EXTENDED EXPERIMENTAL PROCEDURES

Plasmid Construction and Transfections

Mouse TUT1-7 (except TUT4) and RRP6 full-length cDNAs (Open Biosystems) were cloned into the pcDNA3 vector with a N-terminal FLAG tag. Mouse TUT4 and Dis3 cDNAs were reverse transcribed from mouse total RNA, fused to a FLAG tag, and cloned into pCI and pcDNA3 vectors respectively. TUT4, TUT7, RRP6 and DIS3 were also made as fusion proteins with Myc epitope tag. In the case of DIS3 the epitope tags (FLAG or Myc) were fused to the C-terminus of DIS3. Point mutations were introduced with QuikChange XL II Mutagenesis kit (Stratagene) using primers listed in **Table S7**.

Plasmids were transfected with Fugene 6/HD (Promega) in 293T cells. *Dicer*$^{+/fl}$ MEFs were transfected with 18-20 nM siRNAs (except 10 nM for siTUT7) in Lipofectamine RNAiMax (Life Technologies) twice (on Day 0 and Day 2) before MEFs were collected 4 days post transfection.

Construction of HITS Libraries of Full-Length Pre-miRNAs

5'-end radiolabeled RNAs from anti-Ago IPs were resolved on a 10% denaturing PAGE gel. Gel slices containing 20-30 nt (miRNAs) and 50-80 nt (pre-miRNAs) RNAs were recovered. The eluted RNAs were ligated to the miRCat 3' Linker (IDT) using Truncated T4 RNA Ligase 2 (NEB) supplemented with 20% PEG6000 at room temperature for 3-4 hr. Ligation products were gel purified and reverse transcribed with RT primer (pGATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGGCATACGAATTGATG-GTGCCTACAG; p, phosphorylation; /idSp/, 1',2'-Dideoxyribose) using Affinityscript Reverse Transcriptase (Stratagene) at 55 °C for 60 min. The RT products were separated on a 10% denaturing PAGE gel and gel slices containing miRNAs and pre-miRNAs were recovered. The eluted cDNAs were circularized with CircLigase (EPICENTRE) at 60 °C for 60 min and then relinearized with APE 1 (NEB) at 37 °C for 90 min. 1 μl of the relinearized products were amplified with small RNA PCR Primer 1 (5'-CAAGCAGAAGACGGCATA) and Primer 2 (5'-AATGATACGGCGACCACCGACAGGTTC-AGAGTTCTACAGTCCGACG) using Phusion DNA polymerase (NEB) using the following PCR program: 95 °C for 30 sec, 20-25 cycles (95 °C for 10 sec, 60 °C for 10 sec, 72 °C for 5 sec). Then, 2 μl of 1st PCR products were re-amplified as above for 8-16 cycles. The resulting PCR products were separated on a 3% MetaPhor (Lonza) agarose gel. The gel slices containing 100-120 nt (miRNAs) and 130-175 nt (pre-miRNAs) DNAs were recovered and purified with MinElute Gel Extraction Kit (Qiagen). The miRNA, *WT* pre-miRNA and *D2A* pre-miRNA libraries were sequenced from their 5' ends for 101 nts using Genome Analyzer II (Illumina) as per manufacturer's instructions.

Preparation of Sequencing Libraries of Pre-miR-106b and Pre-miR-18a

20 μg of total RNAs from knock-down MEFs were resolved on 11% denaturing PAGE gels. 50-80 nt RNAs were gel purified and ligated to 50 pmole of Linker 1 (IDT) using RNA Ligase 2, Truncated (NEB) and 20% PEG 6000 at room temperature for 3-4 hrs. Ligated RNA products were gel purified, reverse transcribed with RT/PCR REV primer (IDT) using Superscript III (Life Technologies), and amplified with a miR-106b-5p specific forward primer with a BanI site and RT/PCR REV primer. PCR products were digested with BanI-HF (NEB), concatemerized with T4 DNA ligase and polished with Taq DNA polymerase. Concatemers were gel purified, dephosphorylated with Shrimp Alkaline Phosphatase, cloned using TOPO TA Cloning Kit (Life Technologies) and sequenced. Number of pre-miR-106b reads analyzed for SFigure 5D, E: siCtrl, 64; siDis3, 57; siRRP6, 48; siRRP40, 47. To clone pre-miRNAs from Ago-IPs from knock-down cells, the same protocol for HITS library construction was followed except that after the 1st PCR, 2 ul of 1st PCR products were re-amplified with a 5p miRNA specific forward primer with a BanI site (miR-106b TGTAGGCACCTAAAGTGCTGACAGTGCAGAT; miR-18a TGTAGGCACCTAAGGTGCATCTAGTGCAGATAG) and the RT/PCR REV primer (GATTGATGGTGCCTACAG) for another 10-20 cycles. PCR products were subsequently cloned and sequenced.

Immunopurification of Proteins

FLAG-tagged proteins were immunoaffinity purified from transfected 293T cells using anti-FLAG M2 beads (Sigma) in lysis buffer (40 mM HEPES-KOH, pH 7.4, 100 mM KOAc, 1 mM $Mg(OAc)_2$, 1 mM DTT supplemented with 0.1% NP-40 and Complete Protease Inhibitor, EDTA Free (Roche)) at 4 °C for 1 hr. FLAG beads were washed 5 times in lysis buffer before FLAG-tagged proteins were eluted in elution buffer (40 mM HEPES-KOH, pH 7.4, 100 mM KOAc, 1 mM $Mg(OAc)_2$, 1 mM DTT supplemented with 500 mM NaCl and 0.02 mg/ml of FLAG peptide). The eluates were dialyzed against 20 mM HEPES-KOH, pH 7, 100 mM KCl, 1 mM MgCl2, 1 mM DTT, and concentrated before glycerol was added to 50% and stored at -20 °C.

Co-Immunoprecipitations

Lysates from 293T cells co-transfected with FLAG-tagged and Myc-tagged proteins were immunoprecipitated with anti-FLAG M2 beads (Sigma) in lysis buffer RSB200 (20 mM Tris-HCl, pH 7.4,

200 mM NaCl, 2.5 mM $MgCl_2$) supplemented with 1 mM DTT, 0.1% NP-40 and Complete Protease Inhibitor, EDTA Free (Roche) at 4 °C for 1 hr. To detect the interaction b/w RRP6 or Dis3 with TUT4 or TUT7, RSB100 (20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM $MgCl_2$) supplemented with 1 mM DTT, 0.1% NP-40 and Complete Protease Inhibitor, EDTA Free (Roche) was used for the co-IP. FLAG beads were washed 6 times in lysis buffer before western blotting. To determine the dependence of interaction on RNA, after washed 4 times in lysis buffer, immunoprecipitates were split into two halves and one half was incubated with 100 μg/ml of RNase A in PBS at 4^0 C for 90 min before both halves were further washed 3 times in PBS.

To identify proteins that interact with endogenous Ago/pre-miRNAs, immunoprecipitations were performed with the 2A8 anti-Ago monoclonal antibody from Dicer knock-out MEFs reconstituted with catalytically inactive Dicer (D2A) as described previously (Liu et al., 2012). 2A8 immunoprecipitates were resolved by NuPAGE and co-precipitating proteins were excised from the gel and identified by mass spectrometry (Taplin Mass Spectrometry Facility, Harvard) and by Western blots of 2A8 immunoprecipitates with antibodies against endogenous TUT4 (Proteintech 18980-1-AP) and RRP40 (Bethyl A303-909A).

Uridylation Assay

5' radiolabeled pre-miRNAs was incubated with affinity purified FLAG-TUTs in 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 0.5 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 0.33 mM UTP at 37 °C for 2 hours.

Decay Assay

5' radiolabeled pre-miRNAs were incubated with affinity purified DIS3-FLAG in 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 μM $MgCl_2$, 10 μM $MnCl_2$, 1 mM DTT or FLAG-RRP6 in 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 1 mM DTT respectively at 37 °C for 2 hr. Assays in the presence of DIS3-FLAG and FLAG-RRP6 were performed in 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 0.5 mM $MgCl_2$, 0.1 mM $MnCl_2$, 1 mM DTT at 37 °C for 2 hr.

Dicer Processing Assay

5' radiolabeled pre-miRNAs were incubated with affinity purified V5-Dicer in Processing Buffer (20 mM HEPES, pH 7.4, 50 mM KOAc, 5 mM $Mg(OAc)_2$, 1 mM DTT) and 1 mM ATP at 37 °C at 1,000 rpm for 30 min.

Cloning of Drosha Processed Products

In-vitro Drosha processing assay was adapted from (Lee and Kim, 2007). In brief, internally labeled pri-miR-106b was incubated with affinity purified Drosha-FLAG beads in Processing Buffer supplemented with 1 mM $Mg(OAc)_2$ and 1 mM ATP at 37 °C at 1,000 rpm for 1 hr. Drosha processed products were resolved on a 11% denaturing PAGE gel and the gel slice corresponding to pre-miR-106b was sectioned. The eluted RNA was ligated to Linker 1 (IDT) using Truncated T4 RNA Ligase 2 (NEB) supplemented with 20% PEG6000 at room temperature for 3-4 hr. Ligation products were reverse transcribed with RT/PCR REV primer (IDT) using Superscript III (Life Technologies) and amplified with a miR-106b-5p specific forward primer with a BanI site and RT/PCR REV primer. PCR products were cloned using Zero Blunt TOPO Cloning Kit (Life Technologies) and sequenced.

Isolation of total RNAs, Northern Blot and Western Blot

Total RNAs were isolated with TRIzol (Life Technologies). Northern blots and western blots were performed as described previously (Liu et al., 2012). Northern Blots were exposed to phosphorimager and signals were quantitated with Storm scanner. All probes are listed in Table S7.

Real-Time RT-PCR

1 μg of total RNAs were treated with DNase I (Amplification Grade; Life Technologies) as per manufacture's instruction and reverse transcribed with random primers using Superscript III Reverse Transcriptase (Life Technologies) at 50 °C for 1 hr. cDNAs were quantitated by LightCycler 480 Real Time System (Roche) using LightCycler 480 SYBR Green I Master kit (Roche). HPRT1 or β-Actin was used as loading control. All primers are listed in Table S7. To quantitate miRNA levels by Taqman miRNA assays (Life Technologies), 100 ng of total RNAs were used as per manufacture's instruction. snoRNA234 or snoRNA202 was used as loading control.

Analysis of Ago-bound miRNA and pre-miRNA Libraries

In order to map reads that are fully mapped to the genome (fully-mapped reads) as well as reads with genome-matching 5' prefixes and non-genome-matching 3' tails (tailed reads), raw reads were processed and mapped to the mouse genome using a "balanced pipeline" (Zheng et al., 2010). First, raw reads were

aligned to the 3'-adapter using the cross-match program. Reads with ≥ 8 bp aligned at their 3' ends and with ≤8% mismatches were clipped at the insert-adapter boundary whereas reads without 3' adaptors were discarded. Second, to save the computational time and resources, identical clipped reads were collapsed into a single none-redundant tag (NR-tag) and the clone number (number of reads collapsed) of each NR-tag was recorded. We will use the terms "NR-tag" and "read" interchangeably hereafter, but in all situations the clone number of NR-tags is taken into consideration. Third, all NR-tags were mapped to the mouse genome (UCSC mm9 assembly) using the cross-match program. Reads with ≥15 bp aligned and with ≤6% mismatches were considered as fully-mapped reads. To identify reads that are mapped to the genome in the 5' portion with non-genome matching nucleotides at the 3' end (prefix reads), the cross-match alignment files were re-parsed to find all prefix reads. Prefix reads with only one mismatch in a tail longer than one nucleotide and those that can be fully mapped to another locus in the genome were discarded. The rest of the prefix reads were termed tailed reads. Mapped (full-mapped and tailed) reads were filtered and only those with 16-28 nts in the miRNA library and 38-101 nts in the pre-miRNA libraries were retained for further analysis.

Definition of the 5' and the 3' End of Pre-miRNAs

Position specific base composition (PSBC) graphs of individual pre-miRNAs showed that the stacking height of fully-mapped reads usually shows gradual decline with one or more significant drops close to the end of pre-miRNAs. We reasoned that this pattern is shaped by Drosha cleavage, 3'->5' trimming at the 3' end and premature reverse transcription termination at the 5' end. The deepest drop of the stacking height is determined primarily by Drosha cleavage. Therefore, we modeled these enzymatic activities by fitting changes of the stacking height (ΔC) near the 5' and 3' end of pre-miRNAs to Gamma distribution models, which is widely used in biological studies for approximation of multiple enzymatic activities that act independently in an exponential manner. ΔC values were calculated between every consecutive position as:

$$\Delta C_i = C_i - C_{i+3}$$

where ΔC_i is the change of the stacking height and C_i and C_{i+1} are the stacking heights at position i and i+1, respectively. Then ΔC values were fit to the Gamma distribution:

$$P(X = x|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-3} e^{-\beta x}$$

$$P(X \leq x|\alpha, \beta) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

, where x≥0, α,β>0 and α,β are the shape and rate parameters for Gamma-distribution, and $\Gamma(\alpha)$ and $\gamma(\alpha, \beta x)$ are the gamma function and lower incomplete gamma function, respectively. The parameters α and β were estimated by moments estimators as

$$\hat{\alpha} = \left(\frac{\bar{x}}{s}\right)^4$$

$$\hat{\beta} = \frac{s^4}{\bar{x}}$$

, where $\bar{x}$ and s are the sample mean and standard deviation of the ΔC values, respectively. Using the best fitted models, the confidence intervals (C.I.) of ΔC values for the 5' and the 3' end of each pre-miRNA were determined respectively. Then, significant drops were defined as sites with significantly large ΔC values ($p < 0.1$ for the 5' end and $p < 1/3$ for the 3' end, respectively). As the stacking height near the end of pre-miRNAs is controlled by a variety of enzymatic activities that only lead to near-perfect fitting of the Gamma distribution models, we used relative loose p-value cut-offs so that even moderately significant Drosha cleavage sites could be identified. Pre-miRNA boundaries were defined as the outer-most drops with significantly large ΔC values. Pre-miRNAs with fewer than 200 fully-mapped reads were excluded from this analysis.

Prediction of the Secondary Structure of Pre-miRNAs

The secondary structure of defined pre-miRNAs were predicted using the RNAfold program from Vienna package (Hofacker, 2003) and the 3' overhang length of pre-miRNAs was calculated by subtracting the 3' dangling end length by the 5' dangling end length.

Identification of the 3' Peak and the 3' Cliff

PSBC graphs of tailed reads of individual pre-miRNAs showed that tails usually stacked into two patterns: a 3' peak close to the 3' end and a 3' cliff that is > 10 nts away from the 3' end. To computationally

identify the 3' peak, we searched for local maximum stacking heights of tails, within 20 nts upstream of the 3' end of pre-miRNAs. The highest of these local maximum stacking heights that is closer than 7 nts away from the 3' end, was defined as the 3' peak. The highest of these local maximum stacking heights that is more than 6 nts away from the 3' end, with at least three stacking heights followed, was defined as the inner peak. The 3' cliff (Ago2 cleavage site) was defined as the site with the largest change of the stacking height (ΔC) at or upstream of the inner peak.

Small RNA Libraries from Knock-Down MEFs

20-30 nt small RNAs were gel purified from total RNAs and sequentially ligated to Linker-1 (IDT) and 5' barcoded adaptors as described in Table S7. The ligated products were amplified with DSFP5 (AATGATACGGCGACCACCGACTATGGATACTTAGTCAGGGAGGACGATGCGG) and Linker 1-RT (CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG), and sequenced by Hi-Seq 2000 (Illumina). Raw reads were trimmed from the 3'end to remove low-quality bases as described in the BWA program (Li and Durbin, 2009). The 3' adaptor was clipped using cutadapt with a 0.25 acceptable error rate. Reads that match transfected siRNAs, shorter than 15 nts or without 3' adaptor were excluded from further analyses. Filtered reads were aligned to the mouse (mm9) genome using the BWA program (Li and Durbin, 2009). A single mismatch was allowed for the first 18 nts and a maximum of 10 mismatches for the whole read. Different libraries were normalized by upper quartile normalization.

SUPPLEMENTAL REFERENCES

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res *31*, 3429-3431.

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

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Liu, X., Jin, D.Y., McManus, M.T., and Mourelatos, Z. (2012). Precursor microRNA-programmed silencing complex assembly pathways in mammals. Mol Cell *46*, 507-517.

Werner, C., Hanhoun, M., Widmann, T., Kazakov, A., Semenov, A., Poss, J., Bauersachs, J., Thum, T., Pfreundschuh, M., Muller, P.*, et al.* (2008). Effects of physical exercise on myocardial telomere-regulating proteins, survival pathways, and apoptosis. J Am Coll Cardiol *52*, 470-482.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.S., and Gregory, B.D. (2010). Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. PLoS Genet *6*, e1001141.