Mammalian 5′-Capped MicroRNA Precursors that Generate a Single MicroRNA

Mingyi Xie,[1] Mingfeng Li,[2] Anna Vilborg,[1] Nara Lee,[1] Mei-Di Shu,[1] Valeria Yartseva,[3] Nenad Šestan,[2] and Joan A. Steitz[1,*]
[1]Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Boyer Center for Molecular Medicine, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536, USA
[2]Department of Neurobiology, Kavli Institute of Neuroscience
[3]Department of Genetics
Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA
*Correspondence: joan.steitz@yale.edu
http://dx.doi.org/10.1016/j.cell.2013.11.027

SUMMARY

MicroRNAs (miRNAs) are short RNA gene regulators typically produced from primary transcripts that are cleaved by the nuclear microprocessor complex, with the resulting precursor miRNA hairpins exported by exportin 5 and processed by cytoplasmic Dicer to yield two (5p and 3p) miRNAs. Here, we document microprocessor-independent 7-methylguanosine (m^7G)-capped pre-miRNAs, whose 5′ ends coincide with transcription start sites and 3′ ends are most likely generated by transcription termination. By establishing a small RNA Cap-seq method that employs the cap-binding protein eIF4E, we identified a group of murine m^7G-capped pre-miRNAs genome wide. The m^7G-capped pre-miRNAs are exported via the PHAX-exportin 1 pathway. After Dicer cleavage, only the 3p-miRNA is efficiently loaded onto Argonaute to form a functional microRNP. This unusual miRNA biogenesis pathway, which differs in pre-miRNA synthesis, nuclear-cytoplasmic transport, and guide strand selection, enables the development of shRNA expression constructs that produce a single 3p-siRNA.

INTRODUCTION

MicroRNAs (miRNAs) are ubiquitous ∼22-nucleotide (nt) long gene regulators that modulate essential cellular processes at the posttranscriptional level. Accumulating evidence suggests that aberrantly expressed miRNAs lead to various diseases including cancer (Garzon et al., 2009), underscoring the importance of elucidating mechanisms by which miRNAs are expressed and how their biogenesis is regulated. Most metazoan miRNAs are produced from long primary (pri) miRNA transcripts by two cleavage events: first, the microprocessor (Drosha/DGCR8) complex processes the pri-miRNA into a 60–80 nt precursor (pre) miRNA hairpin in the nucleus. Subsequently, the nuclear-cytoplasmic export of the pre-miRNA is mediated by the RNA export factor exportin 5 (XPO5), which recognizes the double-stranded (ds) RNA stem along with the 2 nt 3′ overhang of the pre-miRNA. In the cytoplasm, Dicer cleaves off the terminal loop of pre-miRNA hairpins to produce ∼22 bp mature miRNA duplexes (Kim et al., 2009). The miRNA duplex is then loaded onto Argonaute (Ago) to form an RNA-induced silencing complex (RISC), with one strand usually selected based on the relative thermodynamic stability of duplex ends (Khvorova et al., 2003; Schwarz et al., 2003). Because strand selection is not a stringent process, miRNAs derived from both the 5′ and 3′ arms of a pre-miRNA hairpin, termed the 5p-miRNA and 3p-miRNA, may form functional miRNPs (Yang et al., 2011; Zhou et al., 2012).

Alternative miRNA biogenesis pathways that bypass either microprocessor or Dicer cleavage have been documented, increasing the complexity of the miRNA regulatory network (Yang and Lai, 2011). The first microprocessor-independent miRNA biogenesis pathway identified was the mirtron pathway, in which one or both ends of the pre-miRNA are generated by splicing (Flynt et al., 2010; Okamura et al., 2007; Ruby et al., 2007). In cells infected by several mammalian herpesviruses, microprocessor-independent viral miRNAs are cotranscribed downstream of other noncoding RNAs (ncRNA). The 5′ end of the viral pre-miRNA is therefore generated by 3′ end processing of the ncRNA: RNaseZ cleavage of tRNA-miRNA precursors generates *murine γ-herpesvirus 68* miRNAs (Bogerd et al., 2010), whereas Integrator cleavage of snRNA-miRNA precursors generates *Herpesvirus saimiri* miRNAs (Cazalla et al., 2011). Some miRNAs are derived from abundant ncRNAs such as tRNA and snoRNAs (Ender et al., 2008; Lee et al., 2009), thereby also bypassing the microprocessor. To date, there is only one documented Dicer-independent miRNA biogenesis pathway in which vertebrate miR-451 is generated by Ago2 slicer activity followed by an uncharacterized trimming reaction (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010).

MiR-320 is a PTEN (phosphatase and tensin homolog deleted on chromosome 10)-regulated anti-oncogenic miRNA that functions as a key intermediate in the PTEN-ETS2 (v-ets erythroblastosis virus E26 oncogene homolog 2) tumor suppressor pathway

CrossMark

in stromal fibroblasts (Bronisz et al., 2012). The production of miR-320 has been reported to require neither Drosha nor DGCR8 (Babiarz et al., 2008; Chong et al., 2010). Furthermore, the 5p-miRNA derived from pre-miR-320 hairpin was under-represented in miRNA-seq studies, suggesting that it is a directly transcribed endogenous short hairpin RNA (endo-shRNA), whose 5p-miRNA lacks the 5′-monophosphate required for inclusion in the sequencing libraries (Babiarz et al., 2008). This discovery led to the hypothesis that certain endo-shRNAs, including pre-miR-320, are synthesized by RNA polymerase (pol) III, similar to U6 promoter-driven shRNAs (Yang and Lai, 2011).

Here, we demonstrate that pre-miR-320 is an RNA Pol II transcript that is 7-methylguanosine (m^7G)-capped at its 5′ end, whereas its 3′ end is most likely produced by transcription termination. Using the cap-binding protein, eukaryotic translation initiation factor 4E (eIF4E), we developed a selection method called small RNA Cap-seq and identified a group of m^7G-capped pre-miRNAs in newborn mice. We show that m^7G-capped pre-miRNAs utilize the PHAX (phosphorylated adaptor for RNA export)-dependent exportin 1 (CRM1 or XPO1) pathway for nuclear-cytoplasmic transport, the first example of PHAX-XPO1-dependent miRNA biogenesis. As in the canonical pathway, Dicer processes m^7G-capped pre-miRNAs. However, guide strand selection by Ago is severely biased against the 5p-miRNA because of the presence of the m^7G-cap. This relatively stringent strand selection mechanism has allowed the development of a shRNA construct that expresses only one small interfering (si) RNA, from the 3p arm.

RESULTS

Human Pre-miR-320a Is Synthesized by RNA Pol II

We performed multiple sequence alignment analysis of mammalian miR-320 genes. Human miR-320a (referred to as miR-320a hereafter) was chosen for further analysis because it is predominantly expressed among the eight miR-320 gene family members (Morin et al., 2008), whereas nonprimate mammals possess only one miR-320 gene, which is homologous to human miR-320a. The alignment identified a putative RNA Pol II TATA box ~30 nucleotides upstream of the pre-miR-320 hairpin (Figure 1A).

To test whether mammalian pre-miR-320 is generated directly as an shRNA by RNA Pol II, we cloned and transiently expressed the miR-320a gene in human embryonic kidney (HEK) 293T cells, which otherwise produce low levels of miR-320a (Kim et al., 2008). Indeed, both mature and pre-miR-320a could be detected in 293T cells transfected with the wild-type, but not the TATA-box mutated gene (Figure 1B). We confirmed that miR-320a is transcribed by RNA Pol II using α-amanitin at a concentration that selectively inhibits RNA Pol II transcription (Figure S1A available online).

Human Pre-miR-320a Is 5′ m^7G Capped

A pre-miRNA synthesized by RNA Pol II should cotranscriptionally acquire a 5′ m^7G cap. We exploited the high-affinity m^7G binding of eIF4E by using glutathione S-transferase tagged 4E (GST-4E) bound to Sepharose beads to isolate capped RNAs

from total RNA preparations of 293T cells transfected with the miR-320a gene (Choi and Hagedorn, 2003) (Figure S1B). Pre-miR-320a was efficiently selected by GST-4E beads, whereas a mutant protein deficient in cap binding (GST-4E W102L) failed to interact (Marcotrigiano et al., 1997) (Figure 1C, compare lanes 4 and 7). Conversely, when the pre-miR-320a sequence was inserted into a canonical pri-miR-142 hairpin structure that is processed by the microprocessor to produce a 5′ monophosphate (Figure S1D), its interaction with GST-4E was abolished (Figure 1C, lane 10). Pre-miR-320a was also specifically selected by anti(α)-m^7G antibody (Figure S1E, lane 6). As negative controls, canonical pre-miR-16 (containing a 5′ monophosphate), U6 RNA (containing a 5′ γ-methyl phosphate cap), and U7 snRNA (containing a 5′ 2,2,7-trimethylguanosine cap) were not enriched by GST-4E (Figure 1C) or α-m^7G antibody (Figure S1E). As a positive control, 5′ m^7G-capped GAPDH mRNA was specifically enriched by GST-4E from total RNA extracted from mouse neuroblastoma NIE-115 cells (Figure S1C). 5′ rapid amplification of cDNA ends (RACE) determined that the first nucleotide of the predicted pre-miR-320a hairpin corresponds to the transcription start site (TSS): first, calf-intestinal alkaline phosphatase (CIP) was used to remove the phosphate group on the 5′ ends of undesired RNAs; then, tobacco acid pyrophosphatase (TAP) converted m^7G to monophosphate, enabling the ligation of a 5′ adaptor. Sequencing of two unique clones primed by P1 and P3 (Figure 1D) revealed that the 5′ end was the first G residue of the pre-miR-320a hairpin shown in Figure 1B.

Mouse m^7G-Capped Pre-miRNAs Are Identified by Small RNA Cap-Seq

Because our analyses indicated that human pre-miR-320a is m^7G-capped, we developed a small RNA Cap-seq protocol to identify m^7G-capped mouse (mmu-) pre-miRNAs genome wide (Figure 2A) (see Extended Experimental Procedures). First, 40–100 nt RNAs from a postnatal day 1 mouse were enriched by size fractionation on a polyacrylamide gel. Polynucleotide phosphorylase (PNPase), which efficiently degrades RNAs containing a single-stranded 3′ tail of >7 nt (Chen et al., 2013), was then used to degrade exonucleolytically the bulk of RNAs that contain such a tail, sparing pre-miRNAs because of their 2 nt 3′ overhangs (Figures S2A and S2B). Second, m^7G-capped pre-miRNAs were enriched by GST-4E selection (Figure S2C) and CIP-TAP cloning as described above (see Extended Experimental Procedures). The resulting library was subjected to deep sequencing (Figure S2D), yielding ~14 million reads of 25–45 nt that mapped to the mouse reference genome (NCBI37/mm9). More than 80% of the reads corresponded to the 5′ ends of known m^7G-capped transcripts (Figure 2B), validating the method. Furthermore, a strong yr motif (Y = pyrimidine, R = purine), in which R is the +1 position of the reads, was observed (Figure 2C), consistent with most Cap-seq reads representing genuine RNA Pol II TSSs, which typically start with an R preceded by a Y residue (Carninci et al., 2006).

As expected, Cap-seq reads mapped to the mmu-miR-320 locus, as well as to both TSSs, but not to the body of the divergent gene Polr3d (Figure 2D). Indeed, the first G of the pre-miR-320 hairpin corresponded to the preferred positive strand TSS in vivo (Figure 2E). In previous mouse miRNA-seq studies, the paucity of
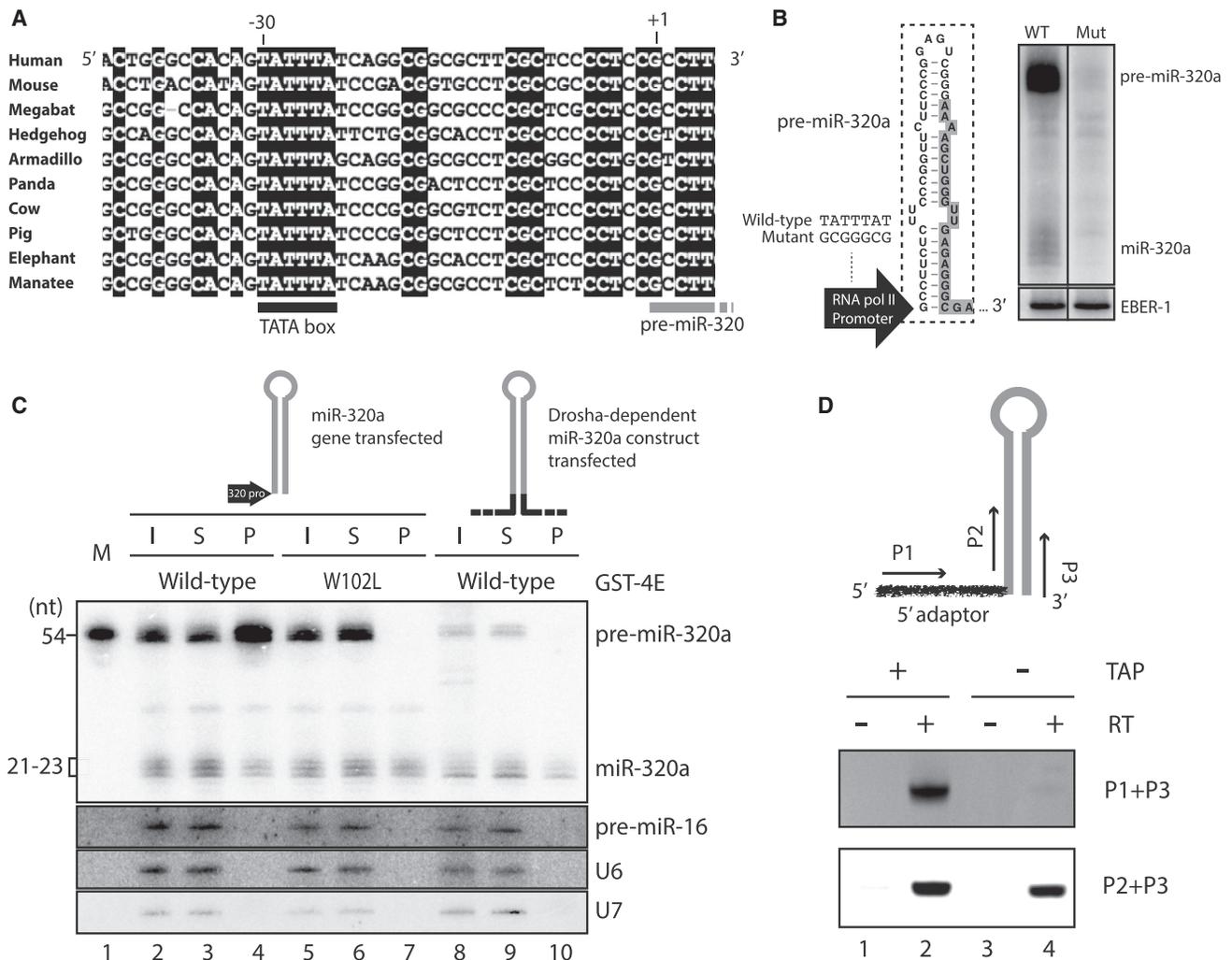
Figure 1. Human Pre-miR-320a Contains a 5′ m7G Cap

(A) Mammalian miR-320 gene promoter sequences are aligned, with conserved nucleotides highlighted in black. +1 denotes the first nucleotide (TSS) of pre-miR-320.

(B) Human pre-miR-320a (identical in sequence to mouse pre-miR-320) is shown, with miR-320a highlighted in gray. The right panel shows a northern blot probed for miR-320a and control EBER-1. Total RNA was isolated from 293T cells cotransfected with either pBS-miR-320a (WT) or pBS-miR-320a with a mutated TATA box (Mut) and a plasmid expressing EBER-1.

(C) Total RNA extracted from 293T cells transfected with either pBS-miR-320a or a plasmid encoding a pri-miR-320a engineered to undergo Drosha-dependent processing was subjected to GST-4E purification. Pre-miR-320a sequences are gray, with added sequences shown in black. Detailed sequence information is given in Figure S1D. GST-4E-selected RNAs were analyzed by northern blotting using probes for transfected miR-320a, endogenous miR-16, U6, and U7 RNAs. M, in vitro-transcribed 54 nt pre-miR-320a marker. A 22 nt RNA marker comigrated with the middle band of the mature miR-320a (not shown); I, input; S, supernatant; P, pellet; I and S were both 5% relative to P. Mature miR-320a seen in lanes 4, 7, and 10 resulted from nonspecific binding to the GST-4E beads.

(D) Small RNAs extracted from 293T cells transfected with pBS-miR-320a were first treated with CIP, then treated with or without TAP, followed by ligation to a 5′ adaptor. After reverse transcription (RT) using a miR-320a-specific primer (P3), PCR was carried out to amplify the 5′-adaptor-ligated pre-miR-320a (P1+P3) or just the pre-miR-320a hairpin (P2+P3).

See also Figure S1 and Table S1.

the 5p-miRNA from pre-mi-R320 was noted (Babiarz et al., 2008), consistent with the fact that m7G blocks 5′ adaptor ligation, a key step in miRNA-seq library preparation. Therefore, aligning Cap-seq with miRNA-seq reads provides another means of identifying m7G-capped pre-miRNAs (Figures 2E and S2E).

Among annotated mouse miRNA genes, we identified 19 candidates that generate m7G-capped pre-miRNAs (see Exper-

imental Procedures), 13 belonging to the murine-specific miR-344 family (Table 1). Although Cap-seq did not capture reads from pre-miR-344b or pre-miR-344d-1, sequence alignment suggests that all 13 mouse pre-miR-344 s are likely to be m7G-capped, as a conserved TATA box is appropriately positioned upstream (Figure S2F). Moreover, five Cap-seq selected pre-miRNAs—miR-320, miR-484, miR-344, miR-344e, and
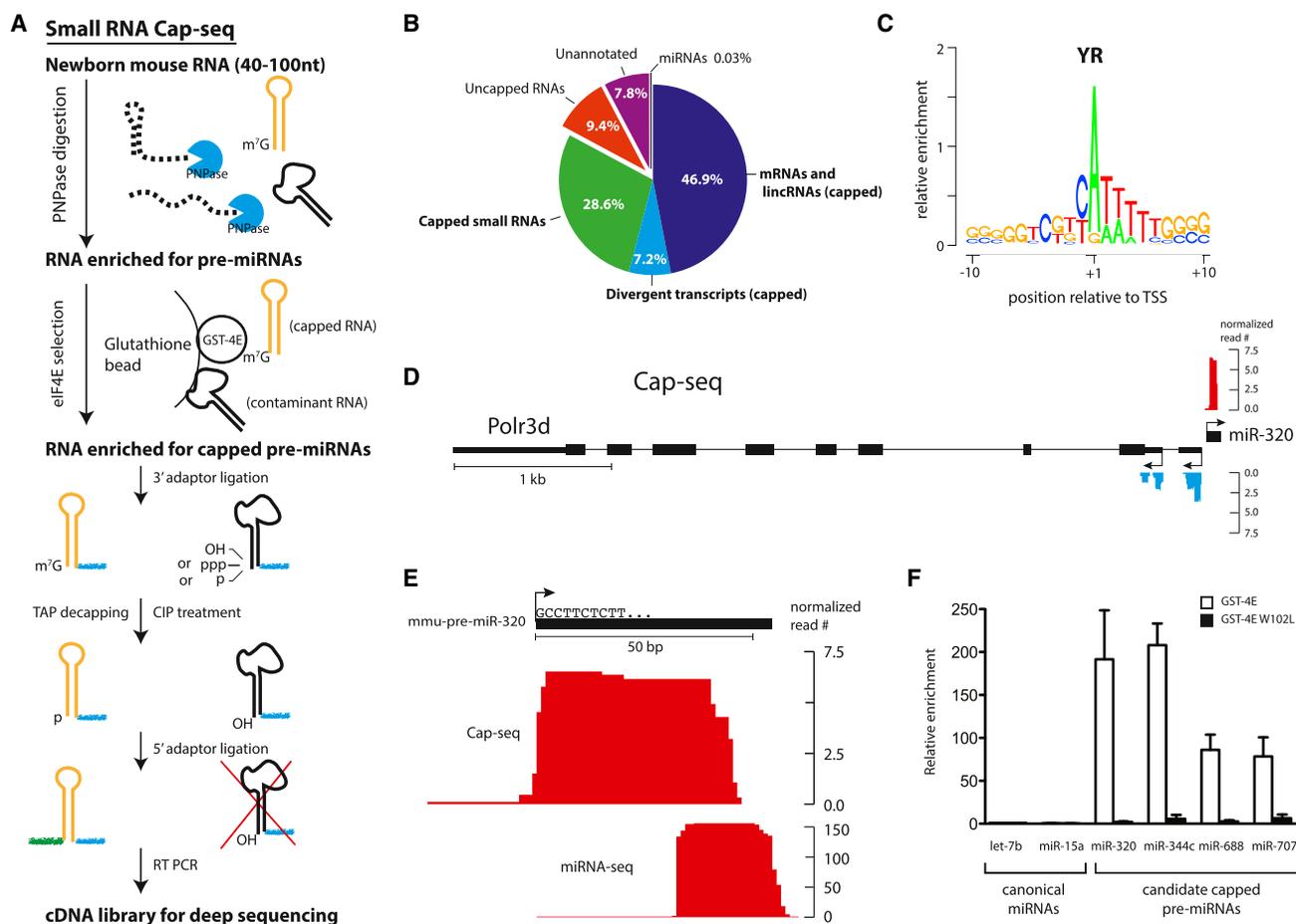
A Small RNA Cap-seq

Figure 2. Small RNA Cap-Seq Identifies Mouse m^{7}G-Capped Pre-miRNAs

(A) Flowchart illustrating the small RNA Cap-seq protocol (see Extended Experimental Procedures).

(B) Pie chart showing the distribution of small RNA Cap-seq reads. More detailed definitions of the different categories are given in the Extended Experimental Procedures.

(C) Motif analysis of nucleotides surrounding putative TSSs (+1 = first nucleotide of Cap-seq reads). The nucleotide height reflects the log2 ratio of the enriched nucleotides relative to genomic nucleotide composition.

(D) The histogram shows Cap-seq reads mapping to the divergent miR-320 and Polr3d loci, with the positive and negative strands colored red and blue, respectively. TSSs and transcription directionality are indicated by black arrows. The height of each bar is proportional to the normalized read number (raw read number per million mapped reads).

(E) Histogram showing Cap-seq reads and miRNA-seq reads (Chiang et al., 2010) that map to the miR-320 locus. The black bar shows pre-miR-320, with the sequence of the first 10 nt given. The 3′ ends of the clipped Cap-seq reads (45 nt) do not represent the 3′ ends of pre-miR-320.

(F) Size-selected RNAs (40 to 100 nt) from mouse NIE-115 cells were subjected to GST-4E selection, followed by RT-qPCR quantitation of six pre-miRNA hairpins. The bars show the relative enrichment of RNAs bound by GST-4E WT (white) and W102L (black), with pre-let-7b set to 1. Error bars represent SD from three experiments. All differences are statistically significant (Student's t test, p < 0.05).

See also Figure S2 and Table S1.

miR-344f—had previously been reported to be produced independently of DGCR8 and Drosha cleavage (Babiarz et al., 2008; Chiang et al., 2010; Chong et al., 2010). To validate the presence of m^{7}G-caps on newly-identified candidate pre-miRNAs, we selected small RNAs from mouse neuroblastoma NIE-115 cells by GST-4E binding followed by RT-qPCR to detect pre-miRNA hairpins. All four candidate m^{7}G-capped pre-miRNAs tested (miR-320, miR-344c, miR-688, and miR-707), but not canonical pre-miRNAs (let-7b and miR-15a), were enriched by GST-4E (Figure 2F). Thus, our small RNA Cap-seq protocol specifically identifies pre-miRNAs that contain a 5′ m^{7}G-cap.

M^{7}G-Capped Pre-miRNAs Are Exported by Exportin 1

We speculated that m^{7}G-capped pre-miRNAs might be exported from the cell nucleus differently from canonical pre-miRNAs because capped RNAs shorter than 200 nt are bound by CBC (cap-binding complex)-PHAX and exported by XPO1 rather than XPO5 (McCloskey et al., 2012). We investigated the requirement for export of m^{7}G-capped pre-miRNAs using siRNA knockdown experiments in 293T cells and transfection of murine-specific miR-344 gene, which is absent in humans. We verified that pre-miR-344 is indeed m^{7}G-capped when transiently expressed in these cells (Figure S3A). 293T cells depleted

Table 1. Murine m^{7}G-Capped Pre-miRNAs Identified by Small RNA Cap-Seq

Number	Genomic Location	miRNA Name	Cap-Seq Reads	Pre-miRNA Hairpin ΔG (kcal/mol)
1	chr16:14159635-14159784	mmu-miR-484	754	−35.1
2	chr8:130883116-130883265	mmu-miR-1903	340	−25.2
3	chr7:69191009-69191158	mmu-miR-344f	136	−39.3
4	chr14:70843284-70843433	mmu-miR-320	133	−36.8
5	chr7:69022606-69022755	mmu-miR-344	89	−43.9
6	chr7:69127118-69127267	mmu-miR-344g	69	−40.5
7	chr7:52105039-52105188	mmu-miR-707	65	−28.0
8	chr15:102502222-102502297	mmu-miR-688	59	−22.1
9	chr7:68982148-68982297	mmu-miR-344c	24	−32.9
10	chr18:6490510-6490659	mmu-miR-1893	19	−34.8
11	chr7:68871080-68871229	mmu-miR-344d-3	18	−39.7
12	chr7:69084860-69085009	mmu-miR-344-2	17	−43.9
13	chr7:68830103-68830252	mmu-miR-344d-2	6	−34.3
14	chr7:68880359-68880508	mmu-miR-344e	6	−38.3
15	chr7:69230070-69230219	mmu-miR-344i	4	−42.2
16	chr7:68884235-68884295	mmu-miR-344h-1	1	−31.3
17	chr7: 68887251-68887311	mmu-miR-344h-2	1	−31.3
18	chr7:68935405-68935467	mmu-miR-344b	0*	−32.5
19	chr7:68828010-68828078	mmu-miR-344d-1	0*	−28.0

Cap-seq reads show the number of raw reads that mapped to each pre-miRNA. *, miR-344b and miR-344d-1 were not captured by Cap-seq but were identified as candidate m^{7}G-capped pre-miRNAs through sequence homology to other miR-344 family members (Figure S2F). Free energy (ΔG) values were calculated using the RNA folding form on the mfold server. The predicted secondary structures of m^{7}G-capped pre-miRNAs are shown in Figures 1B, S2E, and S6D.

for PHAX or XPO5 by siRNA were transfected with three plasmids encoding mouse miR-344, miR-142 (a canonical miRNA whose levels are undetectable by northern blot in 293T cells) (Cazalla et al., 2011) and EBER-1 (Epstein-Barr virus-encoded RNA-1, a viral nuclear RNA) (Fok et al., 2006). The PHAX component of the XPO1 export pathway was chosen for knockdown because it is an RNA-specific export mediator, whereas XPO1 exports a variety of proteins as well (Ohno et al., 2000).

We find that knocking down PHAX in 293T cells reduced mature miR-344 levels 2-fold, without accumulating pre-miR-344 (Figure 3A); some pre-miRNAs have been reported to be subject to decay in the absence of exportins, whose binding counteracts nuclease attack (Yi et al., 2003). In contrast, XPO5 knockdown moderately increased miR-344 levels (Figure 3A), presumably because fewer canonical pre-miRNAs were exported to the cytoplasm to compete for Dicer processing. This is reminiscent of the increase in microprocessor-independent-miRNAs in DGCR8-knockdown cells, which are unable to generate normal levels of canonical pre-miRNAs (Babiarz et al., 2008; Cazalla et al., 2011). Concurrently, levels of the canonical miR-142-3p were reduced to <50% after XPO5 knockdown, but were not affected by PHAX knockdown (Figure 3A). Knockdown experiments with a second set of siRNAs targeting PHAX and XPO5 showed the same results (Figure S3B).

Furthermore, a stable interaction between m^{7}G-capped pre-miRNAs and PHAX was detected in vivo (Figure 3B). From extracts of NIE-115 cells crosslinked with formaldehyde, α-PHAX antibody coimmunoprecipitated precursors of U2 snRNA, miR-320, miR-344c, miR-688, and miR-707, but not the canonical pre-let-7b. In contrast, compared to pre-let-7b, none of the four m^{7}G-capped pre-miRNAs were enriched by α-XPO5 immunoprecipitation (Figure 3B), strongly suggesting that the PHAX-dependent XPO1 pathway participates in m^{7}G-capped pre-miRNAs biogenesis.

Additional evidence that XPO1 is responsible for m^{7}G-capped pre-miRNA export was obtained from *Xenopus* oocyte microinjection experiments. Four ^{32}P-labeled small RNAs were coinjected into the nuclei of *Xenopus* oocytes; EBER-1 is a nuclear RNA, whereas U1ΔSm (deletion of the Sm site prevents nuclear re-import), tRNA^{phe}, and pre-miR-15b are exported to the cytoplasm by XPO1, XPOt (exportin t), and XPO5, respectively (Rodriguez et al., 2004). Unlabeled pre-miR-344 (m^{7}G-capped) was coinjected in an amount previously determined to saturate exportins (Lund et al., 2004). Indeed, unlabeled pre-miR-344 inhibited U1ΔSm export (Figure 3C, compare lanes 3 and 5) to the same extent as unlabeled U1ΔSm (Figure S3C), confirming that m^{7}G-capped pre-miRNAs are exported by XPO1. Surprisingly, XPO5-mediated pre-miR-15b export was also inhibited by coinjecting pre-miR-344 (Figures 3C and S3D). This is unlikely caused by competition from pre-miR-344 molecules that become decapped during the experiment (Figure S3E). Together, these data suggest that when present in excess, m^{7}G-capped pre-miRNAs can be exported by either XPO1 or XPO5. However, the PHAX-XPO1-dependent export is the dominant pathway for m^{7}G-capped pre-miRNAs under physiological conditions (see Discussion).
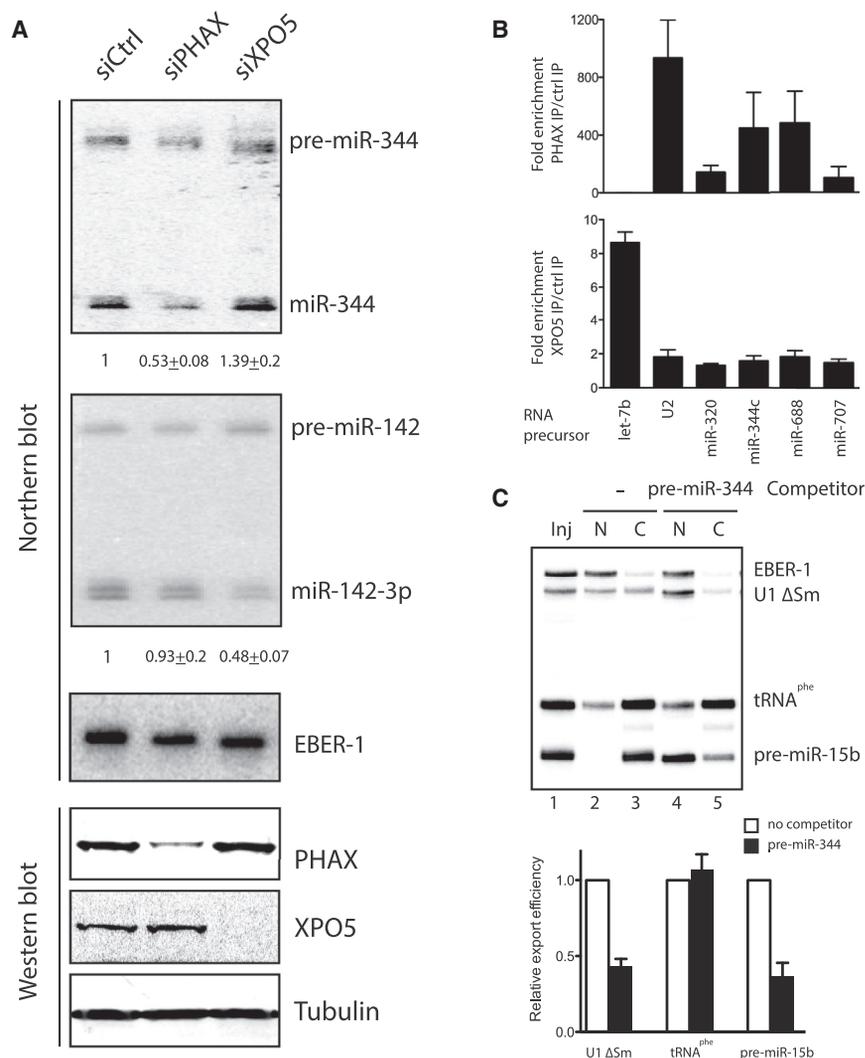
Figure 3. M^{7}G-Capped Pre-miR-344 Is Exported by Exportin 1

(A) Northern blots analyzing the levels of miR-344, miR-142-3p, and EBER-1 in transiently transfected 293T cells treated with a control siRNA (siCtrl) or a siRNA against PHAX (siPHAX) or exportin 5 (siXPO5) for 48 hr prior to cotransfection with plasmids expressing the three RNAs. Quantitations of relative mature miRNA levels (mean ± SD) were derived from three independent experiments. Western blots show the extent of knockdown of PHAX and XPO5, with Tubulin as a loading control.

(B) RNA extracted from NIE-115 cells crosslinked with 1% formaldehyde was subjected to immunoprecipitation using anti(α)-XPO5, α-PHAX antibodies or rabbit IgG and the precipitated RNAs were analyzed by quantitative RT-PCR. The y axis shows the relative fold-enrichment of the indicated precursor RNAs in the PHAX IP (upper panel) or XPO5 IP (lower panel) compared to IgG IP. The means and SD for three experiments are shown.

(C) *Xenopus* oocyte microinjection assay. A mixture of 1–10 fmol ^{32}P-labeled EBER-1, U1ΔSm, tRNA^{phe} and pre-miR-15b with or without 1 pmol of unlabeled pre-miR-344 (m^{7}G-capped) was injected into the nuclei of *X. laevis* oocytes. After 2 hr incubation, six oocytes were manually dissected, and RNAs from the nucleus (N) and cytoplasm (C) equivalent to one oocyte were extracted and analyzed on an 8 M urea-15% polyacrylamide gel. Inj, injected material. The bar graph shows relative RNA export efficiency (cytoplasmic/total) of the indicated RNAs with pre-miR-344 competition (black bars) compared to no competition (white bars). Error bars represent SD from three experiments.

See also Figure S3 and Table S1.

M^{7}G-Capped Pre-miRNAs Produce Single Mature miRNAs from the 3p Arm

The deficit of 5p-miRNA reads derived from m^{7}G-capped pre-miRNAs could arise in several ways (Figures 2E and S2E). One possibility is that 5′-capped miRNAs, even if present in a population, are lost in miRNA-seq studies because a 5′ monophosphate is required for the ligation step of the protocol (Babiarz et al., 2008; Yang and Lai, 2011). To test this explanation, we probed a northern blot of total RNA from 293T cells overexpressing miR-320a for the presence of the predicted miR-320a-5p. Almost none was detected (Figure 4A), nor was miR-344-5p observed in 293T cells transiently expressing miR-344 (Figure S4A).

This lack of capped 5p-miRNAs could be due to impaired Dicer cleavage of the 5p arm of a capped pre-miRNA (Gurtan et al., 2012; Park et al., 2011). Alternatively, the m^{7}G cap might tag the 5p-miRNA as the passenger strand to be degraded after displacement by Ago. To distinguish between these alternatives, we performed in vitro pre-miRNA processing assays and observed that recombinant human Dicer efficiently cleaves its pre-miR-320a substrate with or without an m^{7}G cap (Figure 4B). Note that the released capped 5p-miRNA migrates slightly slower on a polyacrylamide gel than its uncapped counterpart (Figure 4B, upper panel).

Finally, we asked if the m^{7}G cap on a 5p-miRNA influences binding to Ago by transfecting 293T cells with synthetic miR-320a duplexes with or without an m^{7}G cap on the 5p-miRNA. Immunoprecipitation of Ago-containing complexes revealed that m^{7}G-capped miR-320a-5p is much less efficiently loaded onto Ago than uncapped miR-320a-5p (Figure 4C, compare lanes 5 and 8). The bands below Ago-bound m^{7}G-capped miR-320a-5p (arrowhead) represent uncapped miRNAs that are produced upon transfection of synthetic miRNA duplexes (Figure S4B), but not when miRNAs are synthesized via the endogenous pathway (Figures 4A, 4D, and S4A). Heterogeneity of the Ago-bound miR-320a-3p may reflect widespread 3′ end modification (Burroughs et al., 2010) (Figure 4C, lower panel).

We confirmed that the m^{7}G cap of a pre-miRNA is responsible for the observed low levels of the 5p-miRNA by artificially capping the *Herpesvirus saimiri*-encoded pre-miR-HSUR4,
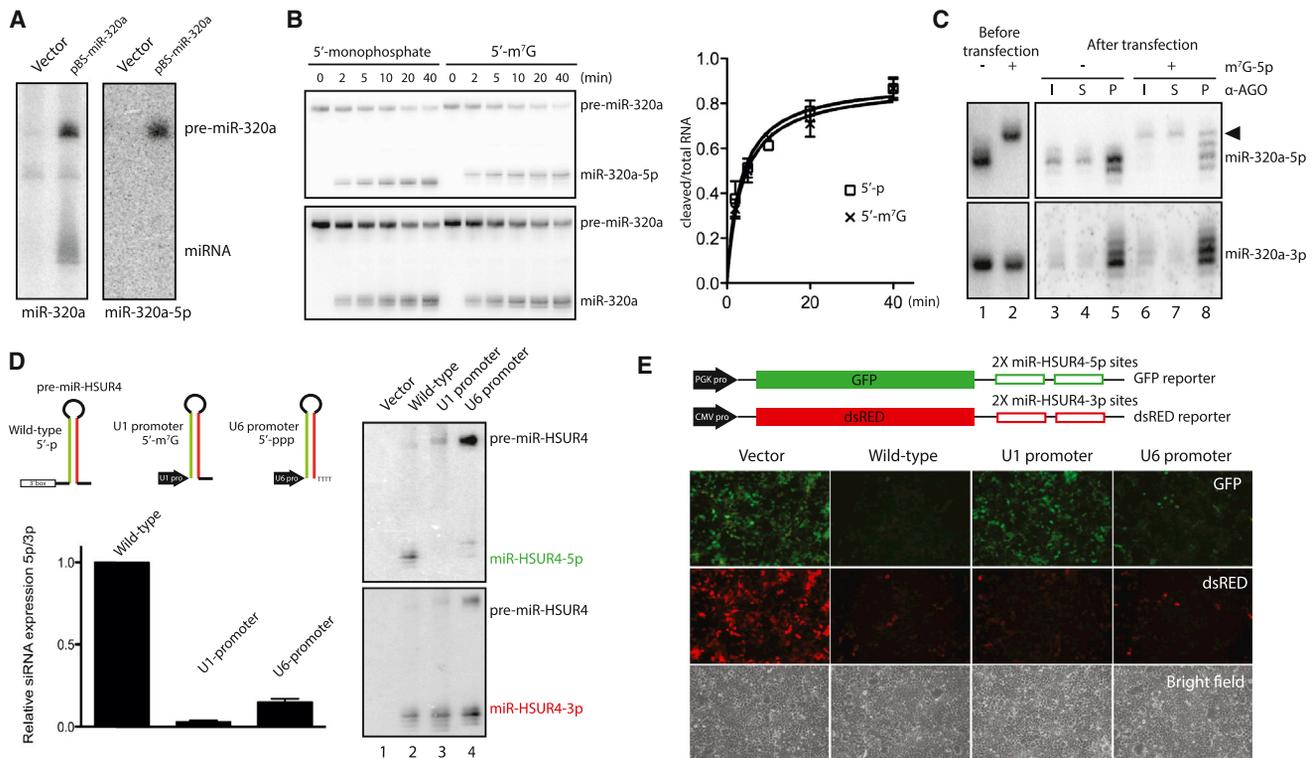
Figure 4. M^{7}G-Capped Pre-miRNAs Generate a Single Mature miRNA from the 3p Arm In Vivo

(A) Northern blots of total RNAs extracted from 293T cells transfected with pBS-miR-320a, probed for miR-320a or miR-320a-5p.

(B) In vitro-transcribed pre-miR-320a with a 5′ monophosphate or a 5′ m^{7}G cap was incubated with recombinant human Dicer for various times and analyzed by northern blot to detect miR-320a or miR-320a-5p. The plot shows cleavage of 5′-p pre-miR-320a (□) or 5′-m^{7}G-capped pre-miR-320a (X) by human Dicer at each time point. Error bars represent SD in three experiments.

(C) Northern blot analysis of miR-320a-5p and miR-320a-3p in the input sample (left) and coimmunoprecipitated with α-Ago antibody (right) from extracts of 293T cells transfected with synthetic miR-320a duplexes with or without an m^{7}G cap on miR-320a-5p. Upper panel: the arrowhead indicates m^{7}G-capped miR-320a-5p, whereas the bands below are uncapped miR-320a produced upon transfection of synthetic duplexes (Figure S4B). I, input; S, supernatant; P, pellet; I and S were both 10% relative to P.

(D) Northern blot analysis of miR-HSUR4-5p (top) and miR-HSUR4-3p (bottom) expressed in 293T cells transfected with either an empty vector or one of the three miR-HSUR4 expression constructs depicted on the left, with miR-HSUR4-5p and miR-HSUR4-3p colored in green and red, respectively. The corresponding 5′ functional groups are indicated. The bar graph shows relative miR-HSUR4-5p levels (normalized to miR-HSUR4-3p) from the three pre-miR-HSUR4 constructs, with the wild-type miR-HSUR4-5p level set to 1. Error bars represent SD in three experiments.

(E) 293T cells were cotransfected with the pGFP-miR-HSUR4-5p and pdsRED-miR-HSUR4-3p reporters (diagramed) and the miR-HSUR4 expression vectors shown in (D). GFP and dsRED fluorescence, as well as bright-field images of the cells after 48 hr, are shown.

See also Figure S4 and Table S1.

which normally generates comparable amounts of 5p-miRNA and 3p-miRNA (Figure 4D, wild-type) (Cazalla et al., 2011). We found that m^{7}G-capped pre-miR-HSUR4 transcribed by RNA Pol II from the U1 promoter did not yield detectable miR-HSUR4-5p (Figure 4D, lane 3). In contrast, when pre-miR-HSUR4 was synthesized instead by RNA Pol III from the U6 promoter, a common way to express shRNAs that contain 5′-tri-phosphates, detectable but low levels of 5p-miRNA were produced (Figure 4D, lane 4). Normalized to 3p-miRNA levels, the U6 promoter produced approximately eight times more 5p-miRNA than the U1 promoter (Figure 4D, bar graph).

M^{7}G-Capping of shRNAs Minimizes Off-Target Effects due to the 5p-siRNA

The above findings suggested that m^{7}G-capped shRNA expression vectors could be designed to produce a single siRNA from the 3p arm, thus minimizing off-target effects due to an undesired 5p-siRNA (Figure S4C). We therefore constructed GFP and dsRED reporters, whose 3′ UTRs contained two sites perfectly complementary to miR-HSUR4-5p and miR-HSUR4-3p, respectively (Figure 4E). These fluorescent protein reporters were cotransfected into 293T cells with an empty vector or one of the three pre-miR-HSUR4 expression vectors pictured in Figure 4D. All three pre-miR-HSUR4 constructs abrogated dsRED fluorescence (via production of miR-HSUR4-3p, see Figure 4D), whereas only the U1 promoter-driven m^{7}G-capped shRNA did not silence GFP (Figure 4E). Swapping the miR-HSUR4-5p and miR-HSUR4-3p target sites on the 3′ UTRs of the fluorescent protein reporters gave comparable results (Figure S4D). We conclude that positioning an RNA Pol II promoter such that a shRNA acquires an m^{7}G cap represents a simple strategy to generate a single (3p) siRNA for targeted silencing.
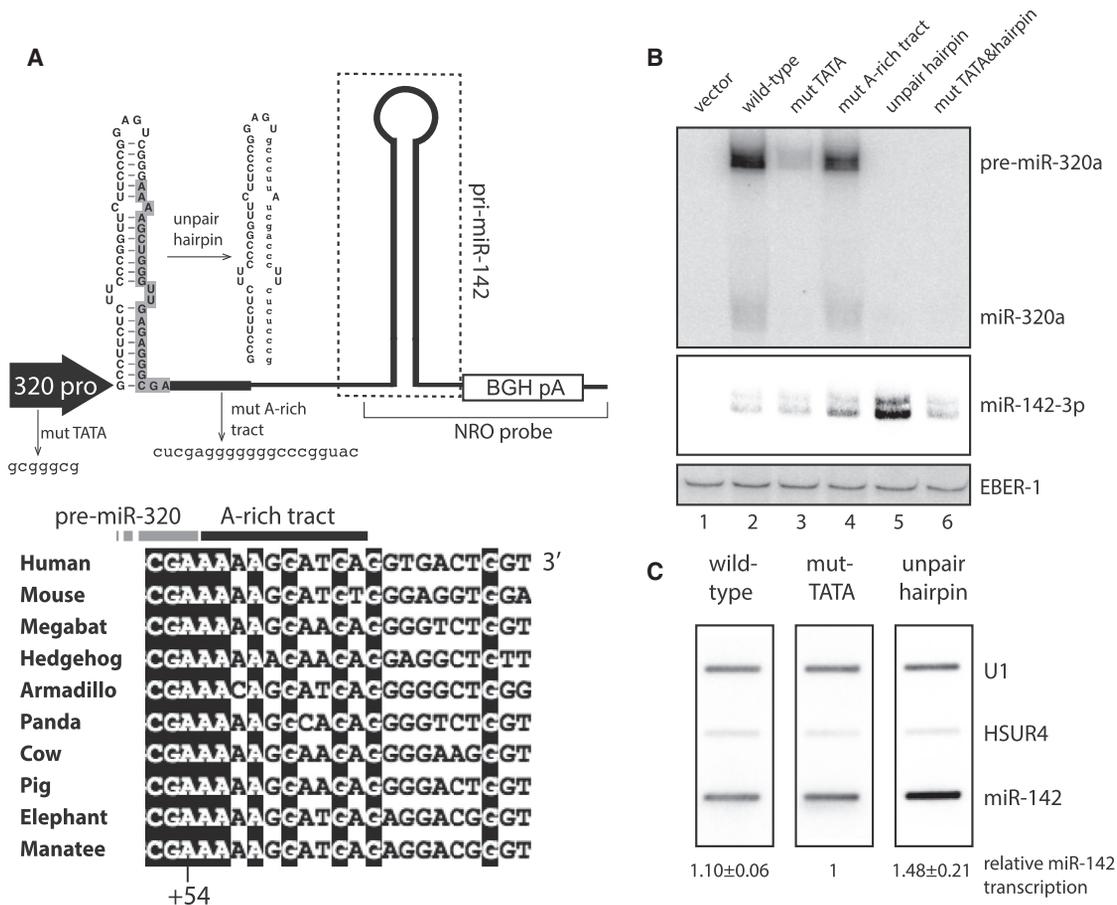
Figure 5. Pre-miR-320a 3′ End Formation Is Linked to Transcription Termination

(A) Schematic of the miR-320a/miR-142 transcription reporter, in which pri-miR-142 and the BGH-polyA signal were inserted downstream of pre-miR-320a. Regions subjected to mutational analysis were the TATA box, the pre-miR-320a hairpin and the A-rich tract; mutated sequences are shown in lowercase letters (detailed in Extended Experimental Procedures). NRO probe, region covered by the NRO probe. Sequences downstream of mammalian miR-320 genes are aligned, with conserved nucleotides highlighted in black. +54 denotes the last nucleotide of pre-miR-320.

(B) Northern blot analysis shows the expression of miR-320a, miR-142-3p and EBER-1 in 293T cells cotransfected with a wild-type or mutant transcription reporter and pEBER-1.

(C) Slot blots show NRO results using nuclei isolated from 293T cells transiently transfected with miR-320a/miR-142 transcription reporters. PCR NRO probes spanned the entire coding region of U1, HSUR4 (a *Herpesvirus saimiri*-encoded small nuclear RNA that is not present in 293T cells) or pri-miR-142 (including the downstream BGH-polyA sequence). Quantification shows relative miR-142 transcription levels (mean ± SE, n = 3), with the transcription ratio of miR-142 to U1 in the mut-TATA NRO experiment set to 1. The miR-142 signals for wild-type and mut-TATA constructs probably result from cryptic transcription (Figure 5B, lanes 2 and 3). See also Figures S5, S6, and Table S1.

M7G-Capped Pre-miRNA 3′ End Formation Is Linked to Transcription Termination

The size of 5′-capped pre-miRNAs is surprisingly uniform (∼54 nt for pre-miR-320a and ∼63 nt for pre-miR-344), suggesting a specific mechanism for 3′ end generation. Interestingly, miR-320 and miR-344 gene alignments revealed conserved downstream adenosine-rich regions (Figures 5A and S2F), resembling the A-rich 3′ end processing signals of several classes of Pol II transcripts, such as the cleavage and polyadenylation signal for mRNAs and the 3′ box for snRNAs (Weiner, 2005). However, siRNA-mediated knockdown of known 3′ end processing enzymes—subunit 11 (Int11) of the Integrator complex for snRNA processing (Baillat et al., 2005), cleavage and polyadenylation specificity factor 73 (CPSF73) for mRNA processing (Mandel

et al., 2006), or nuclear exosome catalytic subunit (Dis3) for 3′ to 5′ trimming (Flynt et al., 2010)—did not affect the patterns of precursor and mature miR-320a or mouse miR-344 (Figure S5A; data not shown).

We noted in published global run-on (GRO-seq) data that strong RNA Pol II promoter-proximal pausing/termination occurs at the 3′ ends of miR-320 and miR-344 (Figures S6A and S6B), whereas transcripts that require 3′ end processing show significant RNA Pol II density downstream of the processing site (Figures S6A and S6C) (Min et al., 2011). Therefore, we asked whether the 3′ ends of pre-miR-320a and other m7G-capped pre-miRNAs are generated by transcription termination, perhaps facilitated by a strong pre-miRNA hairpin abutting the TSS followed by an A-rich tract (Table 1; Figure S6D) (White et al.,
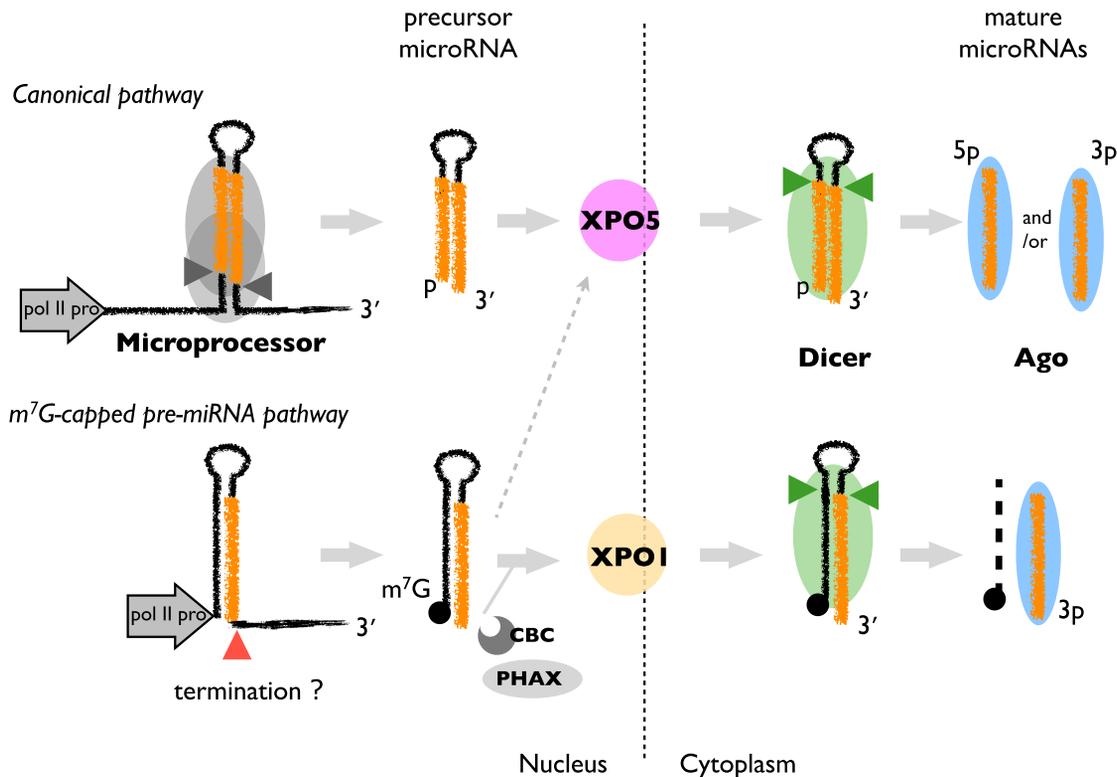
Figure 6. Model for miRNA Biogenesis via an m^{7}G-Capped Pre-miRNA

In this noncanonical miRNA pathway, the 5′ end of the pre-miRNA hairpin is generated directly by transcription initiation, with the 3′ end generated by transcription termination. Therefore, the 5′ end of the pre-miRNA intermediate contains an m^{7}G-cap. Although capped pre-miRNAs have the potential to be exported by XPO5 (dashed line), they are preferentially exported from the nucleus into the cytoplasm by the exportin 1 pathway. After Dicer processing, the 5p-miRNA is inefficiently loaded onto the Ago complex and undergoes degradation. Canonical miRNA biogenesis is illustrated in parallel. Mature miRNAs are highlighted in orange.

2013). To monitor transcription downstream of the pre-miR-320a hairpin, we inserted the canonical pri-miR-142 sequence, which—if synthesized—should produce miR-142 because canonical pri-miRNA processing is cotranscriptional (Figure 5A) (Morlando et al., 2008; Pawlicki and Steitz, 2008). Whereas background levels of miR-142-3p were observed from both the wild-type and mutant miR-320a promoter (mut TATA), better miR-142-3p expression was obtained after mutating the A-rich tract or unpairing the miR-320a hairpin structure by mutating either the 3p or 5p arm (Figure 5B, lanes 4 and 5; Figure S5C, lanes 1, 5, and 6). Mutating the A-rich tract slightly reduced miR-320a production from the same vector, consistent with the hypothesis that impaired transcription termination should produce less pre-miR-320a (Figure 5B, lane 4; Figure S5C, lane 5). Likewise, a stronger signal downstream of pre-miR-320a was observed in a nuclear run-on (NRO) assay (Figure 5C), which detects nascent RNA transcripts, for the unpaired hairpin relative to the wild-type construct. However, on northern blots, we did not detect significant read-through product (pri-miR-142) resulting from downstream termination sites (Figures S5B and S5C). Although rapid processing at the A-rich tract cannot be completely ruled out as a 3′ end formation mechanism, these data add support to the idea that RNA Pol II transcription termination underlies production of the 3′ end of pre-miR-320a.

DISCUSSION

An Unusual miRNA Biogenesis Pathway that Involves m^{7}G-Capped Pre-miRNAs

We have uncovered a biogenesis pathway involving 5′ m^{7}G-capped pre-miRNAs that produces ∼20 miRNAs in human or mouse (including 13 mmu-miR-344 family members) (Figure 6 and Table 1). Two of these miRNAs (miR-320 and miR-484) have been reported to function in the PTEN-ETS2 tumor suppressor pathway and to regulate mitochondrial networking, respectively (Bronisz et al., 2012; Wang et al., 2012). Because their 5′ and 3′ ends are apparently generated directly by RNA Pol II (Figures 1 and 5), m^{7}G-capped pre-miRNAs do not undergo microprocessor cleavage and proceed directly to the cytoplasm for Dicer processing. The presence of the m^{7}G-cap, which is commonly found at the 5′ ends of nascent RNA Pol II transcripts and is known to increase RNA stability and regulate gene expression (Topisirovic et al., 2011), makes this unusual pathway distinct from canonical miRNA biogenesis also in pre-miRNA nuclear-cytoplasmic transport and guide miRNA selection (Figures 3 and 4).

Unlike XPO5-dependent canonical pre-miRNAs, m^{7}G-capped pre-miRNAs are exported by the PHAX-XPO1 pathway (Figure 3). It is worth noting that very few, if any, m^{7}G-capped pre-miRNAs

undergo uncapping to produce canonical pre-miRNAs, as almost no mature 5p-miRNAs containing 5′ monophosphates were detected in miRNA-seq studies (Babiarz et al., 2008). The involvement of XPO1 in exporting pri-miRNAs in *Caenorhabditis elegans* and *Drosophila* was previously reported, although it remains unclear how these pri-miRNAs are subsequently processed into pre-miRNAs in the cytoplasm (Büssing et al., 2010). In mammals, XPO1 can shuttle mature miRNAs between the nucleus and the cytoplasm (Castanotto et al., 2009). To our knowledge, m^7G-capped pre-miRNAs are the only pre-miRNAs that are exported by XPO1. In 293T cells, our knockdown experiments identified XPO1 as responsible for exporting m^7G-capped pre-miR-344. This is consistent with the notion that CBC and PHAX cotranscriptionally bind the cap of nascent transcripts (McCloskey et al., 2012) and direct m^7G-capped pre-miRNAs into the XPO1 pathway prior to their potential interaction with XPO5. In contrast, in *Xenopus* oocytes, excess unlabeled m^7G-capped pre-miRNAs are capable of saturating XPO1 as well as XPO5; structural studies indicate that the 5′ functional group of a pre-miRNA does not interact with XPO5 (Okada et al., 2009). Therefore, if an m^7G-capped pre-miRNA is not efficiently bound by CBC and PHAX or present in excess, perhaps it can alternatively be exported by XPO5, although PHAX-XPO1-dependent export is the physiologically relevant pathway.

The m^7G cap on a pre-miRNA does not affect the efficiency of Dicer cleavage (Figure 4B). This result was unexpected because human Dicer has been reported to require a 5′ monophosphate for accurate and efficient processing of microprocessor-generated pre-miRNAs following the 5′ counting rule: Dicer cleaves ∼22 nt from the 5′ monophosphate end (Park et al., 2011). Furthermore, there exists a human RNA-methyltransferase BCDIN3D, which phospho-dimethylates the 5′ monophosphate of certain pre-miRNAs and has been observed to reduce Dicer processing efficiency in vitro (Xhemalce et al., 2012). Yet, pre-miRNAs with relatively stable base-pairing at the base of the hairpin have been observed to be processed preferentially by Dicer following the 3′ counting rule: Dicer cleaves ∼22 nt from the 2 nt overhang at the 3′ end of the stem (Park et al., 2011). Because this key 3′ determinant is present on the m^7G-capped pre-miRNAs identified in this study, and the base of the m^7G-capped pre-miRNA hairpins appears to be thermodynamically stable (Figures 1B and S2E), we speculate that m^7G-capped pre-miRNAs mainly obey the 3′ counting rule for efficient cleavage by Dicer.

A 5′ m^7G cap is expected to inhibit the association of a mature miRNA with Ago (Figure 4C) as the Ago crystal structure reveals a binding pocket that forms extensive interactions with the 5′ monophosphate of the miRNA (Elkayam et al., 2012; Schirle and MacRae, 2012). Consistent with the finding that m^7G-capped 5p-miRNAs are not favored as functional guide miRNAs during RISC assembly (Figures 4E and S4D), two capped miRNA-sized small RNAs (18–25 nt) encoded by human hepatitis delta virus do not direct Ago-mediated RNA silencing (Haussecker et al., 2008).

It has been suggested that the mirtron miRNA biogenesis pathway may have emerged before the advent of a specialized microprocessor (Ruby et al., 2007). Accordingly, XPO1-dependent m^7G-capped pre-miRNAs may represent a group of ancient miRNAs that appeared before the emergence of XPO5. In contrast to mammals, *C. elegans* does not encode an XPO5 homolog but does express XPO1 (Büssing et al., 2010), and most miRNAs in *C. elegans* are expressed from proximal RNA Pol II promoters (Martinez et al., 2008). Thus, many more m^7G-capped pre-miRNAs may be identified in *C. elegans* and possibly other species. The ability of m^7G-capped pre-miRNAs to be exported by XPO1 may enable their expression in cell types, tissues, or developmental stages that lack XPO5. Therefore, one reason that certain m^7G-capped pre-miRNAs have been maintained during evolution could be that they are required under circumstances where XPO5 is not expressed. Alternatively, maintaining m^7G-capped pre-miRNAs is a clear strategy to diminish the expression of a 5p-miRNA that may be detrimental to the organism. It remains for future studies to determine whether all miRNAs derived from m^7G-capped pre-miRNAs are under similar transcriptional control and/or regulate genes involved in a common pathway.

Promoter-Proximal Transcription Termination Is Linked to m^7G-Capped Pre-miRNA 3′ End Formation

In metazoans, RNA Pol II promoter-proximal pausing/termination occurs ∼30–50 nt after the TSS (Core et al., 2008). Our data suggest that promoter-proximal pausing leads to transcription termination to generate the 3′ end of m^7G-capped pre-miRNAs (∼50 nt long) (Figure 5). In support of a termination model, miR-320a exhibits significant 3′ end heterogeneity (Figures 1 and 2E), whereas Drosha-cleaved miR-320a is more defined in length (Figure 1C, compare mature miR-320a in lanes 2–7 to lanes 8–10). Although we were unable to determine with nucleotide resolution the site of transcription termination or to rule out conclusively the possibility that unknown factors act in rapid 3′ end processing, we favor the model that RNA Pol II transcription termination generates the 3′ ends of m^7G-capped pre-miRNAs.

m^7G-Capped shRNAs in RNAi Technology

Conventionally, constitutive siRNA expression is achieved by expressing RNA Pol III promoter-driven shRNAs that are directly processed by Dicer. Alternative use of an RNA Pol II promoter enables regulated and cell-type or tissue-specific expression of shRNAs, but requires engineering the shRNA into a canonical pri-miRNA context for microprocessor cleavage (Pan et al., 2012). We have demonstrated that specific shRNAs can be generated directly by RNA Pol II because only the 3p-siRNA is produced from 5′ m^7G-capped shRNAs. A similar outcome was reported previously, without revealing that the m^7G cap dictates strand selection (Denti et al., 2004). This finding has important implications for RNAi technology (Figures 4D and 4E), in which many efforts have been made to reduce off-target effects caused by the undesired passenger strand (Gu et al., 2012a). In combination with current strategies that require careful assessment of the thermodynamic stability of siRNA duplex ends and consideration of the "loop counting rule" for Dicer processing (Gu et al., 2012a; Khvorova et al., 2003; Schwarz et al., 2003), expressing m^7G-capped shRNAs from a proximal RNA Pol II promoter constitutes a simple and effective strategy to enhance specificity by producing a single siRNA. Furthermore, the unique ability of m^7G-capped shRNAs to utilize the

PHAX-XPO1 pathway for nuclear export enables their expression in XPO5-deficient cells and in altered cell states (Iwasaki et al., 2013).

EXPERIMENTAL PROCEDURES

Cell Culture, Plasmid Transfection, and Transient RNAi Knockdown

HEK293T cells were maintained in DMEM containing heat-inactivated FBS, penicillin and streptomycin, and L-glutamine. Plasmid transfections were carried out using Lipofectamine 2000 (Invitrogen) or TransIT-293 (Mirus Bio) according to the manufacturers' instructions. After 48 hr, total RNA was harvested using Trizol reagent (Invitrogen) and analyzed by northern blot of transfected or endogenous RNAs indicated in the figures. SiRNAs against specific proteins were transfected with Lipofectamine RNAiMAX reagent (Invitrogen). Forty-eight hours later, plasmids encoding desired RNAs were transfected using Lipofectamine 2000 or TransIT-293 (Cazalla et al., 2011). SiRNAs (Integrated DNA Technology) used in this study are listed in Table S1.

Small RNA Cap-Seq

GST-4E was recombinantly expressed in *Escherichia coli* BL21 (Choi and Hagedorn, 2003) and crude lysate prepared by sonication was loaded onto glutathione Sepharose beads (GE Healthcare). Total RNA was extracted from a P1 mouse (C57BL/6 strain) using Trizol reagent and size-selected after extraction on an 8 M urea-8% polyacrylamide gel to obtain 40–100 nt small RNAs. Purified small RNAs were then incubated with recombinant *Deinococcus radiodurans* polynucleotide phosphorylase (PNPase) (Chen et al., 2013) and incubated with GST-4E- or GST-4E-W102L-coated glutathione beads in NET-2 buffer. Purified RNAs were ligated to a ^{32}P-labeled 3′ adaptor, CIP-TAP-treated, and ligated to the 5′ adaptor. The cDNA library was constructed as described (Cazalla et al., 2011) and sequenced at the Yale Stem Cell Center Genomics Core on an Illumina HiSeq2000 instrument using 50 bp runs. Bioinformatic analysis is described in the Extended Experimental Procedures.

In Vitro Transcription of RNAs

RNAs were synthesized in vitro by T7 RNA polymerase in a 10 μl reaction containing 40 mM Tris-HCl (pH 8.0), 25 mM NaCl, 8 mM MgCl$_2$, 2 mM Spermidine-(HCl)$_3$, 10 mM DTT, 1 mM ATP, CTP, and UTP, 100 μM GTP and 1 mM GMP or m^{7}G cap analog (Promega), 20 U RNase Inhibitor (Roche), and 5 U T7 RNA polymerase. To synthesize ^{32}P-labeled RNAs, 1 mM UTP was substituted with 100 μM UTP and 50 μM α-^{32}P UTP (800 Ci/mmol, Perkin-Elmer). Reactions were carried out at 37°C for 8 hr (or 1 hr for ^{32}P-labeling).

Microinjection into *Xenopus* Oocytes

Stage V and VI oocytes were isolated from female *Xenopus laevis* and the nuclei were injected with 18.4 nl solution containing 20 mg/ml blue dextran, 1–10 fmol ^{32}P-labeled RNAs with or without 1 pmol competitor RNA. After 2 hr, nuclei were manually separated from the cytoplasm in 5:1 + PO$_4$ buffer. Six blue nuclei and the corresponding cytoplasms were collected separately. One-oocyte equivalent of RNA was fractionated on an 8 M urea-15% polyacrylamide gel and analyzed by PhosphoImager.

In Vitro Processing of Pre-miRNAs by Dicer

In a 10 μl reaction, 5 ng in vitro transcribed pre-miRNA substrates were incubated with 0.25 μg recombinant human Dicer in 1× Dicing buffer (20 mM Tris-HCl, pH 7.5, 1.5 mM MgCl$_2$, 25 mM NaCl, 1% glycerol, and 1 mM DTT) at 37°C for 0–40 min. RNAs were ethanol precipitated and analyzed by northern blot.

Immunoprecipitation and Antibodies

α-Ago monoclonal antibody (clone 2A8) was used for immunoprecipitation of extracts prepared from 293T cells transfected with miR-320a duplexes. α-m^{7}G antibody was used for immunoprecipitation of extracts prepared from 293T cells overexpressing miR-320a. For NIE-115 cells, immunoprecipitation with α-XPO5 or α-PHAX antibodies were carried out as described with minor modifications (McCloskey et al., 2012).

ACCESSION NUMBERS

Small RNA Cap-seq data has been deposited in the Sequence Read Archive under accession number SRP032437.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.11.027.

REFERENCES

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. *22*, 2773–2785.

Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265–276.

Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlers, M., and Cullen, B.R. (2010). A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral microRNAs. Mol. Cell *37*, 135–142.

Bronisz, A., Godlewski, J., Wallace, J.A., Merchant, A.S., Nowicki, M.O., Mathsyaraja, H., Srinivasan, R., Trimboli, A.J., Martin, C.K., Li, F., et al. (2012). Reprogramming of the tumour microenvironment by stromal PTEN-regulated miR-320. Nat. Cell Biol. *14*, 159–167.

Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2010). A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. Genome Res. *20*, 1398–1410.

Büssing, I., Yang, J.S., Lai, E.C., and Grosshans, H. (2010). The nuclear export receptor XPO-1 supports primary miRNA processing in C. elegans and Drosophila. EMBO J. *29*, 1830–1839.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Castanotto, D., Lingeman, R., Riggs, A.D., and Rossi, J.J. (2009). CRM1 mediates nuclear-cytoplasmic shuttling of mature microRNAs. Proc. Natl. Acad. Sci. USA *106*, 21655–21659.

Cazalla, D., Xie, M., and Steitz, J.A. (2011). A primate herpesvirus uses the Integrator complex to generate viral microRNAs. Mol. Cell *43*, 982–992.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. (2010). A Dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature 465, 584–589.

Chen, X., Taylor, D.W., Fowler, C.C., Galan, J.E., Wang, H.W., and Wolin, S.L. (2013). An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. Cell 153, 166–177.

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.

Choi, Y.H., and Hagedorn, C.H. (2003). Purifying mRNAs with a high-affinity eIF4E mutant identifies the short 3′ poly(A) end phenotype. Proc. Natl. Acad. Sci. USA 100, 7033–7038.

Chong, M.M., Zhang, G., Cheloufi, S., Neubert, T.A., Hannon, G.J., and Littman, D.R. (2010). Canonical and alternate functions of the microRNA biogenesis machinery. Genes Dev. 24, 1951–1960.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D., et al. (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science 328, 1694–1698.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845–1848.

Denti, M.A., Rosa, A., Sthandier, O., De Angelis, F.G., and Bozzoni, I. (2004). A new vector, based on the PolII promoter of the U1 snRNA gene, for the expression of siRNAs in mammalian cells. Mol. Ther. 10, 191–199.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell 150, 100–110.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. Mol. Cell 32, 519–528.

Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D., and Lai, E.C. (2010). MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol. Cell 38, 900–907.

Fok, V., Friend, K., and Steitz, J.A. (2006). Epstein-Barr virus noncoding RNAs are confined to the nucleus, whereas their partner, the human La protein, undergoes nucleocytoplasmic shuttling. J. Cell Biol. 173, 319–325.

Garzon, R., Calin, G.A., and Croce, C.M. (2009). MicroRNAs in cancer. Annu. Rev. Med. 60, 167–179.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N., and Kay, M.A. (2012a). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of Dicer processing in vivo. Cell 151, 900–911.

Gurtan, A.M., Lu, V., Bhutkar, A., and Sharp, P.A. (2012). In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. RNA 18, 1116–1122.

Haussecker, D., Cao, D., Huang, Y., Parameswaran, P., Fire, A.Z., and Kay, M.A. (2008). Capped small RNAs and MOV10 in human hepatitis delta virus replication. Nat. Struct. Mol. Biol. 15, 714–721.

Iwasaki, Y.W., Kiga, K., Kayo, H., Fukuda-Yuzawa, Y., Weise, J., Inada, T., Tomita, M., Ishihama, Y., and Fukao, T. (2013). Global microRNA elevation by inducible Exportin 5 regulates cell cycle entry. RNA 19, 490–497.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell 115, 209–216.

Kim, D.H., Saetrom, P., Snøve, O., Jr., and Rossi, J.J. (2008). MicroRNA-directed transcriptional gene silencing in mammalian cells. Proc. Natl. Acad. Sci. USA 105, 16230–16235.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. 10, 126–139.

Lee, Y.S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev. 23, 2639–2649.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science 303, 95–98.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature 444, 953–956.

Marcotrigiano, J., Gingras, A.C., Sonenberg, N., and Burley, S.K. (1997). Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell 89, 951–961.

Martinez, N.J., Ow, M.C., Reece-Hoyes, J.S., Barrasa, M.I., Ambros, V.R., and Walhout, A.J. (2008). Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. Genome Res. 18, 2005–2015.

McCloskey, A., Taniguchi, I., Shinmyozu, K., and Ohno, M. (2012). hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export. Science 335, 1643–1646.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 25, 742–754.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 18, 610–621.

Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. Nat. Struct. Mol. Biol. 15, 902–909.

Ohno, M., Segref, A., Bachi, A., Wilm, M., and Mattaj, I.W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell 101, 187–198.

Okada, C., Yamashita, E., Lee, S.J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y., and Tsukihara, T. (2009). A high-resolution structure of the pre-microRNA nuclear export machinery. Science 326, 1275–1279.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 130, 89–100.

Pan, Q., van der Laan, L.J., Janssen, H.L., and Peppelenbosch, M.P. (2012). A dynamic perspective of RNAi library development. Trends Biotechnol. 30, 206–215.

Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5′ end of RNA for efficient and accurate processing. Nature 475, 201–205.

Pawlicki, J.M., and Steitz, J.A. (2008). Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. J. Cell Biol. 182, 61–76.

Rodriguez, M.S., Dargemont, C., and Stutz, F. (2004). Nuclear export of RNA. Biol. Cell 96, 639–655.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature 448, 83–86.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science 336, 1037–1040.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208.

Topisirovic, I., Svitkin, Y.V., Sonenberg, N., and Shatkin, A.J. (2011). Cap and cap-binding proteins in the control of gene expression. RNA 2, 277–298.

Wang, K., Long, B., Jiao, J.Q., Wang, J.X., Liu, J.P., Li, Q., and Li, P.F. (2012). miR-484 regulates mitochondrial network through targeting Fis1. Nat Commun 3, 781.

Weiner, A.M. (2005). E Pluribus Unum: 3′ end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. Mol. Cell 20, 168–170.

White, E., Kamieniarz-Gdula, K., Dye, M.J., and Proudfoot, N.J. (2013). AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination. Nucleic Acids Res. 41, 1797–1806.

Xhemalce, B., Robson, S.C., and Kouzarides, T. (2012). Human RNA methyltransferase BCDIN3D regulates microRNA processing. Cell *151*, 278–288.

Yang, J.S., and Lai, E.C. (2011). Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol. Cell *43*, 892–903.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc. Natl. Acad. Sci. USA *107*, 15163–15168.

Yang, J.S., Phillips, M.D., Betel, D., Mu, P., Ventura, A., Siepel, A.C., Chen, K.C., and Lai, E.C. (2011). Widespread regulatory activity of vertebrate microRNA* species. RNA *17*, 312–326.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. *17*, 3011–3016.

Zhou, H., Arcila, M.L., Li, Z., Lee, E.J., Henzler, C., Liu, J., Rana, T.M., and Kosik, K.S. (2012). Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res. *40*, 5864–5875.

EXTENDED EXPERIMENTAL PROCEDURES

Multiple Sequence Alignment

Mammalian miR-320 and murine-specific miR-344 sequences were obtained from the UCSC genome database (Meyer et al., 2013). The sequence alignment was performed with the BioEdit program using the ClustalW algorithm. For miR-320 alignment, species included are *Homo sapiens* (Human), *Mus musculus* (Mouse), *Pteropus vampyrus* (Megabat), *Erinaceus europaeus* (Hedgehog), *Dasypus novemcinctus* (Armadillo), *Ailuropoda melanoleuca* (Panda), *Bos taurus* (Cow), *Sus scrofa* (Pig), *Loxodonta africana* (Elephant) and *Trichechus manatus latirostris* (Manatee).

Plasmids

For expression of m^7G-capped pre-miRNAs in 293T cells, genomic regions encoding human miR-320a and mouse miR-344, including their upstream promoters (606 nt for miR-320a and 301 nt for miR-344) and downstream sequences (179 nt for miR-320a and 107 nt for miR-344), were cloned into pBluescript (pBS) between the EcoRI and XhoI restriction sites. These plasmids are named pBS-miR-320a and pBS-miR-344, respectively. The putative TATA box for miR-320a was mutated from 5′-TATTTAT-3′ to 5′-GCGGGCG-3′ to generate pBS-miR-320a mut-TATA. To express miR-320a in a Drosha-dependent manner, the 59 nt pre-miR-142 hairpin was replaced by the 54 nt pre-miR-320a hairpin on pcDNA3-pri-miR-142 (Cazalla et al., 2011). The plasmid pGST-4E-K119A for expressing recombinant eIF4E with high cap-binding ability (designated as GST-4E wild-type) in *E. coli* BL21 was a generous gift from Dr. Curt Hagedorn (Choi and Hagedorn, 2003). For GST-4E W102L expression, the codon for tryptophan 102 was changed to a leucine codon by site-directed mutagenesis. Plasmids expressing hvsA-miR-HSUR4, pri-miR-142, or EBER-1 were described in (Cazalla et al., 2011). To generate pU1-miR-HSUR4 for the synthesis of m^7G-capped pre-miR-HSUR4 from a U1 promoter, the 3′ box and flanking sequences were removed from pU1-HSUR4ΔsnRNA (Cazalla et al., 2011) to position the pre-miR-HSUR4 hairpin directly downstream of the U1 promoter. To generate pU6-miR-HSUR4 for the synthesis of pre-miR-HSUR4 from the U6 promoter, a murine U6 promoter followed by the pre-miRNA hairpin was cloned into the pTY lentiviral vector between the SalI and MluI sites. pGFP-miR-HSUR4-5p and pdsRED-miR-HSUR4-3p reporter plasmids were generated from pTY and pBS vectors, respectively; the 3′ UTR of each fluorescent protein gene contained 2 sites perfectly complementary to miR-HSUR4-5p or −3p. MiR-HSUR4-5p and −3p target sites were swapped on the GFP and dsRED reporter plasmids to generate pGFP-miR-HSUR4-3p and pdsRED-miR-HSUR4-5p. To generate the pBS-miR-320a/miR-142 reporter, pri-miR142 and the BGH (bovine growth hormone)-polyA sequence from pcDNA3-pri-miR-142 (Cazalla et al., 2011) were cloned 20 nt downstream of the pre-miR-320a hairpin on pBS-miR-320a using site-directed mutagenesis. The sequence for the schematic shown in Figure 5A is: **GCCTTCTCTTCCCGGTTCTTCCCGGAGTCGGGAAAAGCTGGGTTGAGAGGGCGAAAAAGGATGAGGTGACTGGT***GAATCCCC GTGGACAGACAGACAGTGCAGTCACCCATAAAGTAGAAAGCACTACTAACAGCACTGGAGGGTGTAGTGTTTCCTACTTTATGGA TGAGTGCACTGTGGGCTTCGGAGACCACGCCACGCCGCGGCCCCCTGCCACCGTCCTCCACTCCAACATCAAGGAAACAGGTG AGCATCTCTCACACGCCTCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGC*<u>TAGAGCTCGCTGATCAGCCTCGACTGTGCCTT CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAA AATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT TGGGAAGACAATAGCAGGCATGCTGGGGA</u> (**Bold: pre-miR-320a plus 20 nt downstream sequence**; *Italics: pri-miR-142*; <u>Underlined: BGH-polyA sequence).</u> From pBS-miR-320a/miR-142, five mutant plasmids were derived: 1, mut TATA: the TATA box of miR-320a promoter was mutated as in Figure 1B; 2, unpair hairpin: the 3p arm of the pre-miR-320a sequence was mutated to its complementary sequence to eliminate the hairpin structure; 3, mut A-rich tract: the 20 nt A-rich tract was mutated to 5′-CTCGAGGGGGGGCCCGGTAC-3′; 4, double mutant containing both mut TATA and unpair hairpin; and 5, mut-5p: the 5p arm of the pre-miR-320a sequence was mutated to its complementary sequence.

Cell Culture, Plasmid Transfection, and Transient RNAi Knockdown

HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% heat-inactivated fetal bovine serum (FBS), 1% penicillin and streptomycin, 6 mM L-glutamine at 37°C with 5% CO_2. For plasmid transfections in 6-well plates, 2 μg plasmid and 5 μl Lipofectamine 2000 (Invitrogen) or TransIT-293 transfection reagent (Mirus Bio) were used in one well containing 2×10^5 cells; after 48 hr, total RNA was harvested using the Trizol reagent (Invitrogen) and analyzed by Northern blotting of transfected or endogenous RNAs indicated in figures. SiRNAs targeting specific mRNAs were transfected into 1×10^5 cells/well in 6-well plates using 5 μl Lipofectamine RNAiMAX reagent (Invitrogen). Forty-eight hr later, 1 μg pBS-miR-344 or pBS-miR-320a, 0.8 μg pcDNA3-pri-miR-142 and 0.2 μg pEBER-1 were transfected into the cells using 5 μl Lipofectamine 2000 reagent or TransIT-293. Total RNA was harvested and analyzed by Northern blotting of the transfected RNAs after another 48 hr; total protein was analyzed by western blot. The siRNAs (Integrated DNA Technology) used in this study are listed in Table S1. We noted that knockdown of PHAX significantly reduced cell viability even within the first 48 hr; therefore, because of the low signal for endogenous m^7G-capped pre-miRNAs, it was not possible to assess changes in level by either Northern blot or RT-qPCR analysis. For the miR-HSUR4 fluorescent protein reporter assays, 1.5 μg pGFP-miR-HSUR4-5p and 0.5 μg pdsRED-miR-HSUR4-3p reporter plasmids were co-transfected with either 2 μg pU1-HSUR4 (Cazalla et al., 2011) or 0.4 μg pU1-miR-HSUR4 plus 1.6 μg pBS or 0.2 μg pU6-miR-HSUR4 plus 1.8 μg pBS, using 5 μl Lipofectamine 2000 reagent or TransIT-293 into 2×10^5 293T cells/well in a 6-well plate. Cell images were captured using a fluorescent microscope 48 hr posttransfection. Total RNA was extracted by Trizol and analyzed by

Northern blot. For transfection of the fluorescent protein reporters with miR-HSUR4 target sites swapped, 1.5 μg pGFP-miR-HSUR4-3p and 0.5 μg pdsRED-miR-HSUR4-5p reporter plasmids were cotransfected with the miR-HSUR4 expression vectors described above.

5′ RACE

Total RNA was extracted from pBS-miR-320a-transfected 293T cells and size-fractionated on an 8 M urea-8% polyacrylamide gel to obtain 650 ng of 40–100 nt small RNAs. After treating with 5 U alkaline phosphatase (CIP; NEB) at 37°C for 30 min (CIP converts many 5′ functional groups, such as monophosphate and triphosphates, but not the m^{7}G cap, to hydroxyl, which does not ligate to the 5′ adaptor), small RNAs were phenol/chloroform extracted twice, ethanol precipitated and incubated with 25 U tobacco acid pyrophosphatase (TAP; EPICENTRE) at 37°C for 1 hr (TAP converts an m^{7}G cap to a 5′-monophosphate for subsequent 5′ adaptor ligation). TAP-treated RNAs were phenol/chloroform extracted, ethanol precipitated and ligated with 100 pmole of the 5′ adaptor using 5 U T4 RNA ligase (Fermentas) at 16°C overnight. Reverse transcription was carried out using SuperScript III Reverse Transcriptase (SSIII RT, Invitrogen) with a miR-320a-specific primer (P3, Table S1) according to the manufacturer's instructions. PCR amplification was performed with Phusion polymerase and GC buffer (NEB) using the primers shown in Figure 1D and Table S1. PCR cycles were as follows: initial denaturation 98°C 30 s, [98°C 10 s, 60°C 20 s, 72°C 20 s] for 35 cycles, and finally 72°C 1 min. Desired PCR products were gel purified and blunt-end cloned into the pcDNA3 vector and sequenced.

M^{7}G-Capped Small RNA Purification, Small RNA Cap-Seq Library Preparation, and Deep Sequencing

GST-tagged eIF4E was recombinantly expressed in *E. coli* BL21 (Choi and Hagedorn, 2003) and crude lysate prepared by sonication was loaded onto glutathione sepharose beads (GE Healthcare). For each capped small RNA purification, ~25 μg total RNA from 293T cells overexpressing miR-320a or miR-344 was incubated with approximately 50 μg GST-4E immobilized on 20 μl glutathione sepharose beads in 500 μl NET-2 buffer (150 mM NaCl, 0.05% NP-40, 50 mM Tris-HCl, pH 7.4) for 2 hr at 4°C with gentle agitation. The beads were washed 5 times with 1 ml NET-2 buffer and captured RNAs were phenol/chloroform extracted and subjected to Northern blot analysis.

For library preparation, total RNA was extracted from a P1 mouse (C57BL/6 strain, a gift from Drs. Yongquan Zhang and Sreeganga Chandra) using Trizol reagent and size-fractionated on an 8 M urea-8% polyacrylamide gel to obtain 80 μg of 40–100 nt small RNAs. Forty μg purified small RNAs were then incubated with or without 40 μg recombinant *Deinococcus radiodurans* polynucleotide phosphorylase (PNPase, a gift from Drs. Xinguo Chen and Sandra Wolin) (Chen et al., 2013) in the presence of 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM MgCl$_2$, 10 mM KPO$_4$ pH 7.4 and 2 mM DTT at 30°C for 30 min. About 2.5 μg PNPase-treated or -untreated small RNAs were analyzed on a Northern blot. The remaining small RNAs were incubated with 20 μl GST-4E or GST-4E W102L coated glutathione beads prepared as described above (~12 μg small RNAs and ~50 μg GST-4E per incubation). After 5 washes with NET-2 buffer, 6 pmole of 5′ ^{32}P-labeled 3′ adaptor was added to the beads and ligation was carried out by T4 RNA ligase at 16°C. After 2 hr, 60 pmole unlabeled 3′ adaptor was added and incubation was continued overnight. The ligated RNAs were separated on an 8 M urea-8% polyacrylamide gel and the desired fraction was excised from the gel, eluted and ethanol precipitated. 3′ adaptor ligated RNA (~400 ng) was CIP-TAP treated and ligated with the 5′ adaptor as described in 5′ RACE. The first strand cDNA was synthesized using SSIII RT with a primer that would specifically anneal to the 3′ adaptor (Table S1). Two PCR reactions were performed subsequently using Phusion polymerase and GC buffer (NEB) with cycling as follows: 98°C 30 s, [98°C 10 s, 60°C 20 s, 72°C 25 s] repeated for 26 or 6 cycles, and finally 72°C for 1 min. In the first PCR (26 cycles), cDNA was amplified by primers that anneal to the 5′ and 3′ adaptors. In the second PCR (6 cycles), purified PCR products from the first PCR were further amplified by primers compatible with the Illumina sequencing platform. Adaptor and primer sequences are listed in Table S1. Both PCR reactions were analyzed on an 8 M urea-6% polyacrylamide gel and were not saturated. Finally, PCR products of the expected size from the second reaction were purified and sequenced at the Yale Stem Cell Center Genomics Core on an Illumina HiSeq2000 instrument using 50 bp runs.

Bioinformatic Analysis

Using custom PERL scripts and fastx_clipper program from FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/download.html), 34,478,081 high-quality reads were clipped at the 5′ end to remove the barcode sequence (5′-NNNNG-3′) and at the 3′ end to remove the adaptor sequence. To perform the alignment, we used the software Tophat (version 1.0.3) allowing 2 mismatches and set the multiple hit parameter to 500 to find all probable mapping loci for short reads (Babiarz et al., 2008; Trapnell et al., 2009). In total, 14,481,418 reads of 25-45 nt were aligned to the mouse reference genome (NCBI37/mm9, released July 2007); (http://genome.ucsc.edu/) and were stored in the BAM format for further analyses. The seqLogo function in R software was used for motif analysis and to identify the expected TSS-specific yr motif for the mapped reads (http://www.r-project.org/). First, we calculated the frequency of A/T/C/G at each position for 20 nt flanking each mapped TSS. Next, the nucleotide enrichment level was quantified by calculating the log2 ratio of the observed frequency relative to the expected frequency, which is defined as the nucleotide frequency of all genomic sites for 200 nt flanking all mapped TSSs (Gu et al., 2012b). The enriched nucleotides in each position; i.e., log2 ratio larger than 0, were visualized by seqLogo, in which the nucleotide heights are proportional to their enrichment levels.

The miRNA annotation for the mm9 reference genome was retrieved from miRBase (version 18) (Kozomara and Griffiths-Jones, 2011). Annotations for other genes were retrieved from the Ensembl genome browser (NCBI37/mm9, version 63, June 2011) (Flicek

et al., 2012). Reads that mapped to the first 500 nt of protein-coding gene transcripts or of long noncoding RNAs (lncRNAs) were classified as mRNAs and lncRNAs. Capped small RNAs included snRNAs (except for U6) and U3, U8 and U13 snoRNAs. Uncapped RNAs included more 3′ fragments of mRNAs and lncRNAs (after the first 500 nt), U6 snRNA, snoRNAs (except for U3, U8 and U13), tRNAs and rRNAs. In addition, we manually built the annotation for short divergent transcripts of mRNAs and lncRNAs by including reads mapping up to 100 nt upstream from the TSS on the antisense strand. Subsequently, the software packages SAMtools (version 0.1.17); http://samtools.sourceforge.net/samtools.shtml and RSEQtools (version 0.5); http://info.gersteinlab.org/RSEQtools were used to calculate the number of reads (read count) falling within the genomic regions of these annotation categories (Habegger et al., 2011; Li et al., 2009). Briefly, the BAM format alignment was first converted into SAM format alignment, and then the "sam2mrf" function in RSEQtools was used to convert the SAM format to the Mapped Read Format (MRF), which compactly represents the mapped reads. Second, the "mrfQuantifier" function in RSEQtools was used to quantify the read count for each annotation entry by counting all aligned reads overlapping the given annotation entry. To comparatively visualize the read count data in UCSC genome browser, the "mrf2wig" function in RSEQtools and "wigToBigWig" function in UCSC utilities were used to build a track file with the normalized read count; i.e., the raw read count was divided by the total number per million mapped reads. To identify probable m^7G-capped pre-miRNAs, we downloaded publicly released miRNA sequencing data of newborn mouse whole body samples (Chiang et al., 2010) and aligned these sequences to the mm9 reference genome using the settings described above. Through manual inspection using the UCSC genome browser, pre-miRNAs containing Cap-seq reads that matched within 5 nt of their predicted 5′ ends on miRBase were identified as candidate m^7G-capped pre-miRNAs. Candidates not located in annotated 5′ exons of a mRNA were selected for experimental validation by RT-qPCR.

To investigate off-target effects caused by 5p-siRNA derived from a U6-promoter-driven shRNA, we analyzed the public data set (Luo et al., 2012), in which RNA-seq data were obtained from wild-type human 293T cells or cells expressing a U6-promoter-driven shRNA targeting AFF2 (Thermo scientific, V2THS_113929), AFF3 (V2THS_133851), or AFF4 (V2THS_197522). The data sets were deposited at GEO under the accession number GSE34097. As an example, we chose the four wild-type and two AFF2-shRNA RNA-seq data sets to examine shRNA off-target effects.

We used R bioconductor DESeq to identify the differentially expressed genes between wild-type and AFF2-shRNA expressing cells (Anders and Huber, 2010). The annotated genes were retrieved from GENCODE (released version 10) (http://www.gencodegenes.org/). The SAMtools (version 0.1.17) and RSEQtools (version 0.5) software packages were used to calculate expression of RPKM and reads count as described above (Habegger et al., 2011; Li et al., 2009). To reduce false-positive results, we focused on protein-coding genes with RPKM $\geq$ 1 in at least one biological condition. When performing the comparison, DESeq first receives the mean reads count of biological replicates as a joint expression estimate for both groups, and then calculates the difference as well as the p value for the statistical significance of this change. The adjusted p value was also calculated based on multiple testing with the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), which controls the false discovery rate (FDR). We set FDR < 0.01 to detect reliable DEX genes. In total, we identified 273 down- and 181 upregulated genes.

To investigate whether gene downregulation was caused by the 3p-siRNA (desired strand) or 5p-siRNA (passenger strand that would cause an off-target effect) from AFF2-shRNA, we counted the percentages of downregulated genes whose 3′-UTRs contain perfect complementary sequences for heptamer seed regions (positions 1–7 or 2–8) of the 3p- and 5p-siRNA, respectively. For comparison, we performed the same statistical analyses for the upregulated genes. To simulate random probability, we randomly picked 200 genes, around the mean of up-/downregulated genes, from all protein-coding genes and calculated the percentage of genes containing heptamer seed regions. This operation was repeated 1000 times to obtain the simulated percentages (mean $\pm$ SEM).

RT-qPCR

RNA from mouse neuroblastoma NIE-115 cells (cultured as above for 293T cells) was purified using Trizol reagent and approximately 45 μg size-fractionated 40–100 nt RNAs were subjected to pulldown with either wild-type or W102L mutant GST-4E as described above. After phenol/chloroform extraction and ethanol precipitation, RNA was treated with DNase RQ1 (Promega), subjected to another round of phenol/chloroform extraction and ethanol precipitation, and resuspended in double-distilled water. The same volume (corresponding to < 5 μg of RNA) from each sample was used for cDNA synthesis with SSIII RT and random primers (Invitrogen). The cDNA obtained was analyzed by quantitative real-time PCR using FastStart Universal SYBR Green Master (Rox) master mix (Roche) and the primers listed in Table S1 on a StepOnePlus instrument (Applied Biosystems) in technical triplicates. Primers were designed using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and blasted (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to eliminate off-target products. From each experiment, one control with and one without SSIII RT were included in the qPCR analysis to control for genomic contamination. The pulldown (PD) value for each sample was normalized to its supernatant (SUP). For each experiment, all normalized values were then normalized to that of pre-let-7b. The experiment was performed three times. Statistical significance was calculated by Student's t test and results were considered significant if p < 0.05.

Microinjection into *Xenopus* Oocytes

Stage V and VI oocytes were isolated from female *X. laevis* and maintained in 1X modified Barth's saline (MBS) (1 mM HEPES pH 7.9, 80 mM NaCl, 1 mM KCl, 0.8 mM $MgSO_4$, 2.4 mM $NaHCO_3$ and 0.7 mM $CaCl_2$) at 18°C. Nuclei were injected with 18.4 nL solution

containing 20 mg/ml blue dextran, 1-10 fmole ^{32}P-labeled RNAs with or without 1 pmole competitor RNA. After 2 hr, nuclei were manually isolated from the cytoplasm in 5:1+PO$_4$ buffer (83 mM KCl, 17 mM NaCl, 6.5 mM Na$_2$HPO$_4$, 3.5 mM KH$_2$PO$_4$ and 10 mM MgSO$_4$). Six blue nuclei and the corresponding cytoplasms were collected separately. Total nuclear RNA was extracted directly with Trizol reagent. For cytoplasmic RNAs, the cytoplasms were homogenized, and one-oocyte equivalent of material was treated with 2 mg/ml proteinase K in a 300 μl reaction containing 50 mM Tris-HCl pH7.4, 5 mM EDTA, 300 mM NaCl and 1.5% SDS at 37°C for 15 min. RNAs were extracted with Trizol reagent after proteinase K digestion. One-oocyte equivalent of RNA was separated on an 8 M urea-15% polyacrylamide gel and analyzed by PhosphoImager. To quantitate RNA export efficiency, nuclear and cytoplasmic RNAs were first normalized to the corresponding EBER-1 levels. Then, export efficiency was calculated as cytoplasmic RNA / (nuclear RNA + cytoplasmic RNA). The relative export efficiency of each RNA with competition was normalized to its export efficiency without competition.

NRO Assay

Run-on assays were performed as described (Conrad and Steitz, 2005), except that the probes used in this study were dsDNA generated by PCR. The probes spanned the entire transcribed regions of the HSUR4, U1 and pri-mi142 (including downstream BGH-polyA region).

Immunoprecipitation and Antibodies

α-Ago monoclonal antibody (clone 2A8, kindly provided by Dr. Zissimos Mourelatos) was used for immunoprecipitation from extracts prepared by sonication of 2×10^6 293T cells transfected with 0.5 nM perfectly base-paired miR-320a duplexes with 2 nt 3′ overhangs for 24 hr. α-m^{7}G antibody (kindly provided by Drs. Elisabetta Ullu and Chris Tschudi) was used for immunoprecipitation from extracts prepared by sonication of 2×10^6 293T cells transfected with 2 μg pBS-miR-320a for 48 hr. Each immunoprecipitation reaction contained 15 μl antibody-coated protein G sepharose beads (GE Healthcare) and 500 μl 293T cell extract at 4°C for 4 hr. Coimmunoprecipitated RNAs were extracted by phenol/chloroform and analyzed by Northern blot. For XPO5 and PHAX immunoprecipitation, 4×10^7 NIE-115 cells were washed twice with ice-cold phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 19 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$) and crosslinked in PBS containing 1% formaldehyde for 10 min at room temperature. The reaction was quenched by the addition of glycine to a final concentration of 0.25 M. After two washes with PBS, the cell pellet was resuspended in 1.5 ml RIPA buffer [10 mM Tris-HCl pH7.5, 150 mM NaCl, 2.5 mM MgCl$_2$, 1% Triton X-100, 0.1% SDS, 0.5% Sodium deoxycholate, 1 mM DTT, and 1 X protease inhibitor (Calbiochem)] and cell extract was prepared by gentle agitation at 4°C for 30 min. Supernatant was collected after 10 min centrifugation at 16,000 xg. For each immunoprecipitation reaction, 7.5 μg α-PHAX antibody (Abcam ab157096) or α-XPO5 antibody (Abcam ab131281) was mixed with 1.5 ml cell lysate at 4°C overnight. Fifteen microliters protein A sepharose beads (GE Healthcare) were then added to the antibody-lysate mixture for 2 hr at 4°C. The beads were washed 5 times with NET-2 buffer and incubated with 50 mM Tris-Cl (pH7.5), 5 mM EDTA, 1.5% SDS, 300 mM NaCl, 1.5 mg/ml proteinase K for 30 min at 50°C, and then for 45 min at 70°C. RNAs were recovered and analyzed by RT-qPCR as above.

Antibodies against PHAX (Abcam ab157096), Exportin5 (Abcam ab131281, ab57491), Integrator11 (Abcam ab75276), CPSF73 (Santa Cruz), Dis3 (Abnova H00022894-B01P), GAPDH (Cell Signaling 14C10) and Tubulin (Sigma) were used for western blotting.

SUPPLEMENTAL REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. J. R. Stat. Soc. *57*, 289–300.

Conrad, N.K., and Steitz, J.A. (2005). A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. EMBO J. *24*, 1831–1841.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2012). Ensembl 2012. Nucleic Acids Res. *40* (Database issue), D84–D90.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012b). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell *151*, 1488–1500.

Habegger, L., Sboner, A., Gianoulis, T.A., Rozowsky, J., Agarwal, A., Snyder, M., and Gerstein, M. (2011). RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics *27*, 281–283.

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. *39* (Database issue), D152–D157.

Lemm, I., Girard, C., Kuhn, A.N., Watkins, N.J., Schneider, M., Bordonné, R., and Lührmann, R. (2006). Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. Mol. Biol. Cell *17*, 3221–3231.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol. Cell. Biol. *32*, 2608–2617.

Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., et al. (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. *41* (Database issue), D64–D69.

Preti, M., O'Donohue, M.F., Montel-Lehry, N., Bortolin-Cavaillé, M.L., Choesmel, V., and Gleizes, P.E. (2013). Gradual processing of the ITS1 from the nucleolus to the cytoplasm during synthesis of the human 18S rRNA. Nucleic Acids Res. *41*, 4709–4723.

Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepien, P.P., et al. (2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J. *29*, 2342–2357.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Zhu, Z.H., Yu, Y.P., Shi, Y.K., Nelson, J.B., and Luo, J.H. (2009). CSR1 induces cell death through inactivation of CPSF3. Oncogene *28*, 41–51.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415.
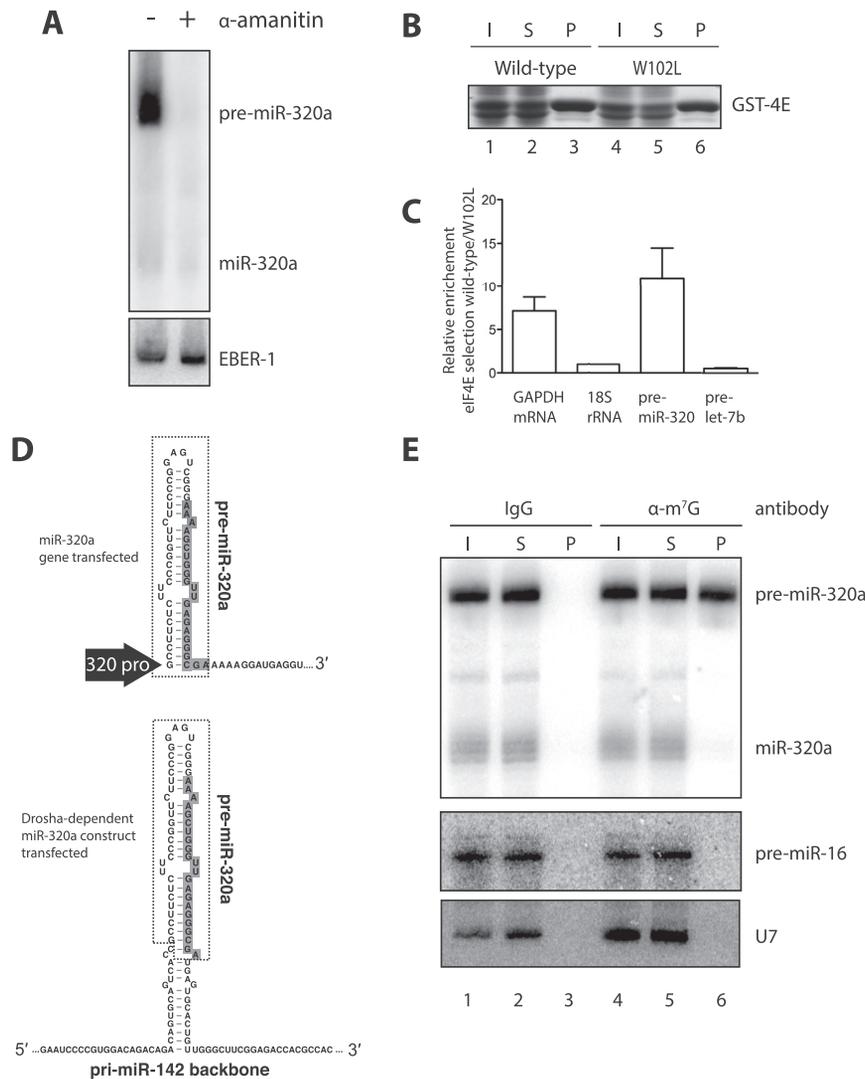
Figure S1. Additional Evidence that Human Pre-miR-320a Contains 5′ m⁷G Cap, Related to Figure 1

(A) Northern blot analysis of the expression of miR-320a and EBER-1 in 293T cells cotransfected with pBS-miR-320a and pEBER-1 for 24 hr and then treated with 1.6% DMSO or 1.6 μM α-amanitin in DMSO for another 24 hr. EBER-1 is a RNA Pol III transcript.

(B) Coomassie blue stained SDS-PAGE shows GST-4E immobilized on glutathione sepharose beads from *E. coli* BL21 lysate overexpressing recombinant wild-type or W102L GST-4E. I: input; S: supernatant; P: pellet; I and S were both 10% relative to P.

(C) Total RNA from mouse NIE-115 cells was subjected to GST-4E selection, followed by RT-qPCR quantitation of the indicated RNAs. The bars show the relative enrichment of RNAs bound by GST-4E WT compared to W102L, with 18S rRNA set to 1. Error bars represent standard deviation from three experiments.

(D) Schematic of miR-320a expression constructs used in Figure 1C. The sequence of pre-miR-320a is boxed with a dashed line, with mature miR-320a shaded gray.

(E) Total RNA extracted from 293T cells transfected with pBS-miR-320a was subjected to immunoprecipitation using α-m⁷G antibodies or rabbit IgG and analyzed by Northern blotting to detect transfected pre-miR-320a (5′ m⁷G-cap), endogenous pre-miR-16 (5′ monophosphate) and U7 RNA (5′ m²,²,⁷G trimethylguanosine cap, TMG). I: input; S: supernatant; P: pellet; I and S were both 5% relative to pellet.
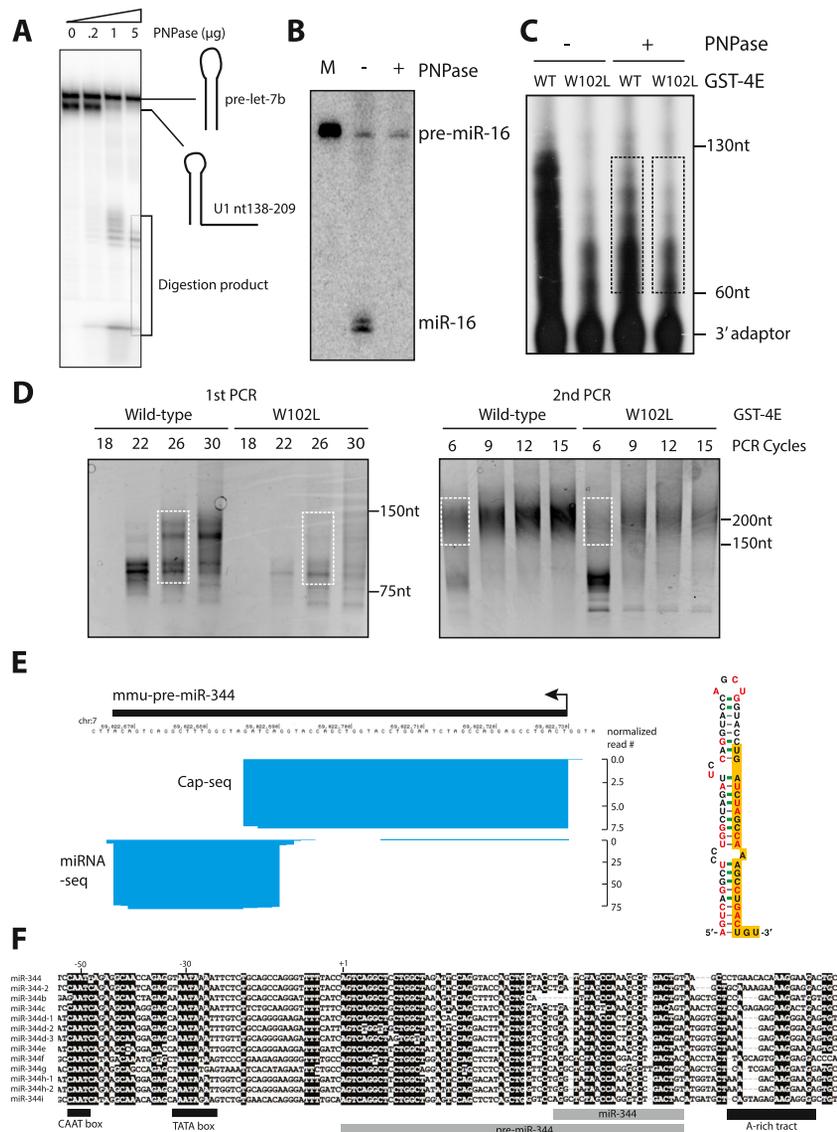
Figure S2. Small RNA Cap-Seq Library Preparation, Related to Figure 2

(A) RNA hairpins with or without a 3′ tail were body-labeled with α-^{32}P UTP and exposed to different amounts of PNPase for 30 min. Secondary structures of the substrates are shown on the right.

(B) Northern blot using a miR-16 probe of size-fractionated (40–100 nt) P1 mouse total RNA with or without PNPase treatment. M: In vitro-transcribed pre-miR-16 marker. Mature miR-16 was observed due to contamination from smaller sized RNAs in PAGE size extraction.

(C) A size-fractionation of mouse total RNA that was untreated or treated with PNPase and then ligated to a ^{32}P-labeled 3′ adaptor. RNAs extracted for subsequent 5′ adaptor ligation were within the dashed-line boxes.

(D) Ethidium bromide-stained denaturing polyacrylamide gels show DNA products from the 1st and 2nd PCR amplification steps of the small RNA Cap-seq library preparation. Amplification cycle numbers are indicated on the top of the gels. DNAs from unsaturated PCR reactions that were extracted for subsequent PCR or sequencing analyses are indicated by dashed-line boxes.

(E) Histogram representing small RNA Cap-seq and mature miRNA reads (Chiang et al., 2010) mapping to the mmu-miR-344 locus. The black bar indicates the pre-miRNA. The transcription start site and transcription directionality are indicated by the black arrow. Reads mapping to the negative strand are colored blue, with normalized read numbers drawn to scale. Due to the short length of the clipped sequencing reads (45 nt), the 3′ ends of the Cap-seq reads do not represent the 3′ ends of the pre-miR-344 RNAs. The predicted pre-miR-344 hairpin is illustrated. 100% conserved residues among 13 miR-344 family members are highlighted in red. Base-pairs supported by covariation are highlighted in green.

(F) Multiple sequence alignment of the 13 members of the mmu-miR-344 family. Residues that are more than 90% conserved are in black. +1 denotes the first nucleotide of the predicted pre-miR-344 hairpins. Conserved promoter elements including a putative CAAT box and a TATA box are indicated with black bars. A downstream A-rich region is noted.
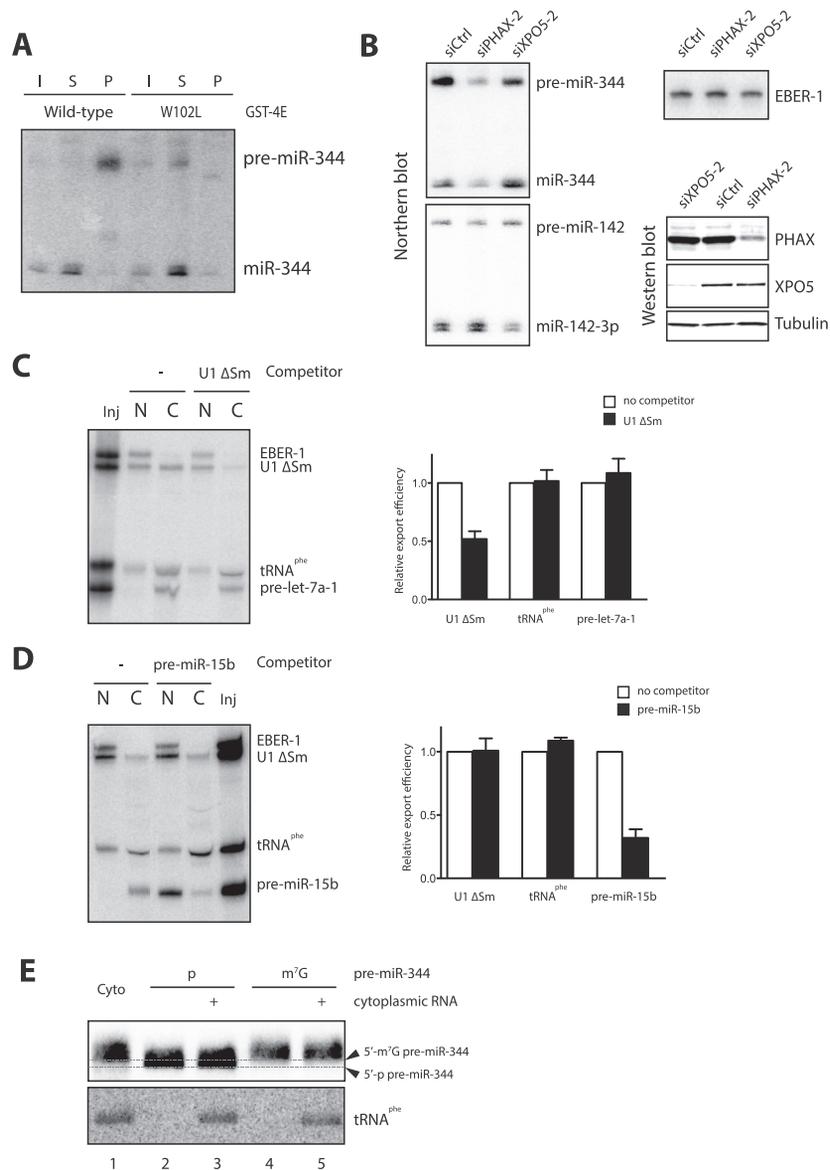
Figure S3. Additional Evidence that m^{7}G-Capped Pre-miR-344 Is Exported by Exportin 1, Related to Figure 3

(A) Total RNA extracted from 293T cells transfected with pBS-miR-344 was subjected to GST-4E affinity purification. Isolated RNA was analyzed on a Northern blot probed for miR-344. I: input; S: supernatant; P: pellet; both I and S were 5% relative to P.

(B) Northern blots analyzing the levels of miR-344, miR-142-3p and EBER-1 in transiently transfected 293T cells treated with a control siRNA (siCtrl) or an siRNA against PHAX [siPHAX-2 (Lemm et al., 2006)] or exportin 5 [siXPO5-2 (Lund et al., 2004)] for 48 hr prior to cotransfection with plasmids expressing the three RNAs. Western blots show the extent of knockdown of PHAX and XPO5, with tubulin as a loading control.

(C) *Xenopus* oocyte microinjection assays. A mixture of 1-10 fmoles in vitro-transcribed ^{32}P-labeled EBER-1, U1ΔSm, tRNA^{phe} and pre-let-7a-1 with or without 1 pmole of unlabeled U1ΔSm were injected into the nuclei of *X. laevis* oocytes. After 2 hr incubation, 6 oocytes were manually dissected, and RNAs from the nucleus (N) and cytoplasm (C) equivalent to one oocyte were extracted and analyzed on a 8M urea-15% polyacrylamide gel. Inj: Injected material. (Right) The black bars indicate relative RNA export efficiency of U1ΔSm, tRNA^{phe} and pre-let-7a-1 with U1ΔSm competition compared to no competition (white bars). Error bars represent standard deviation from four experiments.

(D) Microinjection experiment as in (C), except that pre-miR-15b was used instead of pre-let-7a-1 and the coinjected unlabeled competitor was 1 pmole unlabeled pre-miR-15b. The right panel shows a bar graph of relative RNA export. Error bars represent standard deviations from six experiments.

(E) The majority of exported pre-miR-344 (m^{7}G-capped) remains capped in *Xenopus* oocytes. *Xenopus* oocytes were injected with pre-miR-344 as described in Figure 3C, except that all the RNAs were not ^{32}P-labeled. Cytoplasmic RNAs equivalent to one oocyte were extracted and analyzed by Northern blot. Dashed lines indicate the positions to which 5′-monophosphate pre-miR-344 (bottom) and 5′ m^{7}G-capped pre-miR-344 (top) migrate. In lanes 3 and 5, cytoplasmic RNAs equivalent to one oocyte without injecting pre-miR-344 were mixed with the pre-miR-344 markers.
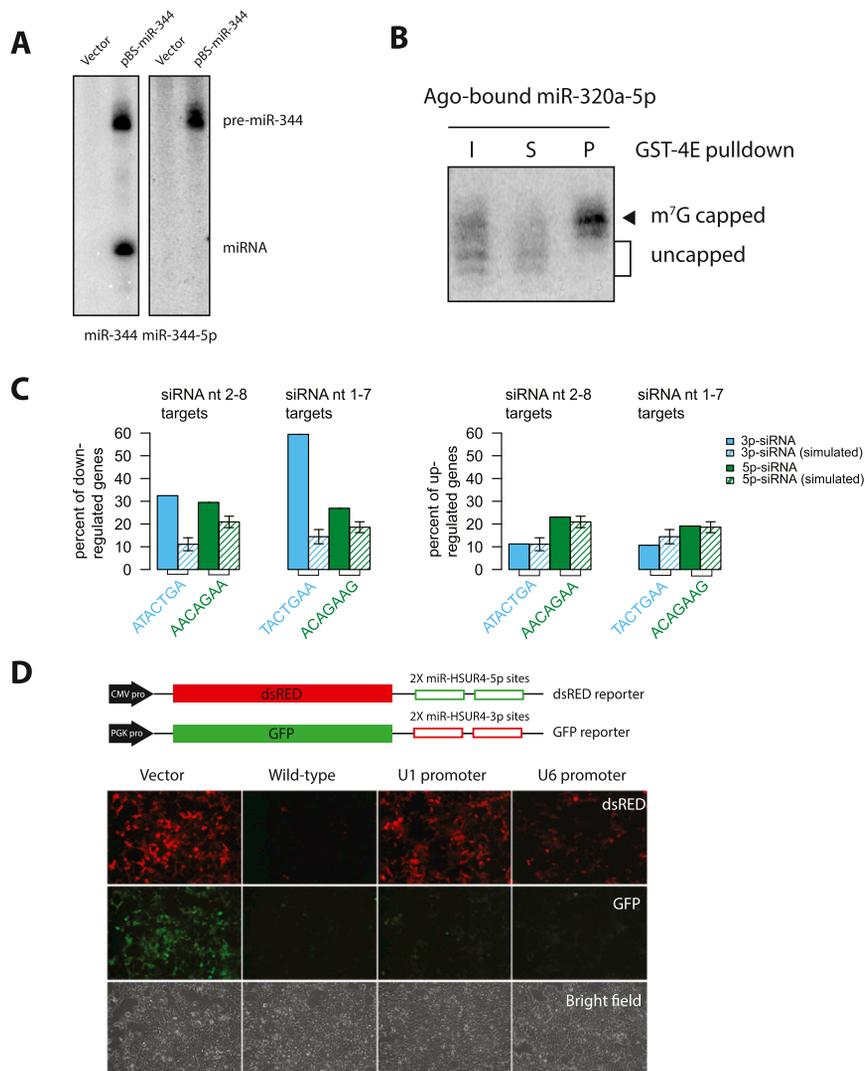
Figure S4. Additional Evidence that m^{7}G-Capped Pre-miRNAs Generate a Single Mature miRNA from the 3p Arm In Vivo, Related to Figure 4

(A) Northern blot analyses of miR-344 or miR-344-5p in total RNA extracted from 293T cells transfected with pBS-miR-344 for 48 hr.

(B) GST-4E affinity purification of Ago-bound miR-320a-5p from Figure 4C, lane 8, showing that the lower bands represent uncapped miR-320a-5p. I: input; S: supernatant; P: pellet; both I and S were 10% relative to P.

(C) Off-target effects caused by 5p-siRNA expressed from a U6-promoter-driven shRNA. The y-axes show the percentage of genes whose 3′ UTRs contain sequences complementary to the 3p- or 5p- siRNA seed region (nt 1-7 and 2-8). The simulated (striped) bars represent the probable appearance of complementary seed region sequences in randomly selected 3′ UTRs. Error bars represent standard error of the mean (SEM) obtained from 1000 simulations (see Extended Experimental Procedures).

(D) Images of 293T cells cotransfected with the pGFP-miR-HSUR4-3p and pdsRED-miR-HSUR4-5p reporters (diagramed) and one of the miR-HSUR4 expression vectors shown in Figure 4D for 48 hr.
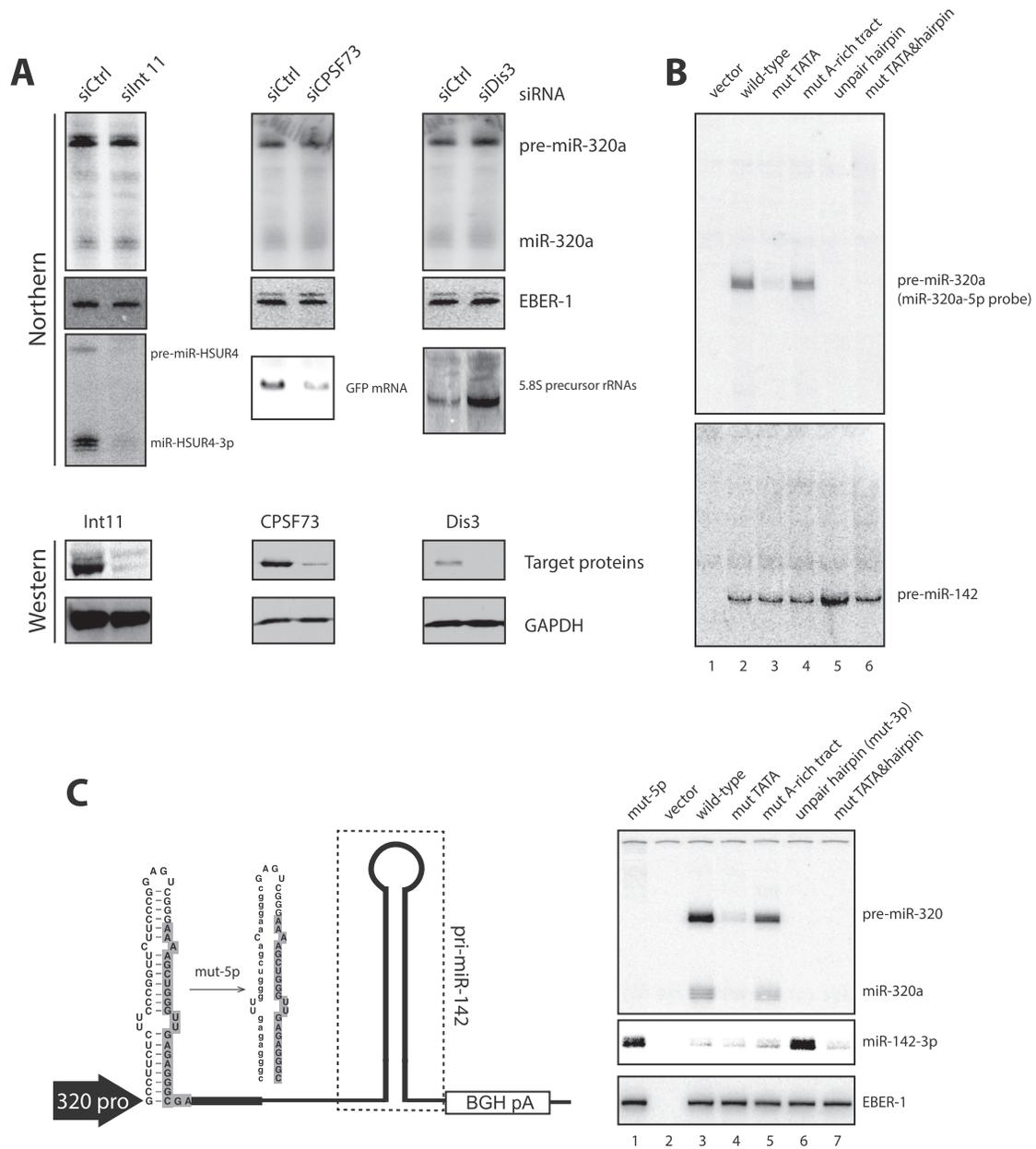
A

siCtrl siInt 11   siCtrl siCPSF73   siCtrl siDis3   siRNA

Northern

pre-miR-320a

miR-320a

EBER-1

pre-miR-HSUR4

GFP mRNA

5.8S precursor rRNAs

miR-HSUR4-3p

Western

Int11   CPSF73   Dis3

Target proteins

GAPDH

B

vector wild-type mut TATA mut A-rich tract unpair hairpin mut TATA&hairpin

pre-miR-320a
(miR-320a-5p probe)

pre-miR-142

1 2 3 4 5 6

C

mut-5p

pri-miR-142

320 pro

BGH pA

mut-5p vector wild-type mut TATA mut A-rich tract unpair hairpin (mut-3p) mut TATA&hairpin

pre-miR-320

miR-320a

miR-142-3p

EBER-1

1 2 3 4 5 6 7

Figure S5. Additional Evidence that Pre-miR-320a 3′ End Formation Is Linked to Transcription Termination, Related to Figure 5

(A) Northern blot analysis of miR-320a, EBER-1, miR-HSUR4-3p, GFP mRNA and 5.8S precursor rRNAs from 293T cells pretreated with a control siRNA (siCtrl) or an siRNA against Integrator 11 (siInt11), cleavage and polyadenylation specificity factor 73 (siCPSF73; Zhu et al., 2009) or nuclear exosome catalytic subunit Dis3 (siDis3; Tomecki et al., 2010). After 48 hr, cells were cotransfected with three plasmids expressing miR-320a, EBER-1 and miR-HSUR4 (for siInt11- and siDis3-treated cells) or pGFP-miR-HSUR4-5p (for siCPSF73-treated cells). miR-HSUR4 biogenesis is Integrator-dependent, providing a positive control for siInt11 (Cazalla et al., 2011). 5.8S precursor rRNAs accumulate upon depletion of Dis3 (Preti et al., 2013). EBER-1 serves as a transfection and loading control. The western blots show the knockdown of Int11, CPSF73 and Dis3, with GAPDH as the loading control.

(B) Northern blot analysis of the total RNAs in Figure 5B, using a miR-320a-5p or miR-142-3p probe. The pre-miR-142 panel shows a blot with adjusted contrast compared to the middle panel of miR-142-3p in Figure 5B.

(C) Left panel, schematic of the miR-320a/miR-142 transcription reporter as in Figure 5A. The pre-miR-320a hairpin was unpaired with miR-320a-5p mutated (shown in lowercase letters). Right panel, Northern blot analysis shows expression of miR-320a, miR-142-3p and EBER-1 in 293T cells cotransfected with a wild-type or mutant transcription reporter and pEBER-1.
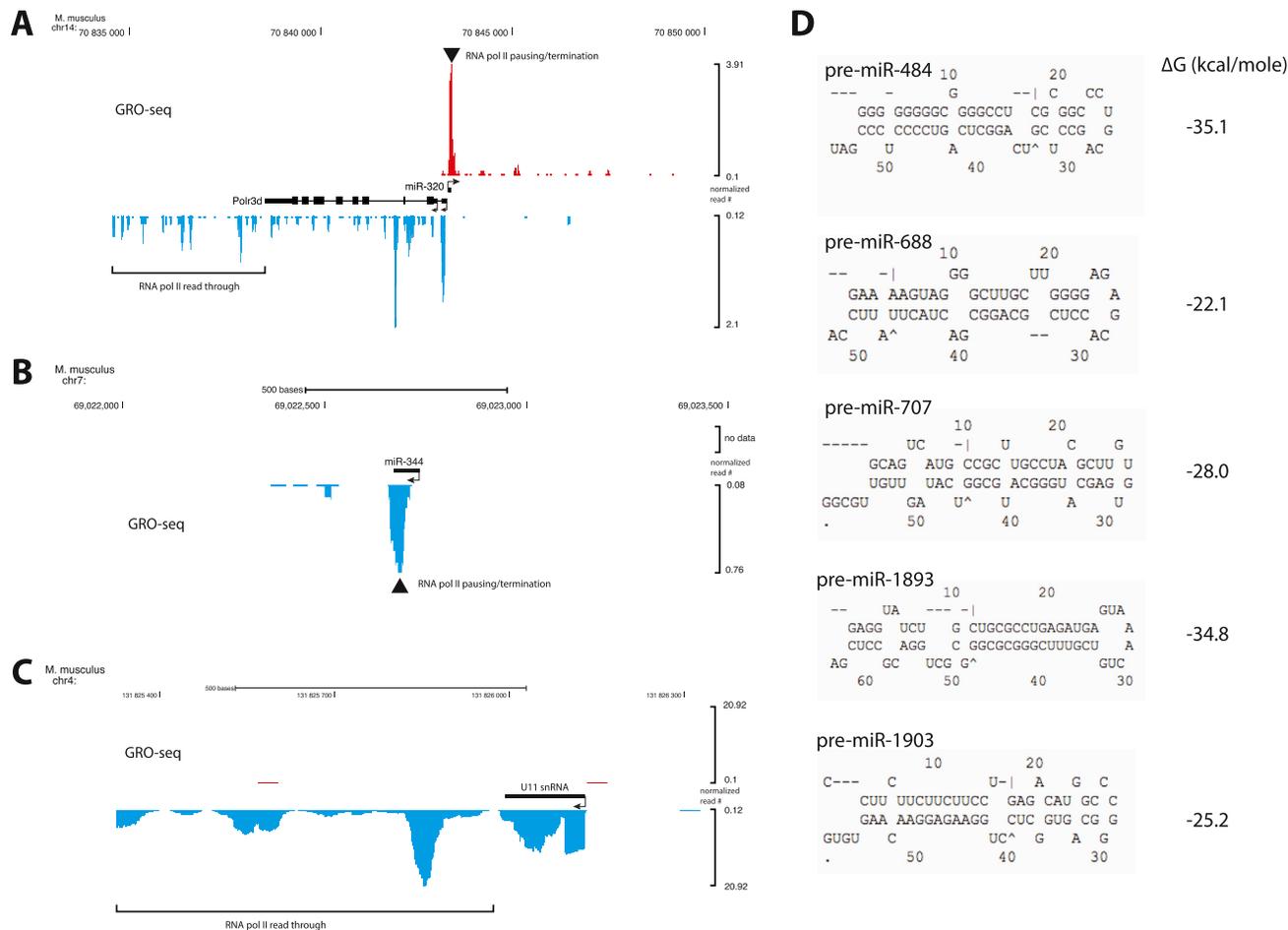
Figure S6. Additional Evidence that m⁷G-Capped Pre-miRNA 3′ End Formation Is Linked to Transcription Termination, Related to Figure 5

(A–C) Histogram showing global run-on sequencing (GRO-Seq) reads (Min et al., 2011) mapping to the mouse loci for the divergent miR-320 and Polr3d (A), the miR-344 (B) and the U11 snRNA genes (C) loci. Exons and introns are illustrated as black bars and lines, respectively. Transcription start sites and transcription directionality are indicated by black arrows. Reads mapping to the positive and negative strands are red and blue, respectively. Normalized read numbers are plotted. Arrowheads indicate proximal RNA Pol II pausing/termination at the miR-320 and miR-344 loci. The bracketed regions on the Polr3d and U11 snRNA loci indicate reads showing that RNA Pol II activity can be detected downstream of the 3′ end processing signals.

(D) The predicted pre-miRNA hairpins of miR-484, miR-688, miR-707, miR-1893 and miR-1903 are shown. Free energy (ΔG) values were calculated using the RNA folding form on the mfold server (Zuker, 2003).