-miRNAs in which the upstream, downstream and pre-miRNA sequences were independently shuffled, preserving dinucleotide frequencies. The numbers of miRNAs that contained a match to the motif in the actual and shuffled cohorts were used to compute an empirical P-value.

Analysis of crosslinked complexes

The *mir-30a* pri-miRNA crosslinking substrate was assembled using T4 RNA ligase 2 (NEB) and a DNA splint to join an in vitro transcribed 5′ fragment to a synthetic 3′ fragment containing a 3′-terminal biotin and a 4-thiouridine within the CNNC motif (Dharmacon). This crosslinking substrate was incubated in Microprocessor lysate and exposed to 1000 mJ of 365 nm UV light in a Stratalinker (Stratagene). For purification of RNA–protein complexes for mass spectrometry, complexes were captured on streptavidin-coated magnetic beads (Invitrogen) and washed twice in Laemmli buffer (4% SDS, 20% glycerol, 125 mM Tris-Cl pH 6.8) and twice in urea buffer (8 M urea, 300 mM NaCl, 25 mM EDTA), then eluted with RNase T1 (Ambion). The eluted complexes were separated on SDS gels, and the corresponding gel slices excised. The complexes were reduced, alkylated and digested with trypsin. After extraction and concentration, peptides were analyzed by HPLC/tandem mass spectrometry using a Waters NanoAcquity UPLC system and a ThermoFisher LTQ linear ion trap mass spectrometer operated in a data-dependent manner. Peptides were identified using SEQUEST and data analyzed with Scaffold (Proteome Software).

For immunoprecipitation, eluted RNA–protein complexes were incubated for 1 h with antibody [either anti-FLAG M2 (Sigma), polyclonal mouse IgG (Millipore), anti-SRp20 (Invitrogen), or anti-9G8 (gift of J. Stévenin)], followed by incubation with protein-G agarose beads (Sigma). After washing the beads three times in at least ten packed-bead volumes of sonication buffer, complexes were separated on SDS gels.

Reanalysis of iCLIP data

Sequencing reads from iCLIP of SRp20 (SRSF3) and SRp75 (SRSF4) were from ArrayExpress (accession ERP000815) (Anko et al., 2012). Although that study did not find enrichment for SRp20-binding sites in miRNAs, the miRNA annotations examined did not extend beyond the miRNA hairpins (i.e., the pre-miRNAs and their basal stems) and thus did not include downstream regions containing the CNNC motif. After adaptor stripping, reads were mapped to the mouse genome using Bowtie (Langmead et al., 2009), allowing for two mismatches and considering only uniquely mapped reads. Each position immediately 5′ to an iCLIP read was considered a crosslink site, and the number of crosslink sites was tallied for each relative distance from the mouse pre-miRNAs confirmed or identified in Chiang et al. (2010). For pri-miRNAs with more than one site, the site supported by the most reads was the one plotted in Figure 6D (distributing fractions of a count to each site in cases in which multiple sites were tied for the most reads).

Cleavage assays with purified SRp20

SRp20 cDNA with an N-terminal 3X-FLAG tag was cloned into the pcDNA3.2-V5-DEST vector (Invitrogen). HEK293T cells were transiently transfected with either the SRp20 construct or an analogous construct expressing N-terminally FLAG-tagged EGFP, using Lipofectamine 2000 (Invitrogen) according to manufacturer’s instructions. After 48 h cells were lysed in sonication buffer. The tagged proteins were immunoprecipitated using ANTI-FLAG M2 magnetic beads (Sigma), washed three times in sonication buffer, and eluted with 150 ng/ul 3X-FLAG peptide (Sigma), then dialyzed against 1000 volumes of sonication buffer using dialysis membrane with a 3 kDa cutoff (Pierce). For cleavage reactions, the Microprocessor complex was first immunoprecipatated from Microprocessor lysate. Biotinylated anti-FLAG-M2 antibody (1:500 dilution, Sigma) was incubated in lysate for 2.5 h, then precipitated with streptavidin-coated magnetic beads (Invitrogen) for 30 minutes. Beads were washed twice in sonication buffer, then incubated at 37ºC with 5′-labeled *pri-mir-16-1* substrates and either SRp20 or EGFP immunopurified from HEK293T cells, at a final volume of 40% beads and 40% either SRp20, EGFP or sonication buffer. Final concentrations were 2 nM pri-miRNA, 100 mM KCl, 1 mM MgCl2, 0.2 mM EDTA, 20 mM Tris-Cl (pH 8.0), 0.7 µl/ml 2-mercaptoethanol, 300 ng/µl yeast total RNA (Ambion) and 0.5% SUPERaseIn RNase Inhibitor (Ambion). After 3 minutes, reactions were stopped in Tri-reagent (Ambion), and products separated on 10% denaturing polyacrylamide gels.

Supplemental Figure Legends

Figure S1. Expression and processing of pri-miRNA hairpins in HEK293 cells, related to Figure 1.

(A) Expression of miR-1 and *pre-mir-1-1* from bicistronic transcripts. HEK293 cells were individually transfected with plasmids bearing a human, *D. melanogaster*, or *C. elegans* pri-miRNAs transcriptionally fused to human *pri-mir-1-1*. Mature miR-1 and *pre-mir-1-1* derived from the transcriptional fusion were detected by RNA blot. (Results from vectors in which *let-7* and *mir-1* were the query pri-miRNAs are shown here but are not shown in Figure 1A because the corresponding mature miRNAs were indistinguishable from those of other transfected vectors after total RNA was pooled for small-RNA sequencing.)

(B) Full membrane images for blots shown in Figure 1B. Total RNA was run on stacked polyacrylamide gels (5% top and 15% bottom) to resolve sizes from 20–1000 nt. Each blot included marker lanes (Century and Decade RNA markers, Ambion) a positive-control lane with 15 fmol *in vitro* transcribed standard derived from the corresponding pri-miRNAs (control).

Figure S2. Confirmation of *hsa-mir-125a* selection results in vitro and in HEK293T cells, related to Figure 2.

(A) Predicted basal stem structure of *mir-125a* variants tested in the experiment.

(B) Competitive cleavage of individual *mir-125a* variants, relative to wild-type *mir-125a*. Variants were mixed with wild-type *mir-125a*, which was longer at its 5′ end, and incubated in Microprocessor lysate. Cleavage products were separated on denaturing gels, and the ratio of wild-type and variant products quantified (blue, geometric mean ± standard error, *n* = 3), together with the relative cleavage inferred from the selection experiment (gray).

(C) Evaluation of *mir-125a* variants in HEK293T cells. Variants were transcriptionally fused to *pri-mir-1-1* and expressed in HEK293T cells, as in Figure S1A. Accumulation of mature miR-125a was quantified by RNA blot and normalized to the level of mature miR-1 (geometric mean ± standard error, $n = 3$).

Figure S3. Analysis of *mir-223* basal stem structure, related to Figure 3.

(A) Wild-type (left) and alternative (right) basal stem structures for *hsa-mir-223*. In the predicted structure of the wild-type the A at +10 is bulged, whereas in the predicted structure of some of the variants the pairing shifts to place nucleotide +10 within a contiguous helix. After sorting the selected variants based on whether or not their predicted secondary structures are consistent with shifted pairing, covariation matrices for both conformations were calculated as in Figure 3A.

(B) Relative cleavage of variants with different lengths of the alternative basal stem. Cleavage values were calculated as in Figure 3B and normalized to the 9 bp stem.

(C) Screen for Watson–Crick pairs involving any two varied positions. For each of the >3000 possible pairs, the degree of Watson–Crick preference was evaluated using a scoring metric that compared the average odds of Watson–Crick pairs to that of non-Watson–Crick alternatives. The number of Watson–Crick candidates is plotted as a function of threshold score, in which a pair is considered a Watson–Crick candidate if its score exceeds the threshold. The number of pairs corresponding to the basal stem is shown (dashed line). In each case, the highest-scoring pairs were those of the basal stem. In the case of *mir-223*, the highest scoring pairs also included the alternative pairs that incorporated the bulged A at +10 into a contiguous helix. For each pri-miRNA, we inspected the next four highest-scoring pairs, and in each case, the covariation matrix did not appear consistent with Watson–Crick pairing (data not shown).

Figure S4. Selection for Microprocessor-binding variants of *hsa-mir-125a*, related to Figure 4.

(A) Schematic of the in vitro selection. Linear variants of *mir-125a* were incubated with immunopurified DGCR8 and catalytically-inactive Drosha (DroshaTN). Bound variants were recovered after nitrocellulose filtration, reverse-transcribed, and amplified for high-throughput sequencing.

(B) Information content after selection for Microprocessor binding. Information content after selection for cleavage (Figure 2D) is reproduced here for comparison. The nucleotides varied in the initial pools are shown for each selection (insets, red inner lines).

Figure S5. Contribution of the CNNC motif in vitro and in HEK293T cells, related to Figure 5.

(A) CNNC odds ratios at alternative positions. Odds ratios were calculated for CNNC dinucleotides starting at the indicated of positions downstream of the Drosha cleavage site. Plotted are odds ratios for all sequences (left panels) and for sequences that lack both wild-type C residues (right panels).

(B) Contributions of the basal UG and downstream CNNC motifs to the accumulation of hsa-miR-30a in HEK293T cells. The listed variants of *hsa-mir-30a* were transcriptionally fused to *hsa-mir-1-1* (top). Predicted secondary structures for variants with non-wild-type structure are shown (center), with the annotated Drosha cleavage sites (purple arrowheads). The accumulation of miR-30a was quantified by RNA blot, normalized to miR-1 (bottom, geometric mean ± standard error, *n* = 3).

(C) Contributions of the basal UG and downstream CNNC motifs to the accumulation of hsa-miR-16 in HEK293T cells, otherwise as in (B).

(D) Contributions of the basal UG and downstream CNNC motifs to the accumulation of hsa-miR-28 in HEK293T cells, otherwise as in (B).

(E) Contributions of the basal UG and downstream CNNC motifs to the accumulation of hsa-miR-129 in HEK293T cells, otherwise as in (B).

(F) Contributions of the basal UG and downstream CNNC motifs to the accumulation of hsa-miR-193b in HEK293T cells, otherwise as in (B).

Figure S6. Immunopurified SRp20, related to Figure 6. 3X-FLAG-SRp20 was expressed in HEK293T cells, captured on anti-FLAG magnetic beads, and eluted with 3X-FLAG peptide. Binding activity of immunopurified SRp20 was measured by crosslinking to a *mir-30a* crosslinking substrate as in Figure 6A, except that the

substrate contained only the mir-30a CNNC motif (pCU(4-S-U)CAAGGG). For comparison, crosslinked complexes were generated from Microprocessor lysate (left).

Figure S7. Selection of functional variants with changes in the apical stem-loop, and rescued processing of *C. elegans* pri-miRNAs in human cells, related to Figure 7.

(A) Schematic of the selection for functional pri-miRNA variants with changes in the apical stem and terminal loop. Linear pri-miRNA variants were incubated in Microprocessor lysate, and cleaved pre-miRNA variants were gel-purified, reverse transcribed, and amplified for high-throughput sequencing.

(B) Relative cleavage of variants with different apical stem lengths. The number of contiguous Watson–Crick pairs was counted and the relative cleavage calculated, normalized to that of the 15 bp stem. For each pri-miRNA, results are shown for both time points (key). For *mir-125a*, 22 bp above the 5p Drosha cleavage site was strongly preferred; longer stems were tolerated, whereas shorter stems were disfavored. Watson–Crick pairing throughout the apical stem was supported by analysis of covariation (data not shown). A 22-pair apical stem was also preferred, albeit more weakly, in *mir-30a*. By contrast, no preference for apical pairing was observed in the stems of *mir-16-1* and *mir-223*. Indeed, lengthening of the *mir-16-1* apical stem at the expense of loop size was detrimental, which was consistent with a previous report (Zhang and Zeng, 2010).

(C) Enrichment and depletion at variable residues in the apical stems and loops. At each varied position (inset, red inner line), information content was calculated for each residue (green, cyan, black, and red for A, C, G, and U, respectively), as in Figure 2D.

(D) Relative cleavage of *mir-30a* variants with the apical UGUG motif beginning at the indicated positions, normalized to variants without the motif. Nucleotides of the mature miRNA are shaded in yellow.

(E) Conservation of the region centered on the apical UGUG of *mir-30a*, otherwise as in Figure 4B.

(F) Enrichment for UGU or GUG trinucleotides in the terminal loops of metazoan pri-miRNAs (Table S2). For each species, pri-miRNA sequences were aligned on the predicted Drosha cleavage site and occurrences of loop UGU or GUG trinucleotides tabulated. Species with a statistically significant enrichment within the indicated window are indicated (asterisk, empirical *p*-value $<10^{-4}$). Although enrichment was observed in fish and insects, the lack of enrichment in several other representative species raises the question of whether the usage of this motif arose independently in multiple lineages or was ancestral and lost multiple times.

(G) Effects of adding human pri-miRNA features to *C. elegans mir-50*. Changes that introduced the listed features were incorporated into *mir-50* within the bicistronic expression vector (left). Secondary structures are shown for changes that were predicted to affect the wild-type basal stem (middle; annotated Drosha cleavage sites, purple arrowheads). After transfection into HEK293T cells, accumulation of miR-50 was assessed on RNA blots, normalizing to the accumulation of the miR-1 control, and increased miR-50 expression is plotted (right; geometric mean ± standard error, n = 3).

(H) Effects of adding human pri-miRNA features to *C. elegans mir-40*, otherwise as in (G).

Supplemental References

Anko, M.L., Muller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J., and Neugebauer, K.M. (2012). The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. Genome Biol *13*, R17.

Bar, M., Wyman, S.K., Fritz, B.R., Qi, J., Garg, K.S., Parkin, R.K., Kroh, E.M., Bendoraite, A., Mitchell, P.S., Nelson, A.M.*, et al.* (2008). MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells *26*, 2496-2505.

Bartel, D.P., Zapp, M.L., Green, M.R., and Szostak, J.W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell *67*, 529-536.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E.*, et al.* (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev *24*, 992-1009.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A.*, et al.* (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res *39*, D876-882.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Han, B.W., Hung, J.H., Weng, Z., Zamore, P.D., and Ameres, S.L. (2011). The 3'-to-5' Exoribonuclease Nibbler Shapes the 3' Ends of MicroRNAs Bound to Drosophila Argonaute1. Curr Biol.

Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.K., Yeom, K.H., Yang, W.Y., Haussler, D., Blelloch, R., and Kim, V.N. (2009). Posttranscriptional crossregulation between Drosha and DGCR8. Cell *136*, 75-84.

Hofacker, I.L., and Stadler, P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics *22*, 1172-1176.

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res *39*, D152-157.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M.*, et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401-1414.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol *14*, 2162-2167.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lee, Y., and Kim, V.N. (2007). In vitro and in vivo assays for the activity of Drosha complex. Methods Enzymol *427*, 89-106.

Moore, M.J. (1999). Joining RNA molecules with T4 DNA ligase. Methods Mol Biol *118*, 11-19.

Pitman, J. (1993). Probability (New York, Springer-Verlag).

Schurer, H., Lang, K., Schuster, J., and Morl, M. (2002). A universal method to produce in vitro transcripts with homogeneous 3' ends. Nucleic Acids Res *30*, e56.

Sontheimer, E.J. (1994). Site-specific RNA crosslinking with 4-thiouridine. Mol Biol Rep *20*, 35-44.

Witten, D., Tibshirani, R., Gu, S.G., Fire, A., and Lui, W.O. (2010). Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. BMC Biol *8*, 58.

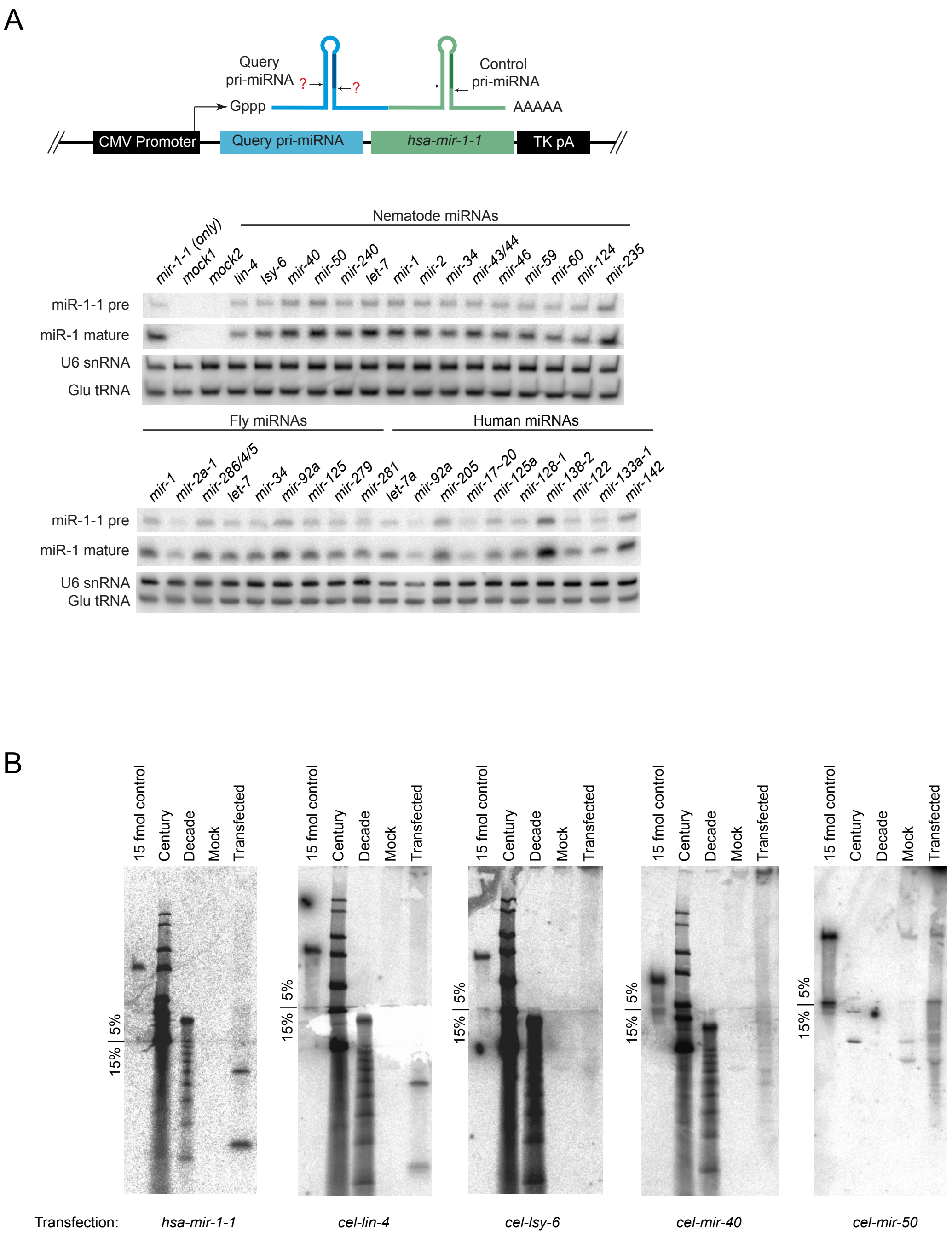
Zhang, X., and Zeng, Y. (2010). The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res *38*, 7689-7697.

Supplemental Figure S1

Click here to download Supplemental Figure: Figure S1.pdf

Auyeung, et al.

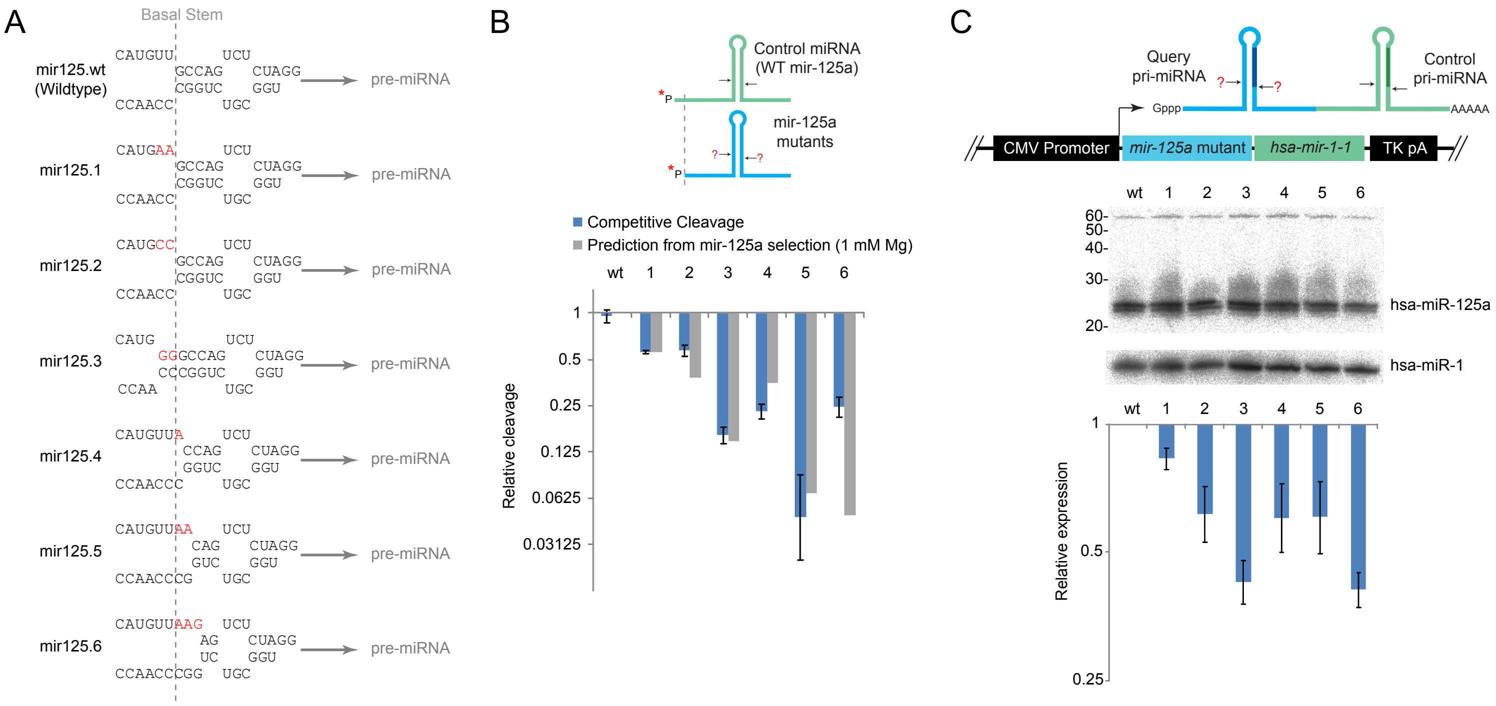
Figure S1



Supplemental Figure S2

Click here to download Supplemental Figure: Figure S2.pdf

Auyeung et al
Figure S2

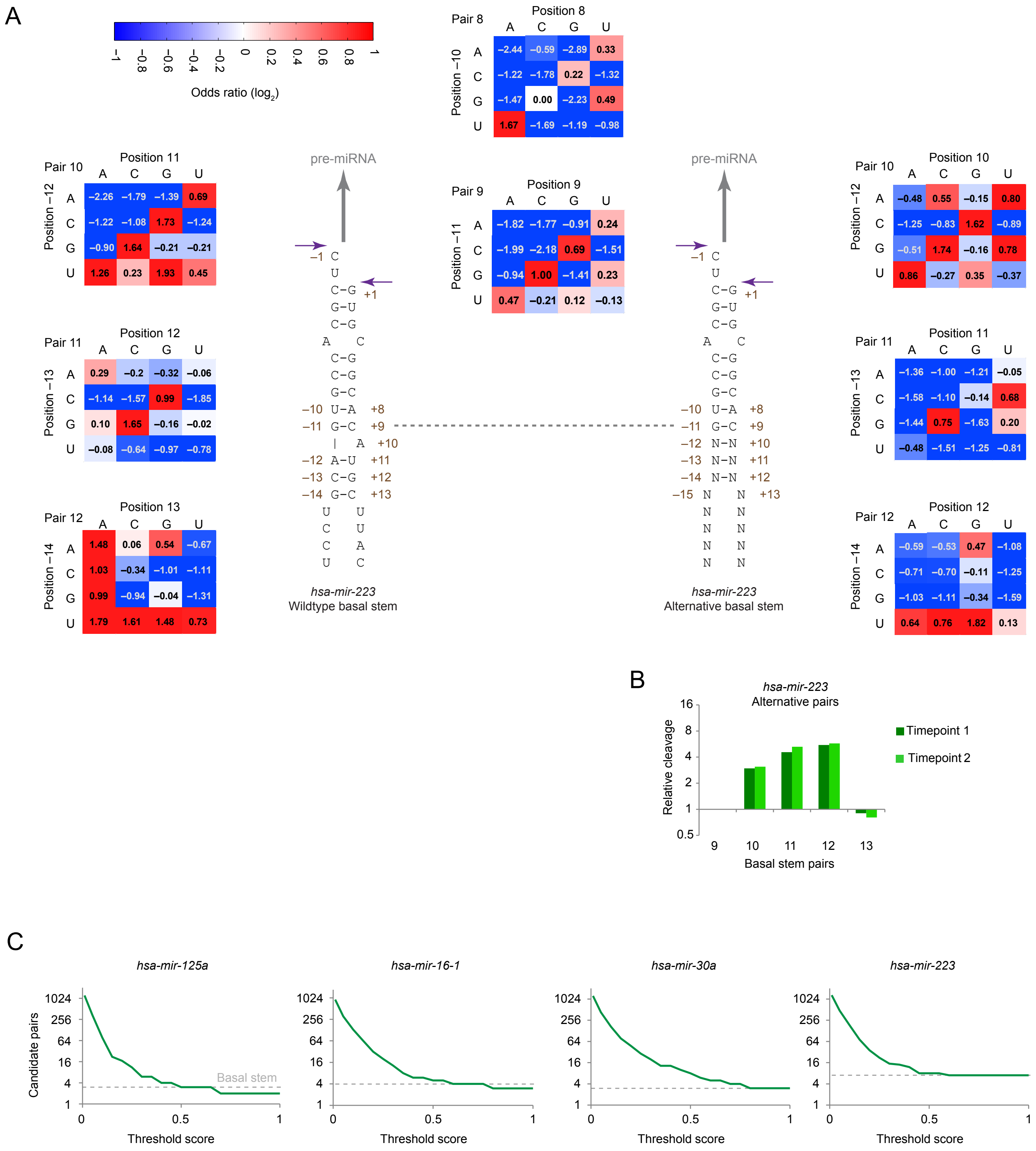


Supplemental Figure S3

Click here to download Supplemental Figure: Figure S3.pdf

Auyeung, et al

Figure S3

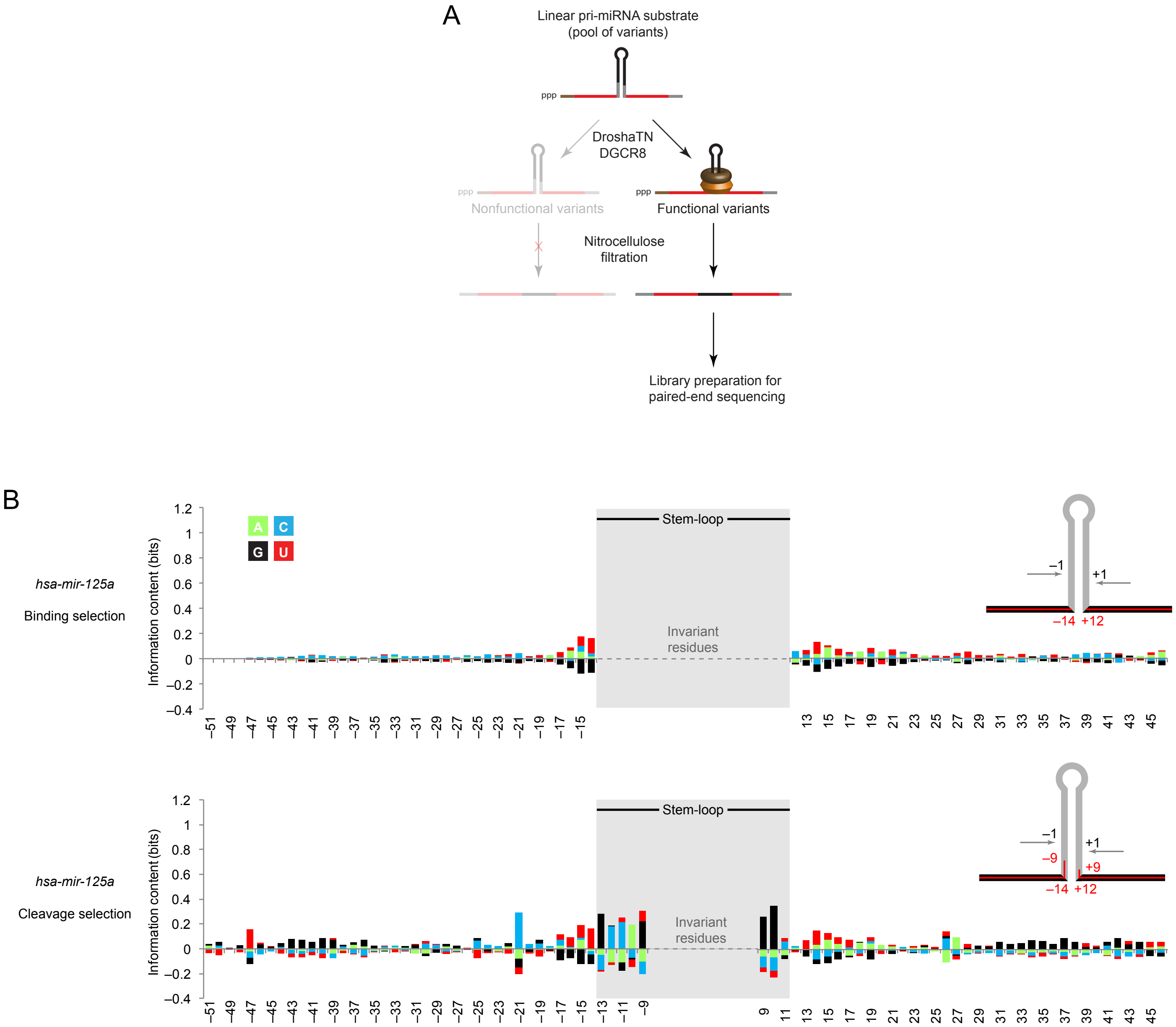


Supplemental Figure S4

Click here to download Supplemental Figure: Figure S4.pdf

Auyeung et al

Figure S4

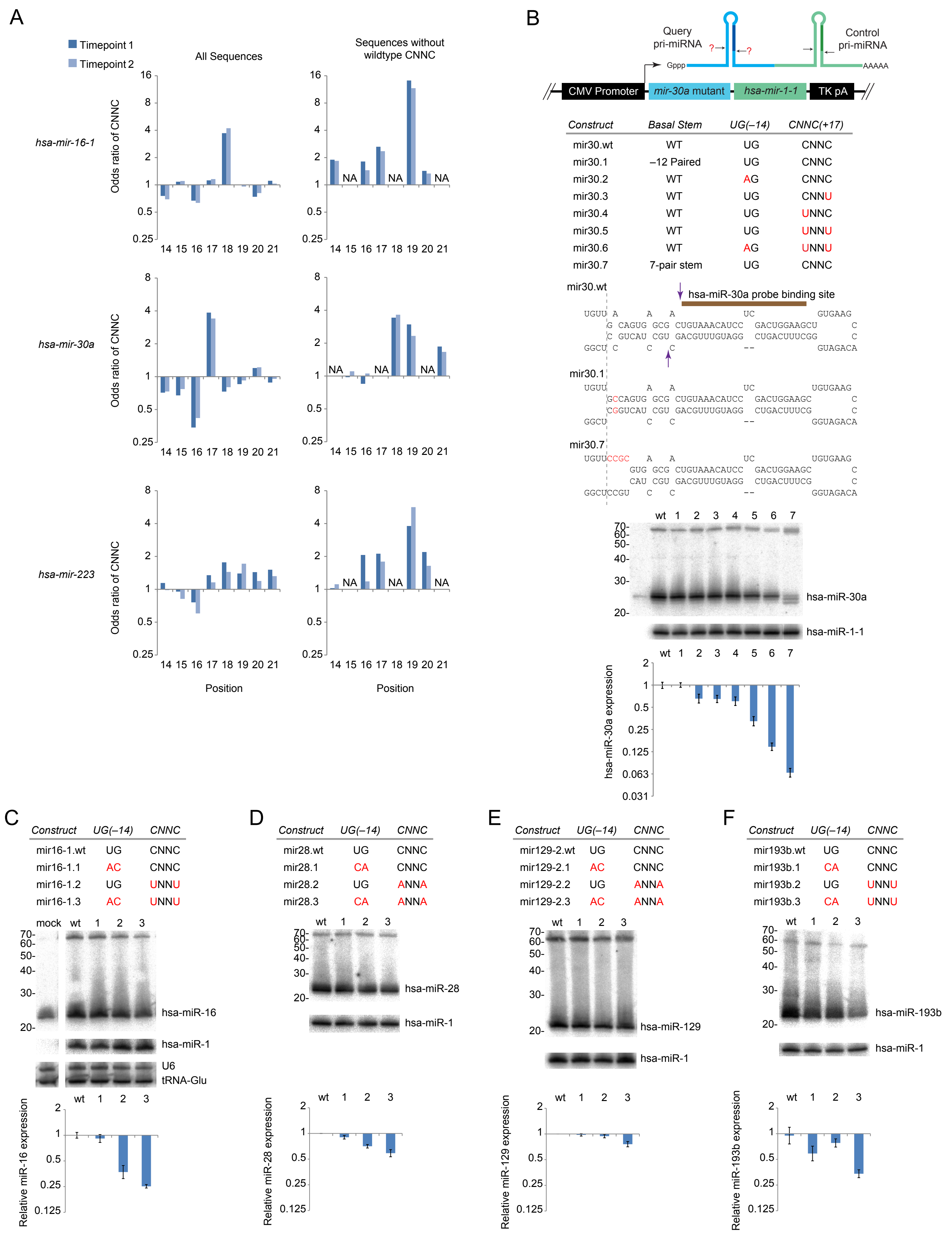


Supplemental Figure S5

Auyeung, et al.

Click here to download Supplemental Figure: Figure S5.pdf

Fig. S5



Construct	Basal Stem	UG(–14)	CNNC(+17)
mir30.wt	WT	UG	CNNC
mir30.1	–12 Paired	UG	CNNC
mir30.2	WT	AG	CNNC
mir30.3	WT	UG	CNNU
mir30.4	WT	UG	UNNC
mir30.5	WT	UG	UNNU
mir30.6	WT	AG	UNNU
mir30.7	7-pair stem	UG	CNNC

mir30.wt

hsa-miR-30a probe binding site

```
UGUU A     A   A           UC           GUGAAG
    G CAGUG GCG CUGUAAACAUCC GACUGGAAGCU      C
    C GUCAU CGU GACGUUUGUAGG CUGACUUUCGG      C
GGCU C     C   C           --           GUAGACA
```

mir30.1

```
UGUU       A   A           UC          UGUGAAG
    GCCAGUG GCG CUGUAAACAUCC GACUGGAAGC       C
    CGGUCAU CGU GACGUUUGUAGG CUGACUUUCG       C
GGCU       C   C           --          GGUAGACA
```

mir30.7

```
UGUUCCGC   A   A           UC          UGUGAAG
        GUG GCG CUGUAAACAUCC GACUGGAAGC       C
        CAU CGU GACGUUUGUAGG CUGACUUUCG       C
GGCUCCGU   C   C           --          GGUAGACA
```

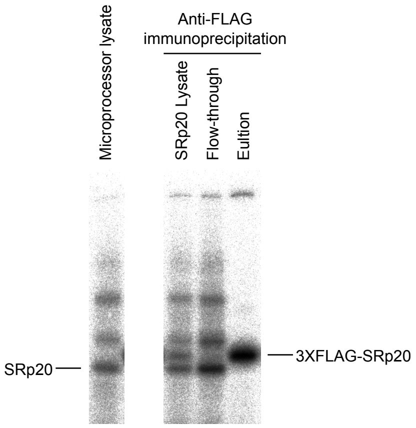
Construct	UG(–14)	CNNC
mir16-1.wt	UG	CNNC
mir16-1.1	AC	CNNC
mir16-1.2	UG	UNNU
mir16-1.3	AC	UNNU

Construct	UG(–14)	CNNC
mir28.wt	UG	CNNC
mir28.1	CA	CNNC
mir28.2	UG	ANNA
mir28.3	CA	ANNA

Construct	UG(–14)	CNNC
mir129-2.wt	UG	CNNC
mir129-2.1	AC	CNNC
mir129-2.2	UG	ANNA
mir129-2.3	AC	ANNA

Construct	UG(–14)	CNNC
mir193b.wt	UG	CNNC
mir193b.1	CA	CNNC
mir193b.2	UG	UNNU
mir193b.3	CA	UNNU

Auyeung et al
Figure S6

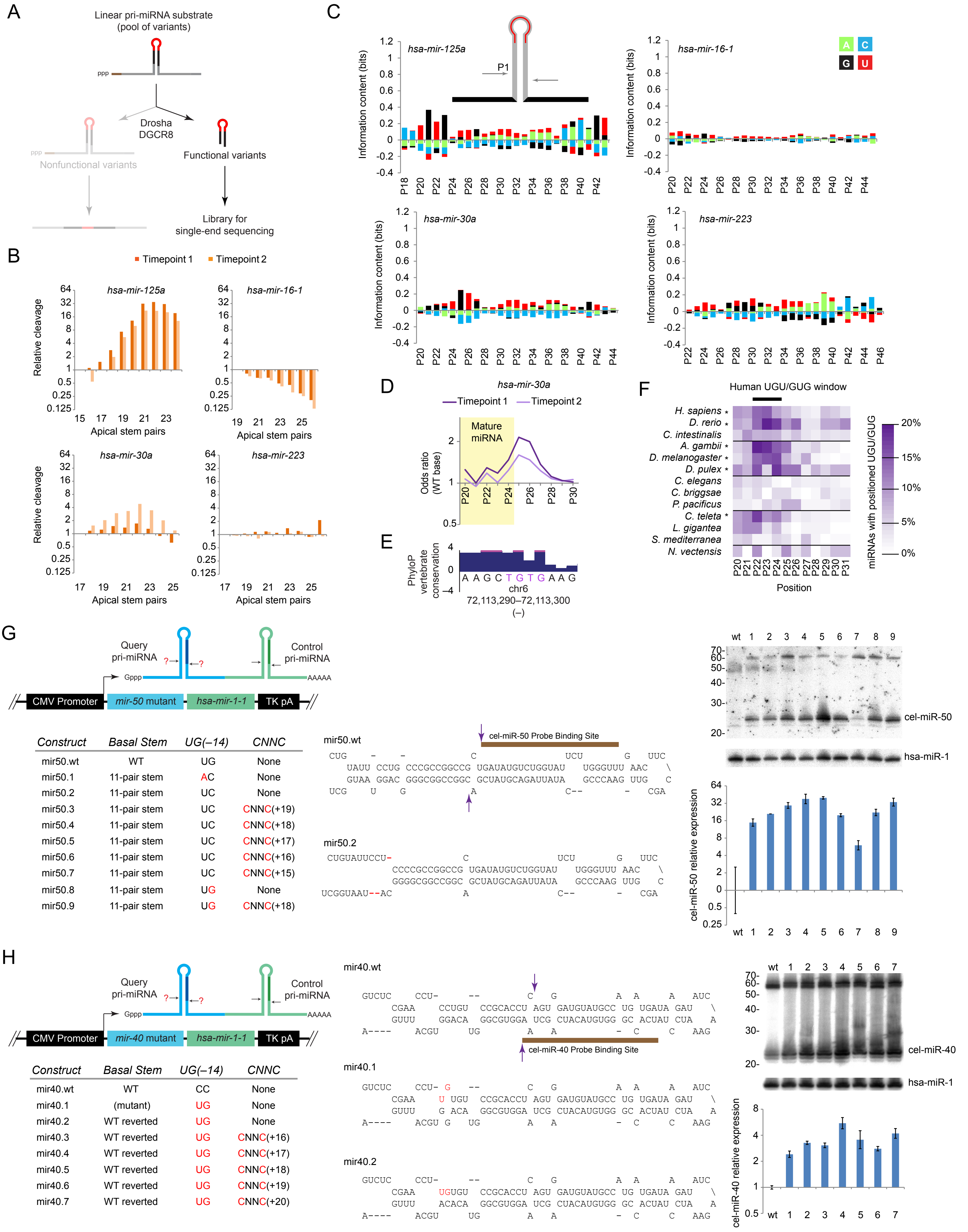


Supplemental Figure S7

Auyeung et al

Click here to download Supplemental Figure: Figure S7.pdf

Figure S7



Construct	Basal Stem	UG(–14)	CNNC
mir50.wt	WT	UG	None
mir50.1	11-pair stem	AC	None
mir50.2	11-pair stem	UC	None
mir50.3	11-pair stem	UC	CNNC(+19)
mir50.4	11-pair stem	UC	CNNC(+18)
mir50.5	11-pair stem	UC	CNNC(+17)
mir50.6	11-pair stem	UC	CNNC(+16)
mir50.7	11-pair stem	UC	CNNC(+15)
mir50.8	11-pair stem	UG	None
mir50.9	11-pair stem	UG	CNNC(+18)

mir50.wt

```
CUG    -    -          C             UCU      G   UUC
   UAUU CCUG CCCGCCGGCCG UGAUAUGUCUGGUAU   UGGGUUU AAC   \
   GUAA GGAC GGGCGGCCGGC GCUAUGCAGAUUAUA   GCCCAAG UUG   C
UCG    U    G          A             C--      -   CGA
```

mir50.2

```
CUGUAUUCCU-            C             UCU      G   UUC
           CCCCGCCGGCCG UGAUAUGUCUGGUAU   UGGGUUU AAC   \
           GGGGCGGCCGGC GCUAUGCAGAUUAUA   GCCCAAG UUG   C
UCGGUAAU--AC           A             C--      -   CGA
```

Construct	Basal Stem	UG(–14)	CNNC
mir40.wt	WT	CC	None
mir40.1	(mutant)	UG	None
mir40.2	WT reverted	UG	None
mir40.3	WT reverted	UG	CNNC(+16)
mir40.4	WT reverted	UG	CNNC(+17)
mir40.5	WT reverted	UG	CNNC(+18)
mir40.6	WT reverted	UG	CNNC(+19)
mir40.7	WT reverted	UG	CNNC(+20)

mir40.wt

```
GUCUC    CCU-    --       C   G          A  A     A   AUC
     CGAA    CCUGU  CCGCACCU AGU GAUGUAUGCC UG UGAUA GAU   \
     GUUU    GGACA  GGCGUGGA UCG CUACAUGUGG GC ACUAU CUA   A
A----    ACGU    UG       A   A          -  C     C   AAG
```

mir40.1

```
GUCUC    CCU- G  --       C   G          A  A     A   AUC
     CGAA    U UGU  CCGCACCU AGU GAUGUAUGCC UG UGAUA GAU   \
     GUUU    G ACA  GGCGUGGA UCG CUACAUGUGG GC ACUAU CUA   A
A----    ACGU G  UG       A   A          -  C     C   AAG
```

mir40.2

```
GUCUC    CCU-    --       C   G          A  A     A   AUC
     CGAA    UGUGU  CCGCACCU AGU GAUGUAUGCC UG UGAUA GAU   \
     GUUU    ACACA  GGCGUGGA UCG CUACAUGUGG GC ACUAU CUA   A
A----    ACGU    UG       A   A          -  C     C   AAG
```