Supplemental Information

Time-resolved small RNA sequencing unravels the molecular principles of microRNA homeostasis

Brian Reichholf[1], Veronika A. Herzog[1], Nina Fasching, Raphael A. Manzenreither, Ivica Sowemimo, and Stefan L. Ameres*

Institute of Molecular Biotechnology (IMBA), Vienna BioCenter (VBC), 1030 Vienna, Austria.

[1] These authors contributed equally
* Lead contact and corresponding author (stefan.ameres@imba.oeaw.ac.at)

Supplemental Information includes **one Supplemental Table** and **six Supplemental Figures**.

Supplementary Table 1l Oligonucleotides used in this study. Related to STAR Methods.

Guide RNA target sequence used for CRISPR/Cas9 genome editing (PAM sequences highlighted in bold).

Name	DNA sequence (5′-3′)
gRNA Nbr KO #1	TATACATGTTCCACTTCTCG**CGG**
gRNA Nbr KO #2	TCTGCGCCAGGAATTTGGAA**GGG**
gRNA Nbr KO #3	ACTTGGGAAGCAGTAGGTCG**CGG**
gRNA Nbr KO #4	GATTTCCCGCATCTTGGCG**AGG**
gRNA Ago2 KO #1	AGCAAAAGCAACAACAACAA**CGG**
gRNA Ago2 KO #2	TGTGGTTGTTGCAGCTGCTG**TGG**

Synthetic RNA oligonucleotides used for quantification of miRNA molecules per cell.

Name	RNA sequence (5′-3′)
synthetic miR-184-3p	UGGACGGAGAACUGAUAAGGGC
synthetic *bantam*-3p	UGAGAUCAUUUUGAAAGCUGAUU

Probes for Northern hybridization.

Name	DNA sequence (5′-3′)
miR-34-5p (21-mer)	AACCAGCTAACCACACTGCCA
bantam-3p (23-mer)	AATCAGCTTTCAAAATGATCTCA
miR-184-3p (22-mer)	GCCCTTATCAGTTCTCCGTCCA
2S rRNA	TACAACCCTCAACCATATGTAGTCCAAGCA

Supplemental Figures

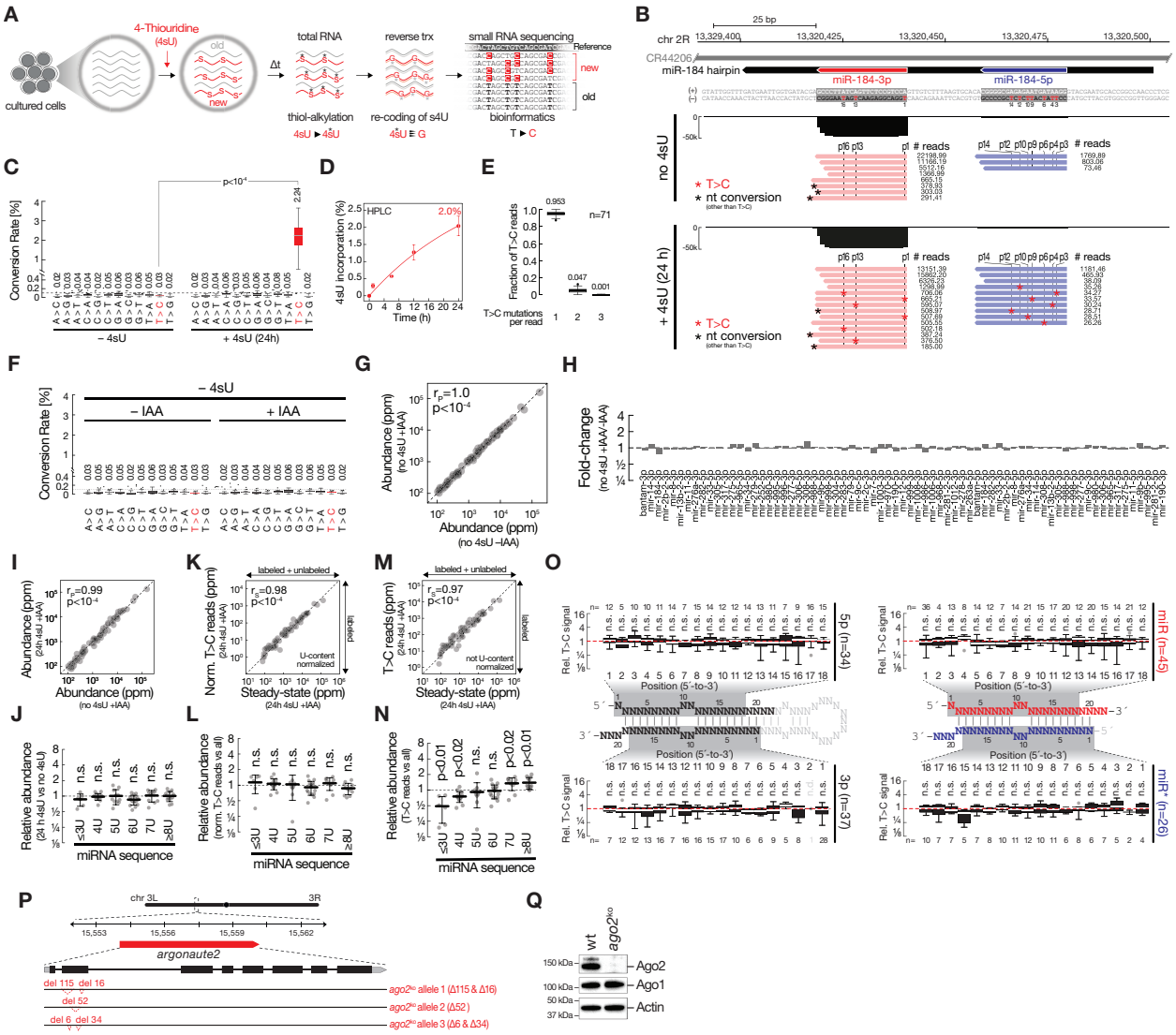

Figure S1. Thiol-linked alkylation for the metabolic sequencing (SLAMseq) of small RNAs in S2 cells. Related to STAR Methods. **(A)** Schematic overview of SLAMseq. Cells are treated with 4-thiouridine (4sU), which – upon cellular uptake – incorporates into newly transcribed RNA. Following 4sU labelling for different time periods and total RNA preparation, 4sU-residues present in newly generated RNA species are carboxyamidomethylated by treatment with iodoacetamide (IAA), resulting in altered base-pairing (i.e. alkylated 4sU preferentially base-pairs to guanine instead of adenine). When combined with well-established small RNA library preparation protocols, the presence of the bulky group at the sites of 4sU-incorporation results in the specific and quantitative incorporation of G across alkylated 4sU during reverse transcription (reverse trx, RT). 4sU-containing sites can be identified bioinformatically in high-throughput sequencing libraries at single nucleotide resolution by calling T-to-C conversions. **(B)** Representative genome browser screen shot for small RNA libraries generated from size-selected total RNA of *Drosophila* S2 cells. A representative area in the *Drosophila melanogaster* genome encoding miR-184 is shown. Nucleotide positions encoding thymine (T, red) are indicated relative to the 5´ end of each small RNA species. Reads representing 99% of all 5´ isoforms are shown for miR-184-3p and -5p and the respective number of reads are indicated in parts per million (ppm). **(C)** Small RNA sequencing libraries generated from total RNA of wild-type *Drosophila* S2 cells before and after 4sU metabolic labeling for 24 h were

mapped to annotated miRNAs and any given conversion rate was determined for abundantly expressed miRNAs (>100 ppm). Tukey boxplots show nucleotide-composition normalized, mean conversions per miRNA (in percent), as determined across positions 1-18. Outliers are not shown. Median observed frequency for each individual conversion are indicated. P-value, as determined by Mann-Whitney test is indicated. **(D)** Quantification of 4sU-incorporation into total RNA after 4sU-metabolic labeling for the indicated time in a pulse labeling experiment in wild-type *Drosophila* S2 cells. Substitution rate of 4sU compared to unmodified uridine determined by HPLC is shown. Values represent mean ± stdev of three independent replicates. Maximum incorporation rate after 24 h labeling is indicated. **(E)** Tukey boxplots show the fraction of T>C conversion containing reads with one, two, or three T>C conversions for each of 71 abundantly expressed (>100 ppm) miRNAs in small RNA libraries prepared from size selected total RNA of wild-type *Drosophila* S2 cells subjected to 4sU metabolic labeling for 24 h. The median fraction of T>C conversion containing reads is indicated. **(F)** Any given conversion rate in small RNA sequencing libraries generated from total RNA of wild-type *Drosophila* S2 cells before and after treatment with iodoacetamide was determined for each abundantly expressed miRNA(>100 ppm). Tukey boxplots show conversions per miRNA in percent. Outliers are not shown. Median observed frequency for each individual conversion is indicated. **(G)** Abundance of miRNAs in small RNA libraries prepared from iodoacetamide-treated or untreated total RNA. Pearson correlation coefficient (r_P) and associated p-value is indicated. **(H)** Fold-change in expression for individual miRNAs in small RNA libraries prepared from iodoacetamide-treated relative to untreated total RNA. **(I-J)** Steady-state abundance of miRNAs is not impacted by 4sU labeling. **(I)** Abundance of miRNAs (>100 ppm, n=71) in small RNA libraries prepared from 4sU labeled (24h 4sU+IAA) or non-labeled (no 4sU+IAA) wild-type *Drosophila* S2 cells. Pearson correlation coefficient (r_P) and associated p-value is indicated. Dashed line represents *x=y*. **(J)** Abundantly expressed miRNAs (>100 ppm, n=71) were grouped according to their U content in the mature miRNA sequence ($n_{\leq 3U}$=6, n_{4U}=15, n_{5U}=11, n_{6U}=17, n_{7U}=10, $n_{\geq 8U}$=12) and their relative abundance in small RNA libraries prepared from 4sU labeled (24h 4sU+IAA) versus non-labeled (no 4sU+IAA) wild-type *Drosophila* S2 cells was analyzed. P-values were determined using Kruskal-Wallis test comparing each subgroup to all 71 miRNAs (n.s., p>0.05). **(K)** Correlation of T>C containing reads (in parts per million, ppm) and all reads (steady-state, in ppm) at steady-state labeling conditions (24h 4sU) of abundantly expressed miRNAs (>100 ppm, n=71) in a small RNA SLAMseq experiment in wild-type *Drosophila* S2 cells. T>C reads were U-content and 4sU-labeling-efficiency normalized (norm. ppm). Spearman correlation coefficient (r_S) and associated p-value is indicated. **(L)** MicroRNAs were grouped according to their U-content as in (B) and the relative abundance of 4sU-labeling-efficiency and U-content normalized T>C reads compared to all reads was analyzed. P-values were determined using Kruskal-Wallis test comparing each subgroup for over- or underrepresentation relative to all 71 miRNAs (n.s., p>0.05). **(M)** Correlation of T>C containing reads (in ppm) and all reads (steady-state, in ppm) at steady-state labeling conditions (24h 4sU) as in (C), but T>C reads were not U-content normalized. Spearman correlation coefficient (r_S) and associated p-value is indicated. **(N)** MicroRNAs were grouped according to their U-content as in (B) and the relative abundance of 4sU-labeling-efficiency normalized T>C reads compared to all reads was analyzed. T>C reads were not U-content normalized. P-values were determined using Kruskal-Wallis test comparing each subgroup for over- or underrepresentation relative to all 71 miRNAs (n.s., p>0.05). **(O)** Over- or underrepresentation of T>C conversions at individual positions of small RNAs that are derived from the 5p- or 3p arm of a miRNA precursor (left), or that constitutes a miR or miR* strand, as defined by selective Argonaute-loading (right). Results are derived from 71 abundantly expressed (>100 ppm) miRNAs (corresponding to 34 5p- and 37 3p-miRNAs, or 45 miR and 26 miR*). Statistically significant differences in relative representation were compared to the total population for the indicated position by Kruskal-Wallis test. n.s., p>0.05; n.d., not determined due to limited data points. **(P)** Schematic representation of the *argonaute 2* (*ago2*) locus in the *Drosophila melanogaster* genome. Zoom-in shows the *ago2* gene (UTRs in grey, exons in black, intron as black line). Disruptive frameshift deletions (del) on three alleles of the *ago2*^{ko} clone used in all following experiments are shown. **(Q)** Western blot analysis of wild-type (wt) S2 cells, or S2 cells stably depleted of Ago2 by CRISPR/Cas9 genome engineering (*ago2*^{ko}). Ago1 levels were unaffected. Actin represents loading control.

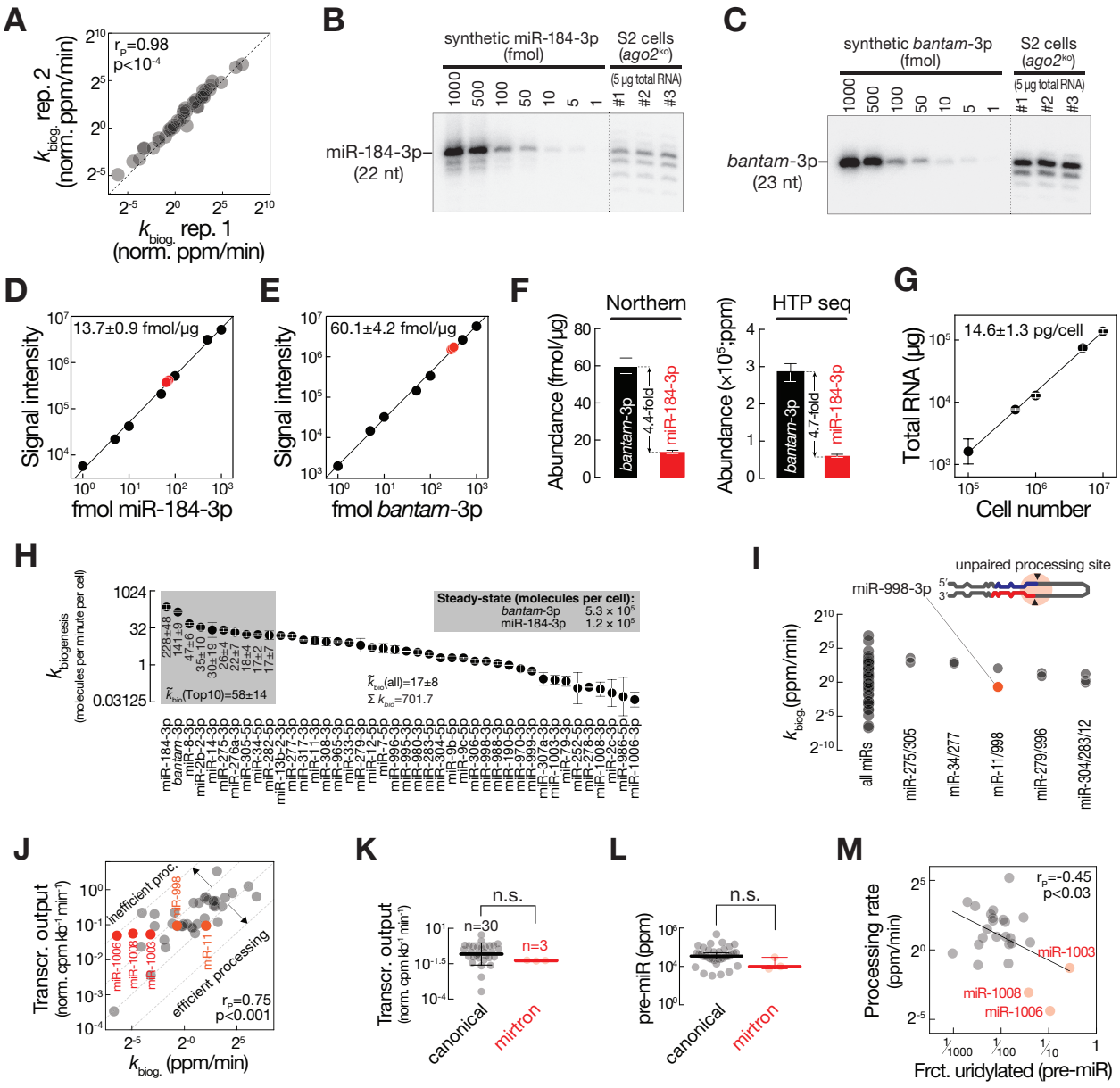

Figure S2. Reproducible measurements of absolute miRNA biogenesis rates by SLAMseq. Related to Figure 1. **(A)** Biogenesis rate ($k_{biog.}$, norm. ppm/min) was assessed for each of the abundantly expressed miRNA (>100 ppm, n=42) from the number of library-depth-, sequence-content-, and 4sU-labeling-efficiency-normalized T>C conversion-containing reads at early time-points of the measurement (i.e. 5 min, 15 min and 30 min) in two independent replicates of SLAMseq small RNA libraries prepared from *ago2*^{ko} S2 cells. Dashed line indicates *x=y*. Pearson correlation coefficient (r_P) and associated p-value is indicated. **(B)** and **(C)** Northern hybridization experiment to assess the number of miR-184-3p (B) or bantam-3p (C) molecules per *µ*g of total RNA. A standard series of titrated amounts (1000 – 1 fmol) of synthetic miR-184-3p (B) or *bantam*-3p (C) and three replicates of 5 *µ*g of total RNA extracted from *ago2*^{ko} S2 cells were loaded on the Northern blot and probed for the indicated miRNA. **(D)** and **(E)** Quantification of the signal intensity of synthetic miR-184-3p (D) or *bantam*-3p (E) dilution series (in black) was fit to linear regression to determine the amount of the respective miRNA in 5 *µ*g of total RNA (in red) as measured by Northern

hybridization shown in (B) or (C), respectively. **(F)** Abundance of *bantam*-3p and miR-184-3p (in fmol per μg total RNA) assessed by Northern blotting (left) or high-throughput sequencing (right, in ppm). Data represent mean ± stdev of three (Northern) or two (HTP seq) independent experiments. Relative abundance of *bantam*-3p and miR-184-3p (as indicated by fold-difference) is similar between the two detection methods. **(G)** Mean total RNA yield from a titration series of the indicated number of *Drosophila* S2 cells was determined to assess the RNA content per cell by linear regression. Data represent mean ± stdev of three independent experiments. **(H)** Absolute biogenesis rates ($k_{biogenesis}$, in molecules per minute per cell) for abundantly expressed (n=42) miRNAs in *ago2*^{ko} S2 cells. Data represent mean ± stdev of two independent experiments. Average biogenesis rate ± stdev of the ten miRNAs with the highest biogenesis rates (shaded grey area) (k_{bio}[Top10]) and all miRNAs analyzed (k_{bio}[all], n=42) is indicated. $\sum k_{bio}$ indicates the total number of miRNA molecules (derived from the 42 most abundant miRNAs in S2 cells) produced per minute per cell. The steady-state abundance (molecules per cell) of *bantam*-3p and miR-184-3p as assessed in (B-G) is indicated.
(I) Cistronic miRNAs tend to share similar biogenesis rates. Biogenesis rate ($k_{biog.}$ in ppm/min) for all abundantly expressed (> 100 ppm) miRs in *ago2*^{ko} S2 cells (n=42, all miRs) or grouped by cistronic cluster. Stem-loop structure of pri-miRNA encoding miR-998-3p (marked in red) as reported in miRbase is indicated (Kozomara and Griffiths-Jones, 2010). Extension of unpaired miRNA loop region into the destined miRNA duplex in miR-998 may explain slow biogenesis rates compared to miR-11, derived from the same primary transcript. **(J)** Correlation of transcriptional output (normalized counts per million (cpm) per kilobase and minute; determined by TT-SLAMseq) and miRNA biogenesis rate ($k_{biog.}$ in ppm/min) for abundantly expressed miRNAs in *ago2*^{ko} S2 cells (n=42). Mirtrons (i.e. miR-1006, miR-1003, and miR-1008) are highlighted in red and exhibit inefficient processing kinetics. Among the miR-11/998 cluster-encoded miRNAs (marked in orange), miR-998 exhibits slower biogenesis rates compared to miR-11. Pearson correlation coefficient (r_P) and associated p-value is indicated. **(K)** Comparison of primary-miRNA transcriptional output (normalized cpm per kilobase and minute; determined by TT-SLAMseq) of canonical miRNA-(n=30) and mirtron-(n=3) producing loci. P-value (Mann-Whitney test; n.s., p>0.05) is indicated. **(L)** Steady-state pre-miRNA abundance in parts per million (ppm) for canonical miRNAs (grey, n=39) or mirtrons (red, n=3), as previously reported (Reimão-Pinto et al., 2015). P-value (Mann-Whitney test; n.s., p>0.05) is indicated. **(M)** Correlation of pre-miRNA uridylation (as determined by (Reimão-Pinto et al., 2015) and miRNA processing rates (i.e. transcriptional output-normalized mature miRNA biogenesis rates; in ppm per minute) for 25 miRs that are abundantly expressed (> 100 ppm) in *ago2*^{ko} S2 cells and show detectable uridylation signature. Mirtrons are highlighted in red. Linear regression is shown. Pearson correlation coefficient (r_P) and associated p-value is indicated.

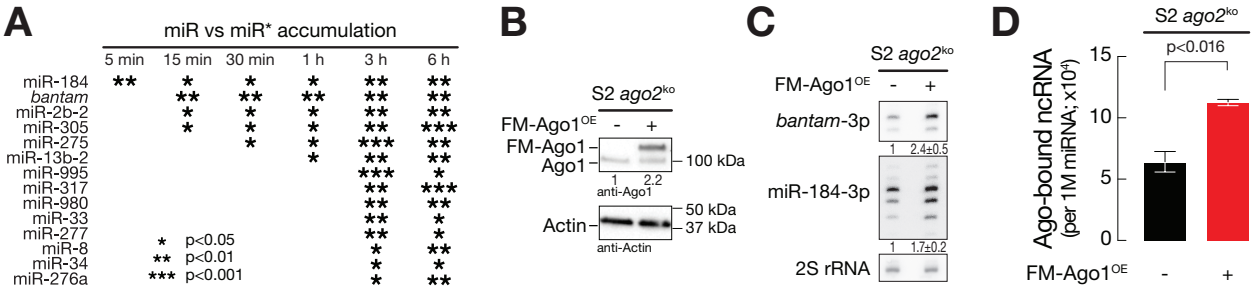

Figure S3. Intracellular kinetics of microRNA loading. Related to Figure 2. **(A)** Ago1-loading rates vary between individual miRNA duplexes. Ago1-loading was deduced by determining statistically significant, consecutive higher accumulation of 4sU-labeled miR strands compared to its miR* partner. P-value (Student's t-test) is indicated. **(B)** Western blot analysis probing for Ago1 protein levels in *ago2*^{ko} S2 cells or *ago2*^{ko} S2 cells expressing FLAG-MYC-tagged Ago1 (FM-Ago1^{OE}). Actin shows loading control. Relative quantification of total Ago1 levels is shown. **(C)** Representative Northern hybridization assay of *bantam*-3p and miR-184-3p in *ago2*^{ko} S2 cells or *ago2*^{ko} S2 cells expressing FLAG-MYC-tagged Ago1 (FM-Ago1^{OE}). Quantification of *bantam*-3p and miR-184-3p normalized to 2S rRNA (loading control) of three independent experiments (mean ± stdev) is reported below the respective blots. **(D)** Ago-bound noncoding RNA levels (ncRNAs, normalized to 1 million miRNAs) is shown for *ago2*^{ko} S2 cells (in black) and *ago2*^{ko} S2 cells expressing FLAG-MYC-tagged Ago1 (in red, *ago2*^{ko}; FM-Ago1^{OE}). Analyzed non-coding RNAs comprise of rRNA, tRNA, snRNA and snoRNA species. Data shows mean ± stdev of two independent replicates. P-value (two-tailed unpaired t-test) is indicated.

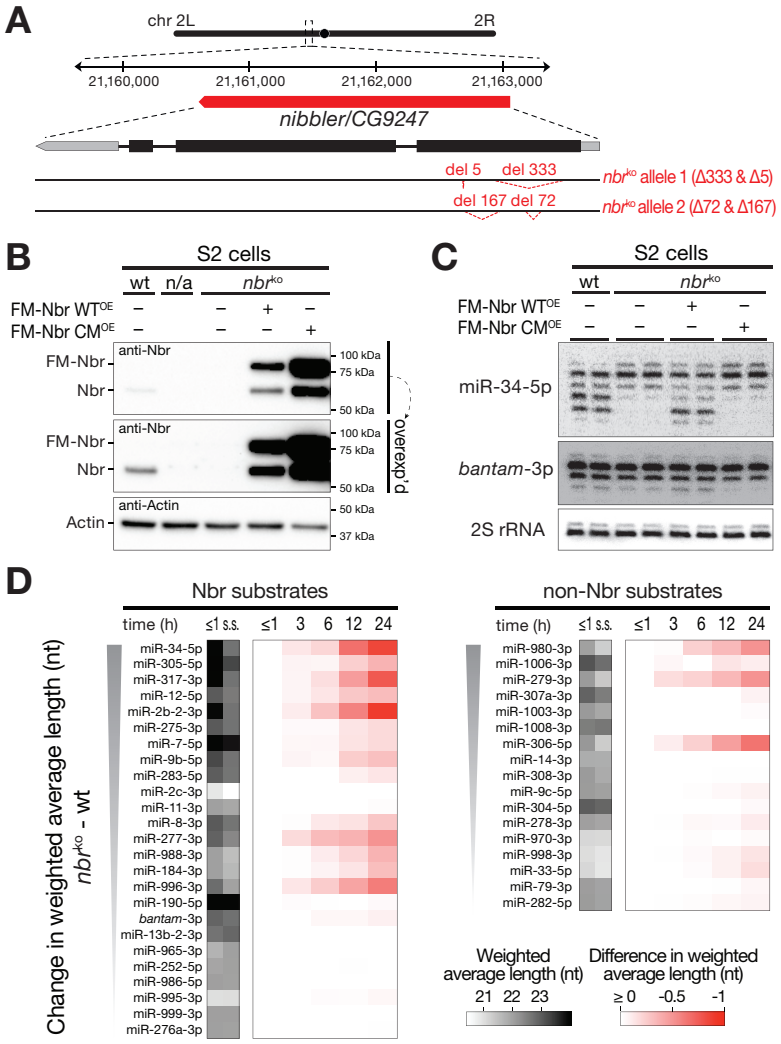

Figure S4. Intracellular kinetics of Nbr-mediated exonucleolytic miRNA trimming. Related to Figure 3 and STAR Methods. **(A)** Schematic representation of the *nibbler* (*nbr*) locus in the *Drosophila melanogaster* genome. Zoom-in shows the *nbr* gene (UTRs in grey, exons in black, intron as black line). Disruptive frameshift deletions (del) on two alleles of the *nbr*^{ko} clone are shown. **(B)** Western blot analysis of Nibbler (Nbr) in wild-type (wt) and *nbr*^{ko} S2 cells, as well as *nbr*^{ko} S2 cells expressing FLAG-MYC-tagged wild-type (FM-Nbr-WT^{OE}) or catalytic-mutant Nbr (FM-Nbr CM^{OE}). Actin serves as loading control. n/a, cell lysate not used in this study. **(C)** Northern hybridization assay for miR-34-5p, *bantam*-3p and 2S rRNA in wild-type (wt) and *nbr*^{ko} S2 cells, as well as *nbr*^{ko} S2 cells expressing FLAG-MYC-tagged wild-type (FM-Nbr-WT^{OE}) or catalytic-mutant Nbr (FM-Nbr CM^{OE}). Two independent biological replicates are shown. **(D)** Heat map represents the weighted average length (in grey scale) of 4sU-labeled miRNAs from early 4sU labeling time-points (≤1h) or of all reads across the entire time-course (steady-state, s.s.) and the difference in weighted average nucleotide length (in red scale) of 4sU-labeled miRNAs across a 4sU labeling time-course in *ago2*^{ko} S2 cells for Nbr substrates (left) or non-Nbr substrates (right). MicroRNAs are ranked according to the change in weighted average nucleotide length in cells, in which Nbr was functionally depleted (*nbr*^{ko} or *nbr*^{ko};FM-Nbr-CM^{OE}), compared to cells containing functional Nbr protein (wt or *nbr*^{ko};FM-Nbr-WT^{OE}). The T>C containing reads from early 4sU labeling time-points (≤1h) were merged prior to the analysis. Nbr substrates were defined by a significant increase in weighted average length (two-tailed Student's t-test, $p<0.01$) in small RNA libraries generated from S2 cells, in which Nbr was functionally depleted (*nbr*^{ko} or *nbr*^{ko};FM-Nbr-CM^{OE}), compared to cells containing functional Nbr protein (wt or *nbr*^{ko};FM-Nbr-WT^{OE}).

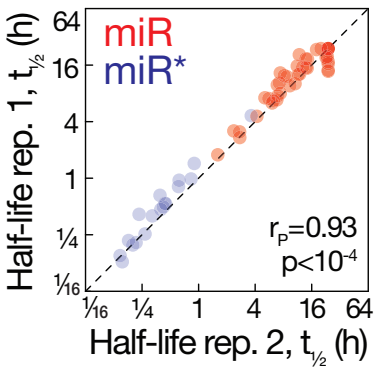

Figure S5. Robust measurements of miRNA stabilities using metabolic miRNA sequencing. Related to Figure 4. Comparison of half-life values determined in two independent biological replicates for 42 miR (red) and 18 miR* (blue) strands. Pearson's correlation coefficient (r_P) and associated p-value are shown.

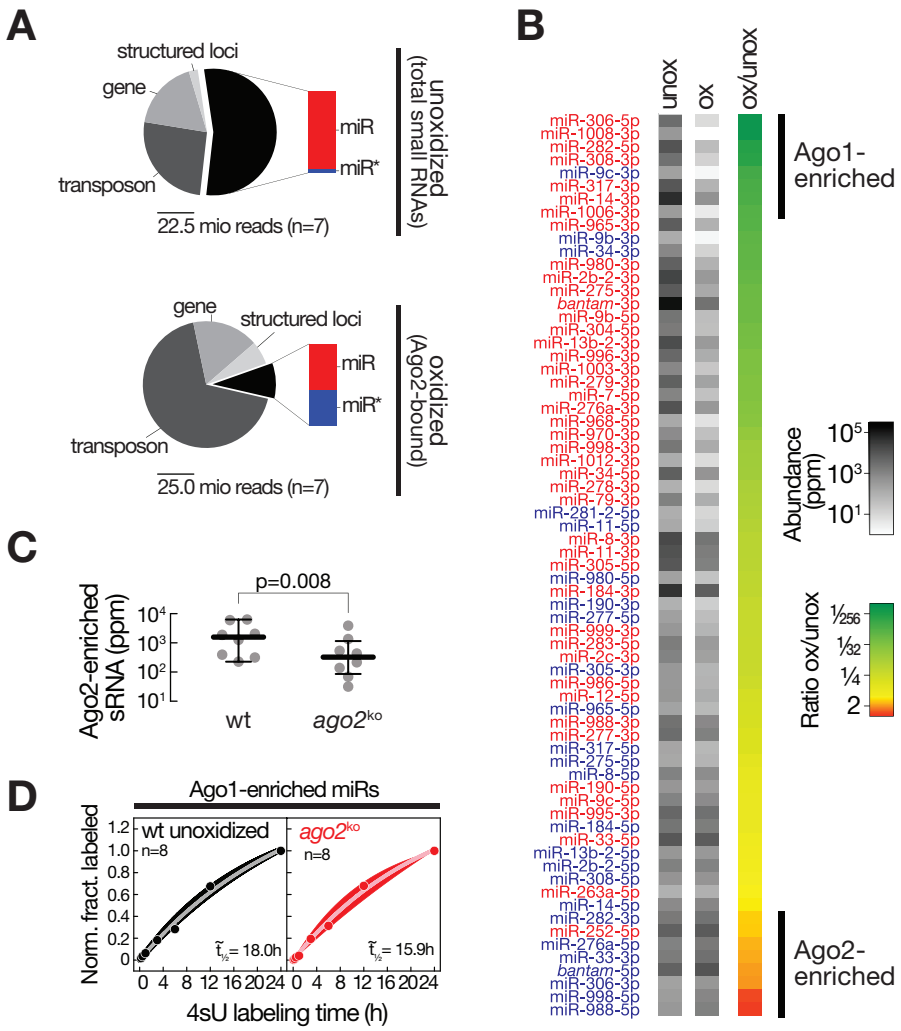

Figure S6. Argonaute protein identity determines small RNA stability. Related to Figure 5. **(A)** Pie charts represent the relative abundance of the indicated endo-siRNA classes and miRNAs in small RNA libraries from wild-type *Drosophila* S2 cells. Results from a standard cloning protocol (unoxidized, upper diagram) and from a cloning strategy that enriches for small RNAs with modified 3′ termini (oxidized, lower diagram) are shown. The fraction of miRs (red) and miR*s (blue) is indicated for both libraries. The average distribution of 7 datasets is shown. The average library depth is indicated. **(B)** Heat maps show the relative abundance (in ppm) of miRs (red), and miR*s (blue) in the indicated libraries (in grayscale). The ratio of relative representations in the libraries indicates preferential association of small RNAs with either Ago1 (green) or Ago2 (red). MicroRNAs classified as Ago1- or Ago2-enriched are indicated. **(C)** Abundance (in ppm) of Ago2-enriched miR and miR* species in wild-type (wt) and *ago2*^{ko} *Drosophila* S2 cells. Median and interquartile range is indicated. P-value (Wilcoxon matched-pairs signed rank test) is shown. **(D)** Median decay kinetics of Ago1-enriched small RNAs (n=8; classified in Figure S11B) in an 4sU metabolic labeling time-course in wild-type (wt, black) or *ago2*^{ko} S2 cells (red). Median and interquartile range of single-exponential saturation kinetics (as specified in main text) are shown. The half-life ($t_{1/2}$) as determined by curve-fitting is indicated.