HOKKAIDO UNIVERSITY

Title	pre-miRNA profiles obtained through application of locked nucleic acids and deep sequencing reveals complex 5'/3' arm variation including concomitant cleavage and polyuridylation patterns
Author(s)	Burroughs, A. M.; Kawano, Mitsuoki; Ando, Y.; Daub, C. O.; Hayashizaki, Y.
Citation	Nucleic Acids Research, 40(4), 1424-1437 https://doi.org/10.1093/nar/gkr903
Issue Date	2012-02
Doc URL	http://hdl.handle.net/2115/48581
Type	article
File Information	40(4)_1424-1437.pdf

Instructions for use

pre-miRNA profiles obtained through application of locked nucleic acids and deep sequencing reveals complex 5′/3′ arm variation including concomitant cleavage and polyuridylation patterns

A. Maxwell Burroughs*, Mitsuoki Kawano*, Yoshinari Ando, Carsten O. Daub and Yoshihide Hayashizaki

Omics Science Center (OSC), RIKEN Yokohama Institute, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama-shi, Kanagawa, 230-0045, Japan

ABSTRACT

Recent research hints at an underappreciated complexity in pre-miRNA processing and regulation. Global profiling of pre-miRNA and its potential to increase understanding of the pre-miRNA landscape is impeded by overlap with highly expressed classes of other non coding (nc) RNA. Here, we present a data set excluding these RNA before sequencing through locked nucleic acids (LNA), greatly increasing pre-miRNA sequence counts with no discernable effect on pre-miRNA or mature miRNA sequencing. Analysis of profiles generated in total, nuclear and cytoplasmic cell fractions reveals that pre-miRNAs are subject to a wide range of regulatory processes involving loci-specific 3′- and 5′-end variation entailing complex cleavage patterns with co-occurring polyuridylation. Additionally, examination of nuclear-enriched flanking sequences of pre-miRNA, particularly those derived from polycistronic miRNA transcripts, provides insight into miRNA and miRNA-offset (moRNA) production, specifically identifying novel classes of RNA potentially functioning as moRNA precursors. Our findings point to particularly intricate regulation of the let-7 family in many ways reminiscent of DICER1-independent, pre-mir-451-like processing, introduce novel and unify known forms of pre-miRNA regulation and processing, and shed new light on overlooked products of miRNA processing pathways.

INTRODUCTION

Micro RNAs (miRNAs), 20–23 nt short RNAs regulating stability and translational efficiency of transcribed mRNAs through complementary binding of target mRNA transcripts (1), are produced via transcription from the genome as primary (pri-miRNA) transcripts encoding either single or multiple (polycistronic) miRNA precursor hairpin-like regions, excision of hairpin regions, termed precursor miRNAs (pre-miRNAs) via the microprocessor complex containing the RNaseIII enzyme DROSHA (2,3), transport of pre-miRNAs into the cytoplasm via 3′-overhang recognition (4,5), duplex generation via hairpin loop removal by the RNaseIII enzyme DICER1 (6), and selection of a single strand of the duplex (the 'mature' strand) for association with a member of the AGO family (7,8). AGO-miRNA association forms functional RNA-induced silencing complexes (RISC) which bind to and regulate mRNA transcripts.

New research is revealing diverse regulatory pathways influencing levels of mature miRNA (9–13), some of which act directly on pre-miRNA hairpins. LIN28A binds a conserved nucleotide sequence motif in the hairpin loop region of the pre-let-7 miRNA family (14–17) acting as a processivity factor (18) in the untemplated addition of polyuridine (poly(U)) tails to the 3′-ends of pre-miRNAs via the ZCCHC11 enzyme, a member of the TRF family in the DNA polymerase β-like superfamily of

ribonucleotidyltransferases (19,20); LIN28A binding-induced structural changes and/or poly(U)-tailing blocks DICER1 uptake (21–23). Similarly, MBNL1-binding to a distinct motif in the hairpin region of pre-mir-1 regulates miR-1 expression (24). An additional form of regulation was observed in AGO2-mediated endonucleolytic cleavage ∼9–11 nt from the 3′ pre-miRNA end (25). This cleavage is an essential step in the recently described DICER1-independent pre-mir-451 processing pathway wherein the 3′-cleaved pre-mir-451 hairpin is unwound and subject to polyuridylation. This poly(U) tail appears to act as a signal for exonucleolytic degradation which proceeds until reduction to ∼23 nt, the remaining length likely shielded from exonuclease activity by AGO2 (26–28).

With emerging research suggesting that pre-miRNAs, far from being static intermediates in the pathway to mature miRNA production, are subject to diverse forms of regulation the need to better understand the global landscape of pre-miRNA sequences has increased. However, deep profiling of pre-miRNA sequences faces a substantial obstacle: the length range overlap of pre-miRNAs with other, far more numerous classes of ncRNA, including tRNA and snoRNA. Our group has previously successfully reduced expression of deep sequencing artifacts in the form of adapter-dimers, increasing the yield of genuine RNA sequences in a given library (29). Additionally, next-generation transcriptome sequencing data displays a wide range of tag expression levels; some tags are expressed much higher relative to others (30,31). Synthesizing the former technique with the latter observation, we have developed a novel approach to increase pre-miRNA yield during small RNA library construction using locked nucleic acid (LNA)-based antisense oligonucleotides which specifically hybridize to the most abundant endogenous sequences in a library. The resulting data set presented here represents the first *in vivo*, complete full-length profiles of nuclear, cytoplasmic, and total cellular populations of human pre-miRNA. Analysis of this data set reveals that pre-miRNAs are subject to far more complex regulatory processes than previously realized and potentially links previously described aspects of pre-miRNA processing and regulation.

MATERIALS AND METHODS

Cell culture and RNA extraction

HeLa cells were purchased from RIKEN BioResource Center and cultured in DMEM (Invitrogen, Carlsbad, CA, USA) and 10% FBS in a 5% CO_2 at 37°C. Cultured cells were collected, washed twice with cold PBS and incubated in Solution A (50 mM Tris-HCl pH 7.5, 0.8 M Sucrose, 150 mM Potassium chloride, 5 mM Magnesium chloride, 6 mM β-mercaptoethanol and 0.5% NP-40) for 10 min on ice (32). Cytoplasmic extracts were cleared by centrifugation at 16 000g for 15 min at 4°C and cytoplasmic RNAs were extracted with TRIzol LS (Invitrogen) and FastPure RNA kit (Takara Bio, Ohtsu, Shiga, Japan) from the extracts.

Pellets were washed twice with Solution A, suspended with TRIzol (Invitrogen) followed by RNA extraction with FastPure RNA kit (yielding nuclear RNAs). Total RNAs were extracted with TRIzol and FastPure RNA kit as previously described (33).

Small RNA library construction and deep sequencing

Small RNA cDNA library were generated from 1.2 µg of HeLa cell RNAs (total, cytoplasmic and nuclear fraction RNAs) as previously described (29) with ∼2 µM each LNA/DNA oligonucleotide (GeneDesign, Ibaraki, Osaka, Japan) for the most highly expressed sequences in the reverse transcription reaction at 47°C. Nucleotide sequences are shown in Supplementary Table S1. Deep sequencing was performed using an Illumina GAIIx sequencer (Illumina, San Diego, CA, USA) with a maximum read length of 115 nt. Sequencing data are deposited in the DNA Data Bank of Japan (DDBJ) under accession number DRA000455.

Selection of targets for LNA/DNA oligo treatment

Targets were selected by extracting the fifty most abundantly sequenced 3′-ends in untreated libraries. The 27 3′-ends showing the highest relative rankings across the three libraries were selected for targeting (Supplementary Data set S1). LNA/DNA oligos were designed as described previously (29), using the 3′-ends of the target RNA species as the template for hybridizing the 3′-ends of the LNA/DNA oligos (Figure 1).

Analysis of libraries

Sequences from each library were processed and mapped to the human genome (hg18 assembly) using software provided by Illumina with standard settings. Artifact sequences were further filtered using the TagDust program (34) and quality of the data independently monitored with the SAMStat program (35). pre-miRNA expression counts were determined by identifying sequence genome overlap with known miRNA loci [ver.16, miRBase (36)]. miRBase definitions occasionally include degradation products of other ncRNA instead of *bona fide* miRNA sequences (37), we detected several such sequences which were likely snoRNA degradation products based on distributions across compartments and size fractions (e.g. large numbers of long reads but few or no reads in short fractions). These sequences (pre-mir-3607, pre-mir-3651, pre-mir-3647 and pre-mir-3653) were manually culled from our list, ensuring the calculated numbers were as conservative as possible. Given the large number of sequences that appeared to be polyuridylated in the data and the potential for such extended tails to prevent proper mapping, we filtered the raw data, removing extended regions at the 3′-end which appeared to harbor poly(U) tails. Tags with removed poly(U) tails were re-mapped to the genome and then checked for concordance with known miRNA positions on the genome. While this procedure extracted quite a few tags which then mapped to the genome, a total of two tags mapped to miRNA loci, indicating that this filtering had little effect on pre-miRNA sequence counts.
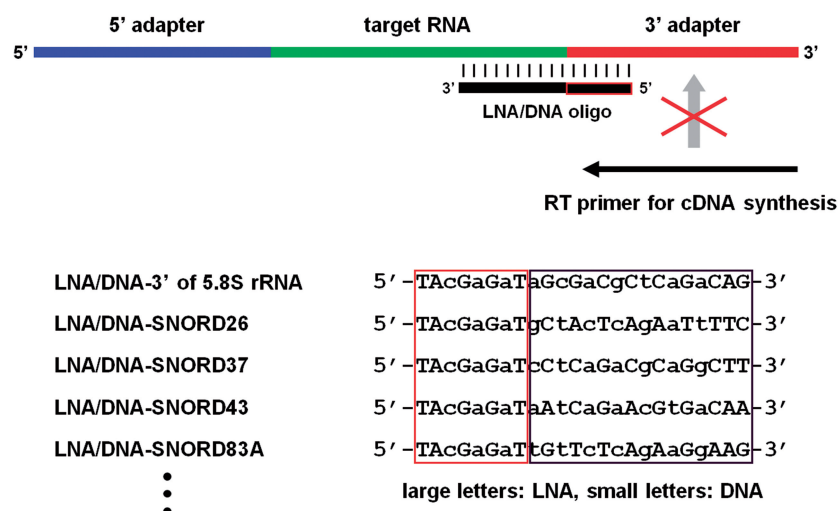
Figure 1. Overview of LNA/DNA targeting technique. LNA/DNA oligos are added to the RNA library preparation prior to the cDNA synthesis step and are designed to interact directly with RNA targets, inhibiting the reverse transcription (RT) reaction as they cannot be used as RT primers. Examples of some designed LNA/DNAs are provided at the bottom of the figure; the complete list is available in Supplementary Data set S1.

Poly(U) tails were identified through successive removal of 3′ nt of all sequenced reads not exactly matching the genome. When such a truncated read exactly matched the genome, the removed nucleotides were tested under the following criteria to determine if the read contained a 3′ poly(U) tail: (i) total nucleotide length of the poly(U) tail was ≥3 and (ii) 80% of the nucleotides within the poly(U) tail-like regions were uridine. The second requirement is necessary as a hedge against the known propensity of the Illumina sequencers to introduce sequencing errors at the 3′-ends of long reads (38). The above process ensures that all identified poly(U) tails are (i) located at the 3′-end of the read and (ii) outside of the tail region, match exactly to the genome. These poly(U) tails are, therefore, most likely the result of post-transcriptional processing. In a similar manner, any read identified as having 3′ or 5′ cleavage sites were required to exactly match the genome. Cleavage sites were required to be positioned five or more nucleotides internal to 5′ or 3′-ends to guard against inclusion of differential cutting mediated by DROSHA.

Primary pre-isomirs, the isomirs with the largest number of sequence reads at individual miRNA loci, were identified as previously described for mature isomiRs (33). Pair probability calculations were also calculated as previously described for duplex structures (33,39) substituting the RNAfold program for the RNAcofold program to reflect the differences in calculating pairing probability for a single hairpin strand versus two strands forming a duplex and with the average value taken across all pair probability values calculated for the first 5, 10, 15, 20 or all nucleotides in the hairpin. Comparisons were made across the set of all pre-miRNA loci with 3′ (or 5′) cleavage events and those lacking any evidence of cleavage. pre-miRNA lengths underlying comparisons across fractions and miRBase hairpin definitions were calculated by weighting all pre-isomir lengths observed at individual loci according to their expression, yielding a single length for all loci. Lengths were compared to the set of miRBase hairpin lengths for which at least one tag was observed in either the total, cytoplasmic or nuclear fraction libraries. A polycistronic miRNA locus was defined as a locus with <200 nt between itself and a neighboring miRNA locus. To identify ppiRNA and fpRNA, we constructed sets of genome coordinates bridging the distance between the polycistronic miRNA loci identified above and extending 200 nt upstream and downstream of all non-polycistronic miRNA loci, respectively. Large-scale genome analyses were carried out using bedtools and samtools software packages (40,41). Statistical analyses were performed using the R language and environment for statistical computing.

RESULTS

Enrichment in pre-miRNA sequences following LNA/DNA treatment

Deep sequencing of HeLa cells in the control condition provided targets for LNA/DNA treatment (see 'Materials and Methods', Figure 1 and Supplementary Data set S1). ncRNA classes are visibly affected following LNA/DNA treatment (Supplementary Table S1 and Supplementary Figure S1). Moreover, individual species of RNA targeted by LNA/DNA treatment are efficiently reduced (Supplementary Figure S2). Comparison of pre-miRNA sequences in the LNA(−) versus LNA(+) libraries revealed a marked increase from a few hundred tags in each library to as many as 20 000 tags in the cytoplasmic LNA(+) library (Figure 2A and B). In addition to increasing total sequence counts, the number of pre-miRNA loci covered also roughly doubled (Figure 2B).
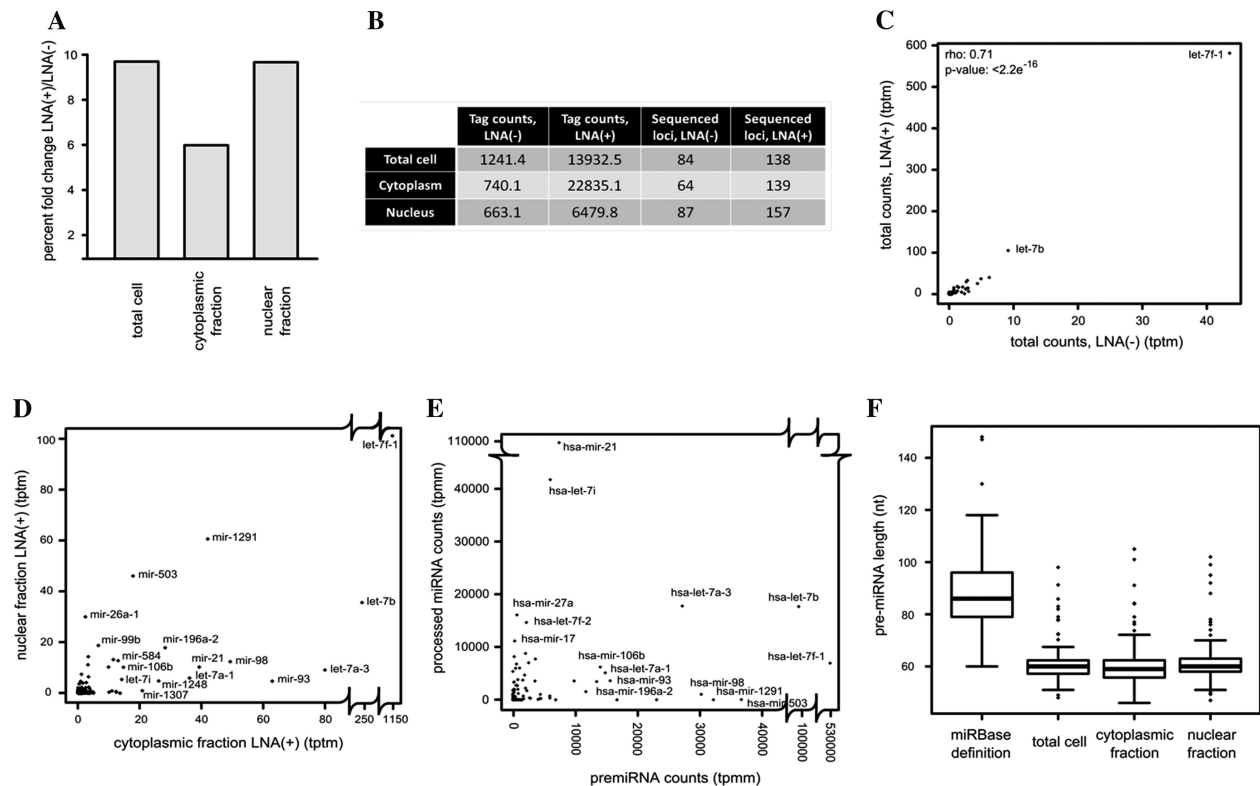
Figure 2. General features of LNA(+) libraries. (**A**) Relative enrichment of pre-miRNA sequences in LNA(+) versus corresponding LNA(−) library as a percentage of the total number of sequences within a library. (**B**) Table showing raw number of pre-miRNA sequences and the number of pre-miRNA loci with at least one sequence identified in each library. (**C**) Comparison of pre-miRNA sequence expression normalized to tags per ten million (tptm) across LNA(−) and LNA(+) conditions in the total cell fraction, see Supplementary Figure S3 for comparisons across cytoplasmic, nuclear and mature fractions. (**D**) Comparison of pre-miRNA sequence expression (tptm) across nuclear and cytoplasmic fractions. (**E**) Comparison of LNA(+) pre-miRNA expression in the total fraction against publicly available total short-read miRNA expression in HeLa cells (42) (summing mature miRNA and miRNA* sequences within individual loci) normalized to the tags per million miRNA within a library (tpmm). (**F**) Comparison of length distributions across different libraries alongside miRBase reference lengths (36) ('Materials and Methods' section).

LNA/DNA treatment does not affect miRNA expression

Possible effects of LNA/DNA treatment on pre-miRNA and mature miRNA were examined by comparing LNA(+) and LNA(−) sequence counts. We observed high correlations across each fraction (rho = 0.64–0.71) with increases in the LNA(+) condition (Figure 2C and Supplementary Figure S3A). We also observed high correlation across mature libraries (rho = 0.90), suggesting that LNA/DNA treatment does not affect mature miRNA sequencing (Supplementary Figure S3B).

Little correlation was observed across cytoplasmic and nuclear compartments. Notably, LNA/DNA treatment increases the dynamic range across the two compartments, clarifying relationships between locus counts in nuclear and cytoplasmic compartments (Figure 2D, Supplementary Figure S4). Little correlation was also observed when comparing total cellular pre-miRNA loci counts to total mature miRNA counts (42) (Figure 2E), likely related to some combination of the following three factors: (i) the influence of different regulatory pathways on pre-miRNA processing (see below), (ii) misannotation of sequences as pre-miRNA hairpins in miRBase (37) and (iii) differences in the relative sequencing depth between the two populations. Sequence counts for pre-miRNA

and mature miRNA loci are provided in Supplementary Tables S2–S4, alignments of all sequences to pre-miRNA loci in Supplementary Data set S2.

General features of sequenced pre-miRNA

While the composition and lengths of mature miRNA are well-characterized through deep sequencing and more targeted approaches, the precise genome boundaries of pre-miRNA precursors are difficult to unequivocally establish, particularly in cases where mature expression from one arm of the pre-miRNA is low (43). Our approach enables precise definitions for such transcripts (Supplementary Data set S2). On a global scale, lengths of sequenced pre-miRNAs were collected and compared across compartments and with corresponding miRBase hairpin definitions (Figure 2F; 'Materials and Methods' section). Our data indicate that dispersion of pre-miRNA lengths is tightly clustered around 60 nt, with no significant length differences observed across compartments. miRBase definitions were longer in median length (86 nt) and considerably more widely dispersed (Figure 2F). To elucidate discrepancies between miRBase and sequenced pre-miRNAs lengths, we mapped positions of sequenced pre-miRNAs relative to miRBase start and

end positions. The majority of sequence 5′ start sites in our data were within 5 nt of the miRBase-defined start site while 3′-ends display an unusual gradual increase with the majority of reads located within 10–15 nt of the miRBase end site (Supplementary Figure S5).

Mature animal miRNAs display substantial 5′/3′-end heterogeneity as revealed by deep sequencing (31,43,44) and detailed biochemical probing of pre-miRNA structures (45,46), yielding multiple distinct sequence 'isomiRs' from a single locus. While the DICER1 enzyme contributes substantially to this heterogeneity through differential cleavage, DROSHA also plays a role in enhancing heterogeneity at the 3′ terminus of mature miRNAs derived from the 3′ arm of the hairpin structure (31,43–45,47). We calculated positional variation at the 5′ and 3′-ends of all unique sequenced pre-miRNAs from the total cell fraction using the most frequently sequenced pre-isomir (hereafter the 'primary pre-isomir') as a reference and compared this with mature miRNA

variation (42) (Figure 3A). Similar heterogeneity is observed in pre-miRNA and mature miRNAs when considering all unique isomiRs; however, when comparing only isomiRs mapping exactly to the genome (thereby removing effects of nucleotidyltransferase-mediated 3′ addition events (33,48,49), heterogeneity in pre-miRNA sequences decreases (Figure 3A). The same trend is evident in the set of all sequenced tags (Supplementary Figure S6), suggesting modifications following DROSHA-mediated cleavage contribute to pre-miRNA end heterogeneity. Consistent with this, a distinctive 'tail' is observed in the region downstream of the 3′-end of pre-isomirs (Figure 3A). Comparison of end heterogeneity across nuclear and cytoplasmic compartments revealed slightly greater heterogeneity at 3′-ends stemming from post-cleavage modifications likely occurring in the cytoplasmic fraction (Supplementary Figure S7). Heterogeneity at the 3′-ends of unique mature miRNA sequences derived from only the 3′-pre-miRNA arm
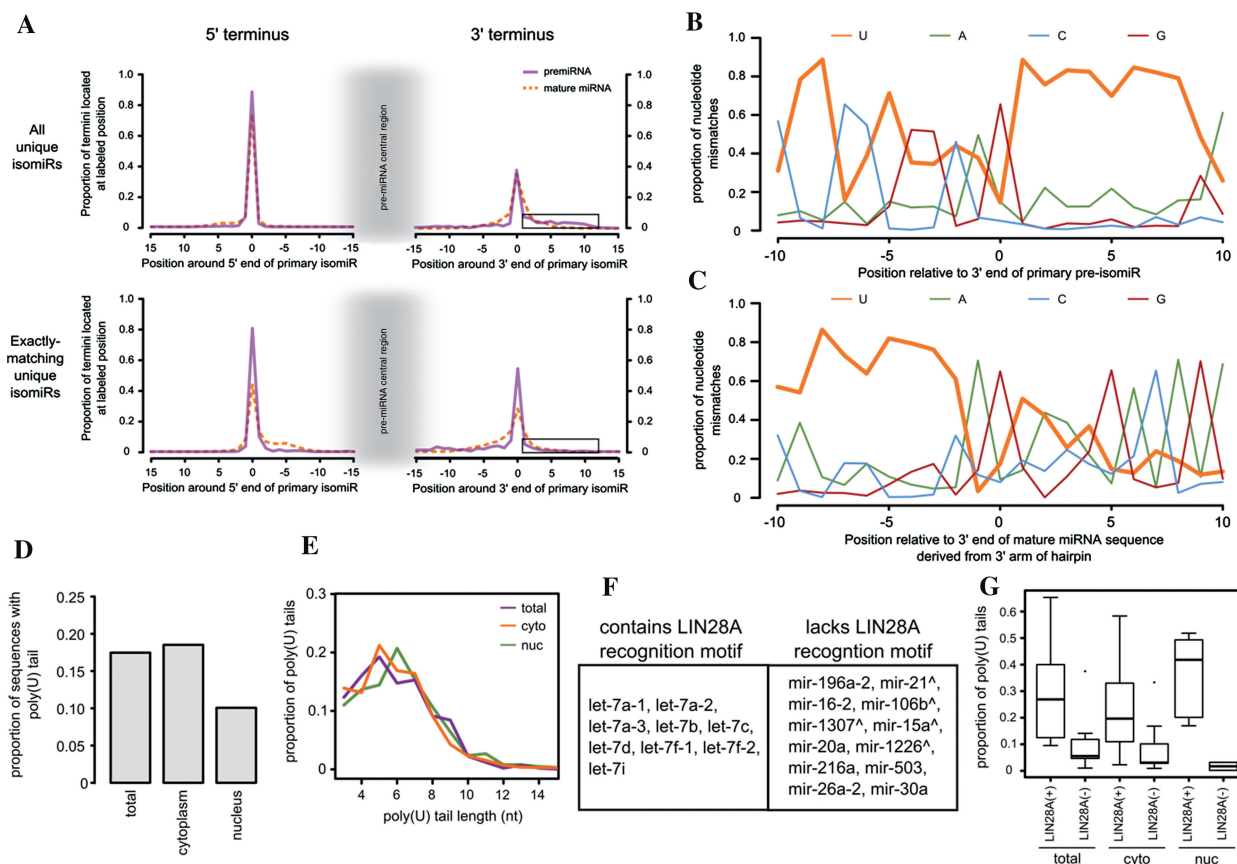

Figure 3. Analysis of pre-miRNA sequence features. (**A**) Analysis of heterogeneity at the 5′ (left side) and 3′ (right side) ends of pre-miRNA relative to mature miRNA considering unique sequences in total cellular fractions. Proportions of sequences in a given library are plotted against the location of their 5′ and 3′ ends relative to the primary pre-isomir (pre-miRNA) or primary isomiR (mature miRNA) normalized to the zero point in all line charts. Negative numbers refer to positions internal to the pre-miRNA hairpin. The top plot shows proportions for all unique sequences in the libraries, the bottom charts show proportions for all exactly mapping unique sequences in the libraries. A black box highlights the extended region of 3′ end variation resulting from poly(U)-tailing when examining all sequences (top right) and the lack of this feature when examining only exactly matching sequences (bottom right). (**B** and **C**) Plotting the proportion of nucleotide mismatches in pre-miRNA sequences from the total cellular fraction at labeled positions around a zero point normalized to (B) the 3′ end of the primary pre-isomir and (C) the miRBase-defined 3′ end point of the mature or miRNA* sequence derived from the 3′ arm of the pre-miRNA hairpin. (**D** and **E**) Proportion of sequences with poly(U) tails (D) and poly(U) tail length distributions (E) in each cellular fraction. (**F**) List of loci with identified poly(U) tails, divided into loci with the LIN28A recognition motif in the experimentally determined relevant location of the pre-miRNA hairpin and those lacking such a motif. '^' denotes loci containing poly(U) tails at the miRBase-defined 3′ end. (**G**) Proportion of poly(U) tails occurring across the two sets of loci defined in (F) in each cellular fraction.

cannot be entirely explained by DROSHA cutting, suggesting contribution from unidentified nucleases (Supplementary Figure S8).

Over-representation of uridine mismatches

To investigate the source of heterogeneity in Figure 3A, we systematically tallied nucleotide mismatches to the genome for pre-miRNA sequences at positions surrounding the 3′ terminus of the primary pre-isomir. A clear over-representation of uridine mismatches extends roughly eight basepairs downstream of the 3′ terminus of the primary pre-isomir (Figure 3B), suggesting the presence of 3′ terminal poly(U)-tailing. However, when setting the 3′ terminus for each pre-miRNA locus as the last base in miRBase-defined 3′ hairpin-derived miRNA (mature or miRNA*) sequence, we surprisingly observed a dramatic shift in uridine mismatches to positions upstream of the 3′-end (Figure 3C), suggesting that poly(U)-tailing events detected in the libraries are primarily internal relative to canonical, miRBase-defined pre-miRNA hairpin structures and the primary pre-isomir of at least some loci have truncated 3′ arms.

Characterization of widespread poly(U)-tailing

Further analysis of poly(U)-tailing determined 18–20% of all tags in total and cytoplasmic fractions and 10% in the nuclear fraction harbored poly(U) tails (Figure 3D; 'Materials and Methods' section). Length distributions of poly(U) tails were similar across all three fractions, centered on 5–7 nt (Figure 3E); these lengths are more consistent with poly(U)-tailing observed following AGO2-cleavage in pre-mir-451 (26–28) than with the ~14 nt poly(U) tails found at 3′ termini of let-7 family pre-miRNAs (14). Polyuridylation affects two groups of loci: (i) LIN28A-binding motif-containing pre-miRNAs (see above) and (ii) LIN28A binding motif lacking pre-miRNAs (Figure 3F, Supplementary Table S5). The relative concentration of poly(U) tails in loci belonging to the former group is significantly higher than loci belonging to the latter group (Figure 3G) indicating poly(U)-tailing is concentrated in LIN28A-binding motif-containing pre-miRNAs.

Poly(U) tailing and 3′ cleavage

With the bulk of poly(U) tails originating from points internal to the 3′ hairpin terminus, we postulated poly(U) tails may be related to 3′-arm nuclease activity. An analysis of all pre-miRNA tags revealed exceptionally high rates of probable 3′ nuclease-mediated cleavage: 44% in the total fraction, 46% in the cytoplasmic fraction and 19% in the nuclear fraction (Figure 4A). We examined the relationship between 3′ cleavage events and poly(U)-tailing in three sets of pre-miRNAs: LIN28A-binding motif-containing, pre-miRNAs with poly(U)-tailing but lacking canonical LIN28A recognition motifs (see below), and pre-miRNA with 3′ cleavage events lacking poly(U)-tailing (see Supplementary Discussion).

LIN28A binding motif-containing pre-miRNA. LIN28A associates with a 'GGAG' sequence motif positionally restricted to the 3′-end of the hairpin loop structure of pre-miRNAs (18,21). This motif is conserved across the let-7 family and is found in several other pre-miRNAs (18). Of the known pre-miRNAs harboring this motif, only let-7 family members were expressed in our libraries; let-7 family 3′-end positions were plotted relative to unique origin sites of poly(U) tails in all three cellular fractions (total cell in Figure 4B, nuclear, cytoplasmic fractions in Supplementary Figure S9). In addition to the expected peak at the 3′ terminus, a striking periodicity is observed in 3′-end positions with peaks centered at −10, −20 and −30 nt positions internal to the pre-miRNA structure with evidence of 'tiling' between the peaks (Figure 4B, Supplementary Figure S9 and S10). Remarkably, the distribution of poly(U) tail origin sites mirrors this pattern of periodicity and tiling with one exception: poly(U) tail formation is rarely observed prior to the −10 nt position (Figure 4B, Supplementary Figure S9 and S10). The peak centered at −10 nt is consistent with *in vivo* cleavage mediated by AGO2 slicer activity [indeed, pre-let-7a is specifically targeted by *in vivo* AGO2-mediated 3′ cleavage (25)] suggesting that AGO2-like cleavage events precede internal poly(U) tail formation.

Potential sources of the second and third cleavage peaks are less clear. Intriguingly, mapping end positions directly onto let-7 family structures reveals the second peak is positioned just downstream of the LIN28A binding site (Figure 4C, Supplementary Figure S11). While LIN28A may not be highly expressed in HeLa cells, a similar RNA-binding factor could block pre-miRNA hairpins from exonuclease activity beyond the −20 nt position; disassociation of such a factor could contribute to the tiling and peaks observed around the −30 nt position. Importantly, while the −30 nt peak is clearly identified when examining unique sequences in the library, it is infrequent in the context of the complete set of library tags, indicating that these cleavage and poly(U)-tailing events are rare relative to AGO2-mediated cleavage events (Supplementary Figure S9–S10). However, comparison across nuclear and cytoplasmic compartments reveals that a striking number of poly(U) tails originate at the −20 nt position in the nuclear fraction (Supplementary Figure S10, Discussion).

pre-miRNA with poly(U)-tailing and no LIN28A-like recognition motif

Several loci lacking LIN28A recognition motifs in the hairpin loop contain 3′ cleavage and poly(U)-tailing. The frequency of these events is lower relative to LIN28A-binding motif-containing pre-miRNAs (Figure 3F and G, Supplementary Table S5) but display similar periodicity (Supplementary Figure S10 and S12) suggesting that the cleavage/polyuridylation patterns are controlled by a less-efficient version of the same regulatory process. Mutations to the LIN28A recognition site weaken but do not ablate LIN28A association with let-7 family members (18) and LIN28A binds to and regulates expression of pre-mir-1, which lacks the canonical binding site (24), suggesting that LIN28A or a LIN28A-like factor could contribute to the observed events.
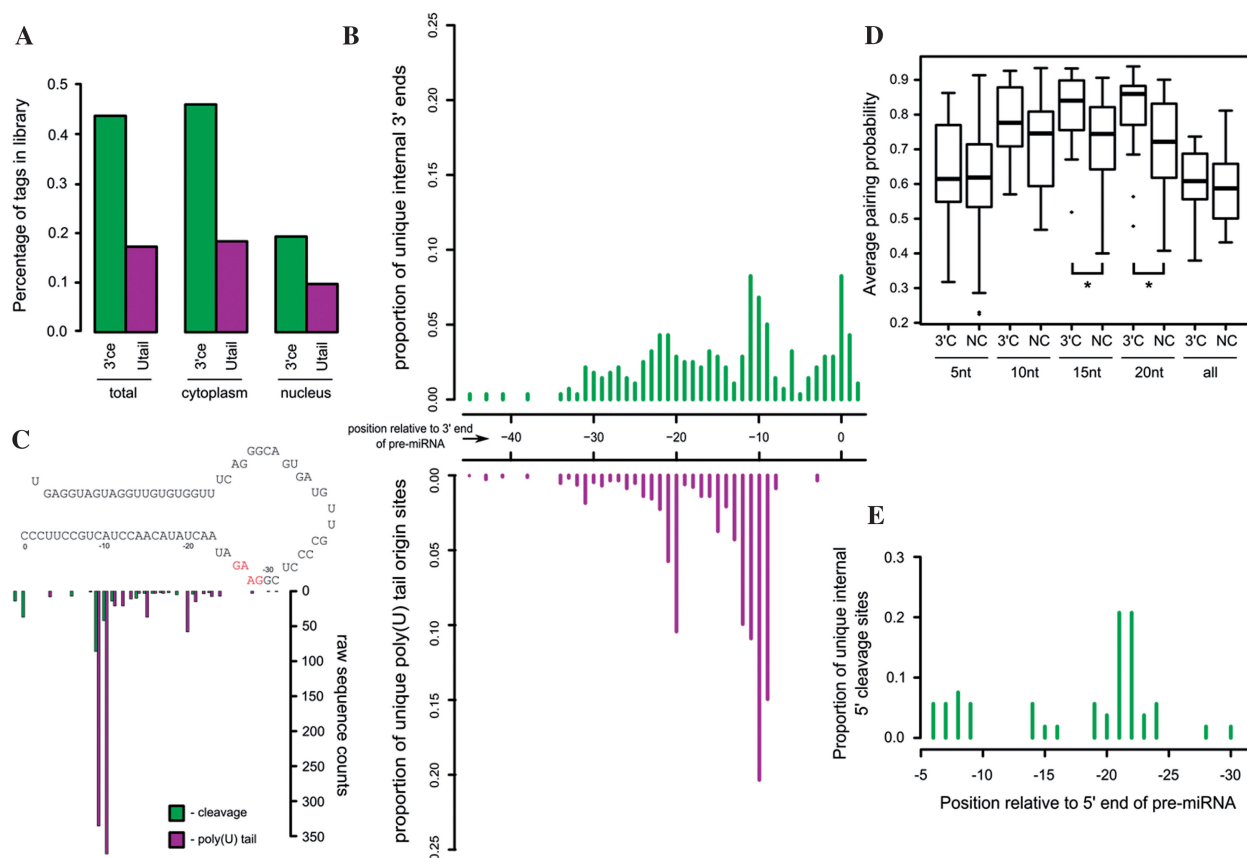
Figure 4. Concomitant pre-miRNA cleavage and polyuridylation. (**A**) Frequency and enrichment of sequences with 3′ nuclease activity, compared with polyuridylation events. (**B**) Histogram plotting the proportions of 3′ ends of unique sequences and unique sites of polyuridylation initiation in LIN28A binding motif-containing pre-miRNA sequences, revealing concomitant periodicity at −10, −20 and −30 nt peaks. Zero point in the histogram refers to 3′ end of the mature miRNA/miRNA* sequence derived from the 3′ pre-miRNA arm defined by miRBase. Negative values refer to points internal to the pre-miRNA hairpin. See Supplementary Figures S9–10, S12 for comparisons involving set of all sequences and across nuclear and cytoplasmic fractions. (**C**) Predicted hairpin structure of pre-let-7b with barplot representing the total number of raw counts in the total cellular fraction with 3′ cleavage events (green) and polyuridylation initiation sites (purple). LIN28A recognition site is colored in red (see also Supplementary Figure S11). (**D**) Boxplot comparing average nucleotide pairing probability for set of pre-miRNAs with 3′ cleavage events ('3′C') against those lacking cleavage events ('NC'). While little difference is observed in the average pairing probabilities for the first five nucleotides when counting from the 5′ base of the stem, as more nucleotides are included in the calculations the differences are significant (at 15 nt, Wilcox rank sum test, $P = 0.0048$; at 20 nt, $P = 0.0024$). (**E**) Histogram depicting proportion of 5′ cleavage events at given locations across all pooled libraries, revealing clear peak in the 20–23 nt range (see also Supplementary Figure S14 and Table S6).

AGO2-mediated 3′ cleavage events have been linked to base-pairing in the initial nucleotides of the pre-miRNA stem (25) and highly complementary base pairing along the stem of the hairpin miRNA structure (26–28). We failed to uncover evidence of bias in initial base-pairing but comparison to the set of loci lacking 3′ cleavage events supports a role for general complementary base-pairing along the hairpin stem (Figure 4D).

A handful of pre-miRNA sequences in this group show genuine 3′ poly(U) tails including pre-mir-21, 106b, 15a, 1307 and 1226 (Supplementary Table S5). It is possible that these tails are involved in blocking DICER1 uptake; however, given none of these hairpins contain properly positioned LIN28A or other conserved recognition motifs (21–23), poly(U) tails could instead be involved in distinct regulatory processes. For example, pre-mir-1226 is a mirtron; mirtrons are excised as introns from precursor mRNA transcripts and fold into hairpin structures independent of DROSHA processing

(50–52). pre-mir-1226 is predicted to fold into a hairpin structure with a rare 5′ overhang; in this case 3′ poly(U)-tailing could provide the 3′ overhang necessary for cytoplasmic export (Supplementary Figure S13) (4,5). An additional 3′ tail observed in pre-mir-1307 is the only identified instance of a poly(A) tail (Supplementary Figure S13).

5′ pre-miRNA cleavage

We also observe 5′ cleavage events scattered across a diverse set of pre-miRNA loci. 3′ cleavage events associated with internal polyuridylation almost exclusively occur in pre-miRNA loci giving rise to mature miRNA from either the 5′ arm or both arms; 5′ cleavage events occur in pre-miRNA giving rise to mature miRNA from both 5′ and 3′ hairpin arms (Supplementary Tables S5 and S6). Mapping the distribution of 5′-ends of all unique tags pooled from all three fractions reveals a distinctive peak in

the range of −20–23 nt internal to the pre-miRNA hairpin, but no peak in the −10 nt region (Figure 4E, Supplementary Figure S14). Of the 93 (73%) 5′ cleavage events in the −20–23 nt range, 68 are derived from a single pre-miRNA locus, pre-let-7i; 23% of all tags at the pre-let-7i locus undergo 5′ cleavage (Supplementary Data set S2) matching the 5′ lengths of mature let-7i miRNA. Generation of single-nick pre-miRNA hairpin products via recombinant DICER1 has been independently observed *in vitro* by two groups (53,54); the above observations suggest *in vivo* processing at a limited number of loci. The unusual hairpin loop structure harbored by let-7i may contribute to high levels of potential single-nick DICER1 processing (Supplementary Figure S15) which in turn explains distinctive pre-let-7i expression patterns relative to other let-7 family members (Figure 2E).

pre-mir-21 also undergoes considerable 5′ cleavage in the −5 to −10 nt range (Supplementary Table S6, Supplementary Data set S2). miR-21 is derived from the 5′ arm, indicating that such cleavage could affect mature miRNA production. Inspection of deep-sequenced mature miRNA data (42) suggests that this cleavage persists in processed products, indicating that these 5′-shortened hairpins are substrates for DICER1 processing. Investigation of the functional effects and molecular basis of this 5′ hairpin cleavage, which appears specific to pre-mir-21, may increase understanding of the oncogenic role of miR-21 in various cancer types (55).

Sequences flanking pre-miRNA loci

As pre-miRNAs are typically processed from pri-miRNA precursors, investigation into regions surrounding pre-miRNA sequences in LNA(+) libraries can provide insight into miRNA biogenesis. Related to this, recent research has identified a novel class of small RNAs (18–23 nt in length) located immediately downstream and upstream of pre-miRNA hairpins, termed miRNA-offset RNA (moRNA). Accumulating evidence suggests that moRNAs are produced through directed production pathways (56,57) and are not merely byproducts of miRNA biogenesis. Two proposals have been suggested for moRNA production: (i) double-stranded cleavage of extended hairpin regions on pri-miRNA transcripts via secondary DROSHA1 processing (56) and (ii) exonucleolytic activity on precursor transcripts (58,59).

Flanking pre-miRNA (fpRNA) sequences (∼60–115 nt in length) are less abundant than total pre-miRNA sequences and are enriched in the nucleus, consistent with ratios of moRNAs relative to mature miRNAs and compartmental moRNA enrichment (57), suggesting a possible link between fpRNA and moRNA processing (Figure 5A–C, Supplementary Table S7, Figure S16A). Interestingly, while moRNAs are strongly biased for derivation from the 5′ region of pre-miRNAs (57) (Supplementary Figure S17), fpRNA display a converse bias for derivation from the 3′ region (Figure 5C) possibly reflecting increased moRNA processing efficiency in 5′-derived fpRNA. Similar to previous research, we observe no correlation between moRNA and associated

mature miRNA transcripts (60); this extends to fpRNA and associated pre-miRNA sequences (Supplementary Table S7).

moRNA sequence abundance was significantly higher in pre-miRNAs derived from polycistronic miRNA sequences (Figure 5D, 'Materials and Methods' section and Supplementary Figure S16B). Reads flanking polycistronic sequences bridge pre-miRNA loci and therefore cannot be assigned to 5′- or 3′-ends of any single pre-miRNA locus. Considered independently of fpRNA sequences, polycistronic pre-miRNA intervening RNA (ppiRNA) (∼50–115 nt in length) are also enriched in the nuclear fraction; however, unlike fpRNA, ppiRNA is substantially more abundant than associated pre-miRNA sequences (Figure 5E and 5A, Supplementary Figure S16B, Table S8). The high observed ppiRNA sequence counts possibly result from the unique structure/sequence features of the intermediates furnished by processing of polycistronic pri-miRNA transcripts including the lack of a 5′ cap structures or 3′ poly(A) tails, which would tend to be present in the fpRNA.

To gain further insight into moRNA processing, heterogeneity within individual moRNA loci was examined. moRNAs derived from 5′ and 3′ regions flanking pre-miRNAs display broad length distributions inconsistent with a DROSHA-based cleavage mechanism (Figure 5F and G). We analyzed the ends of fpRNAs corresponding to DROSHA cleavage sites, reasoning that targeted moRNA production would result in moRNA-removed fpRNA intermediates beginning 18–23 nt downstream of fpRNA start sites (Figure 5H and I). DROSHA cleavage sites display variation consistent with 3′–5′ exonuclease activity in fpRNA derived from the 5′ region of the pre-miRNA transcript (Figure 5H). Similarly, fpRNA derived from the 3′-end of pre-miRNA transcripts in the nuclear fraction appear subject to low-level 5′–3′ exonucleolytic activity (Figure 5I). fpRNA ends not affected by DROSHA were also examined; 5′-ends (presumed sites of transcription initiation) show little variation while 3′-ends appear subject to exonucleolytic activity (Supplementary Figure S18A and B). Similar analyses of ppiRNA revealed no clear signal indicative of moRNA processing (Supplementary Figure S18C and D).

Shared nuclear compartment enrichment and comparable moRNA/miRNA and fpRNA/pre-miRNA enrichment levels suggest that fpRNA sequences may act as precursors for moRNA production (Figure 5A and B). However, several observations argue that moRNA production may be more complex than previously thought: (i) the persistence of fpRNA derived from the 3′-end of the pre-miRNA (Figure 5C) suggests that the 5′ arms (from which the bulk of moRNAs are produced) could be processed independent of 3′ arms and thus independent of double-stranded cleavage mechanisms. (ii) The observed enrichment of moRNAs flanking pre-miRNA loci belonging to polycistronic transcripts (Figure 5D, Supplementary Figure 14) and the enrichment of ppiRNAs relative to their flanking pre-miRNA sequences (Figure 5E) could suggest that moRNAs from polycistronic and non-polycistronic transcripts are produced
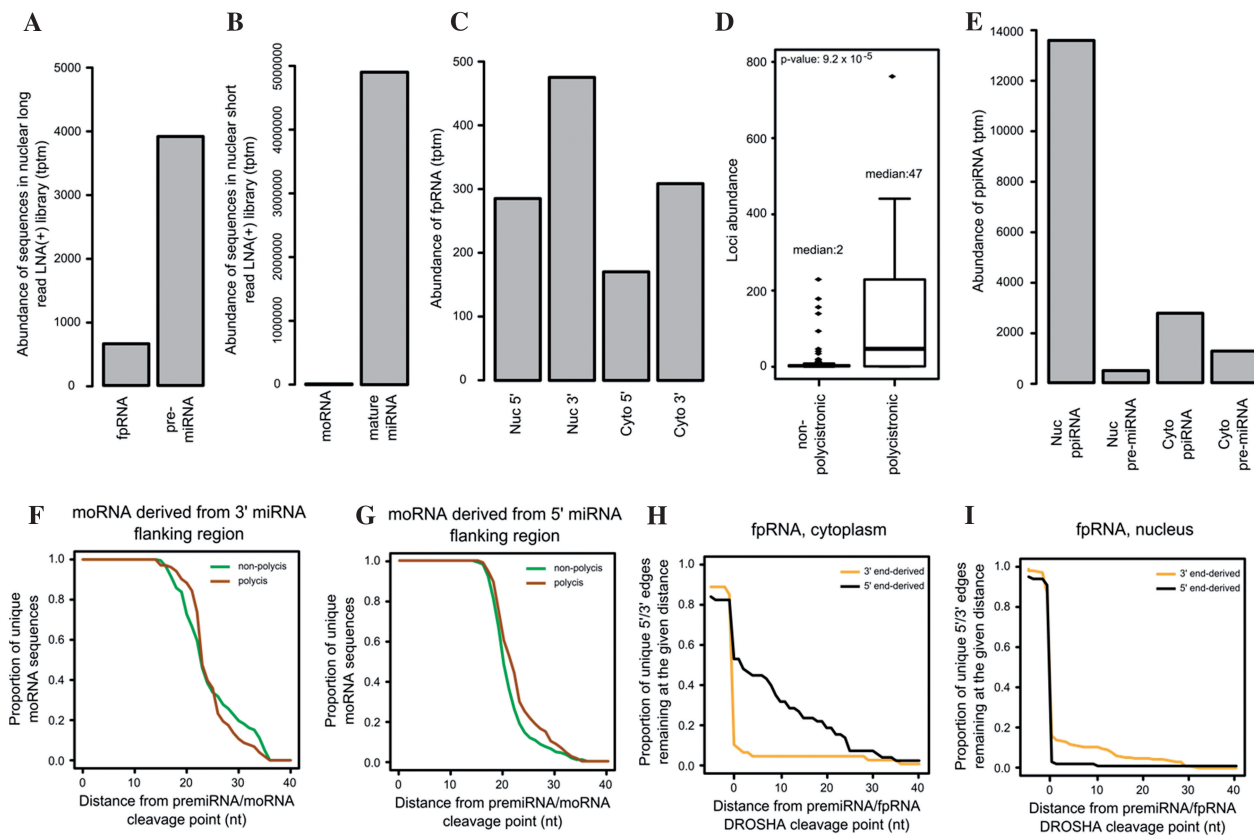
Figure 5. fpRNA and ppiRNA in relation to moRNA. (**A** and **B**) Relative abundance of fpRNA and moRNA to their cognate pre-miRNA and mature miRNA sequences. (**C**) Slight enrichment is observed in fpRNA located adjacent to 3′ arm of the pre-miRNA in both nuclear and cytoplasmic fractions, tpm normalized to facilitate comparison across the fractions. (**D**) Boxplot depicting the significant enrichment (Wilcox rank sum test, $P = 9.2e^{-5}$) of moRNA sequences derived from polycistronic pri-miRNA transcripts. (**E**) Enrichment of ppiRNA sequences in both the nuclear fraction and relative to pre-miRNA sequences derived from polycistronic pri-miRNA transcripts, tpm normalized. (**F** and **G**) Line plots depicting proportion of moRNA sequences relative to distances from the inferred DROSHA cleavage point for moRNA derived from polycistronic (brown) and non-polycistronic (green) pri-miRNA transcripts from regions adjacent to the 3′ (**F**) and 5′ (**G**) ends of pre-miRNA hairpins. Little difference is observed, suggesting similar modes of processing while the range of lengths in moRNA sequences argues against consistent cleavage points. (**H** and **I**) Line plots depicting proportions of fpRNA sequences with 3′ (orange) or 5′ (black) edges remaining at the indicated distances from the most-frequently occurring, inferred DROSHA-mediated cleavage point separating pre-miRNA from fpRNA sequences in the cytoplasmic (**H**) and nuclear (**I**) fractions. (**H** and **I**) are both indicative of possible 3′–5′ exonucleolytic processing likely unrelated to moRNA production.

through distinct pathways. (iii) The wide range of lengths observed in 'mature' moRNA sequences in this (Figure 5F and G) and other research (57) argue against a strictly measured cleavage mechanism. (iv) The puzzling lack of observed intermediates in fpRNA and particularly ppiRNA corresponding to the expected lengths of moRNA sequences and the possibility of exonuclease activity in some locations and cellular fractions (Figure 5H and I, Supplementary Figure S15). This final point, however, is hampered by lack of sequencing depth: it is possible intermediates derived from targeted endonuclease activity are escaping detection, particularly given the low abundance of moRNA transcripts relative to mature miRNA counterparts (Figure 5B). The DROSHA-mediated cleavage argument for moRNA production has been bolstered by computational meta-survey of pooled small RNA libraries in Drosophila (61) and more targeted studies in the basal chordate *Ciona* (56). Further characterization, both computational and experimental, will be required to sort out the roles of different

nucleases and possible differences between the mammalian and other animal moRNA biogenesis pathways. The striking stability of ppiRNA could indicate additional functions beyond an moRNA precursor role.

DISCUSSION

Here, we present the first complete pre-miRNA deep sequencing profiles in total, nuclear and cytoplasmic HeLa cellular fractions. These data sets form an invaluable resource for understanding global trends in pre-miRNA regulation. The analysis presented here and summarized in Figure 6 reinforces the coalescing notion that pre-miRNA hairpins are subject to a wide range of diverse and targeted regulatory processes (14–17,24,25). Notably, the data (i) connect previously described AGO2-catalyzed pre-miRNA 3′ hairpin cleavage (25) to polyuridylation and concomitant exonuclease activity, (ii) suggests that the hairpins targeted by AGO2 cleavage are intricately tied to interaction with potential
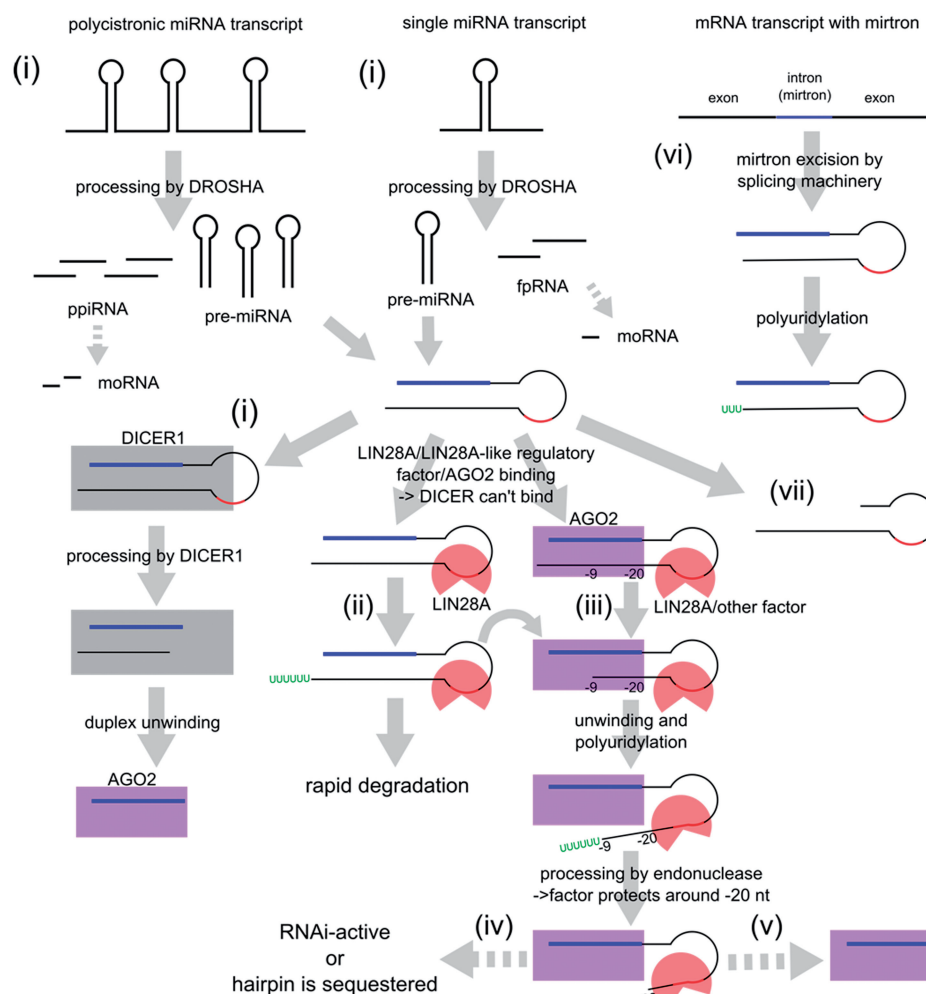
Figure 6. Overview of pre-miRNA processing. (**i**) Canonical pre-miRNA processing pathway, beginning with DROSHA processing of polycistronic or single miRNA transcripts and ending with formation of RISC complex with DICER1-processed mature miRNA. The polycistronic pre-miRNA intervening RNA (ppiRNA) and flanking pre-miRNA (fpRNA) products of DROSHA processing are further processed into miRNA offset RNA (moRNA) by undetermined mechanisms (dashed arrows). (**ii** and **iii**) Potential processing pathways involving LIN28A-like regulatory factor association: (ii) LIN28A blocks DICER1 uptake and recruits the polyuridyltransferase ZCCHC11 leading to 3′ poly(U) tail formation and pre-miRNA degradation and (iii) LIN28A or other factor association coinciding with AGO2 association leads to AGO2-mediated single-stranded hairpin cleavage followed by repeated polyuridylation and exonuclease activity restricted to the −20 nt position by factor binding (iii). While the let-7 family is most strongly affected by the process beginning at (iii), consistent with the low frequency of 3′ arm-derived let-7* sequences observed in miRBase (36), a wide range of additional loci are also affected (Supplementary Table S5). (**iv**) The end products of (iii) could either be RNAi-active, capable of binding target mRNA sequences, or could regulate pre-miRNA activity by sequestering inactivated hairpins. (**v**) On relatively rare occasions, the pre-miRNA hairpin could be processed through polyuridylation and subsequent polyuridylation activity after factor disassociation to lengths consistent with DICER1-processed mature sequences (Supplementary Figure S19 and S20). (**vi**) Mirtrons excised from mRNA transcripts via the splicing machinery may result in hairpins with 5′ overhangs (Supplementary Figure S13). 3′ polyuridylation could provide 3′ overhangs for cytoplasmic export. (**vii**) Some pre-miRNA loci are subject to 5′ cleavage, possibly processed by single-nick DICER1 activity (Supplementary Figure S15).

RNA-binding factors like LIN28A, and (iii) provides evidence linking a broad set of miRNA loci to the pre-mir-451-like, DICER1-independent processing pathway, raising the possibility that in several loci an additional pathway exists for production of the RNA component associating with AGO2 during RISC formation. Given the conservation of LIN28 in the crown group eukaryotes (62), it is possible that this pathway is active in both animal and plant miRNA processing. The fundamental evolutionary implication, however, is that RNAi could be widely active in only the presence of an AGO family member (19) and an RNA hairpin, consistent with

the observation that AGO-centric RNAi is remarkable in its ability to incorporate various RNA and DNA sources derived from distinct processing pathways as guide strands which target mRNA across all three superkingdoms of Life (63–71). However, as the specific AGO clade of the AGO superfamily is a later innovation in eukaryotes (as opposed to the more ancestral PIWI clade) (62,72), the complete phyletic distribution of such an AGO-RNA hairpin processing pathway remains unclear. In addition to providing insight into processing pathways incorporating AGO-based cleavage, the data presented in this manuscript provide insight into moRNA

production and suggests the presence of novel hairpin regulatory pathways including (i) mirtron hairpin processing through poly(U)-tailing, (ii) 5′, single-strand endonucleolytic cleavage in generation of mature miRNAs and (iii) exonucleolytic activity on the 5′-end of hairpins (Figure 6).

Reduction of LIN28A has previously been shown to increase mature let-7 family miRNA expression (21,22) and decrease substrate miRNA target expression (22) through a proposed pathway entailing let-7 family hairpin-binding, polyuridylation at the 3′ terminus, and rapid degradation (14,21). Our data suggest the presence of an additional contributing pathway wherein a factor like LIN28A, instead of binding directly to the pre-miRNA hairpin, binds to the hairpin after or in conjunction with AGO2-mediated 3′ pre-miRNA cleavage (Figure 6, Supplementary Figure S19). Reduction of the factor would facilitate exonucleolytic degradation of the hairpin similar to miR-451 down to the length range (20–23 nt) protected by AGO2 association, increasing mature let-7 expression and decreasing target expression. Intriguingly, comparison of length distributions of mature miR-451, let-7 family, and miRNA not affected by 3′ cleavage identifies a significant difference in the distributions of mature-451 and miRNA not affected by 3′ cleavage but not with the let-7 family, suggesting that small amounts of let-7 could be processed in a manner similar to mir-451 (Supplementary Figure S20; 'Materials and Methods' section). These distributions appear consistent with lengths of miRNA observed in DICER1-knockout mice models (73). Alternatively, 3′ cleavage/poly(U)-tailing may divert pre-miRNA from functional RNAi activity as a kind of hairpin 'sequestering'; a function with the same end result as a signal for degradation (Figure 6). Analyzing the effects of LIN28A and DICER1 knockdown on mature and hairpin deep-sequence profiles could discriminate between these possibilities.

Extreme 3′ terminus poly(U)-tailing was not observed in large quantities. Two possible reasons for this absence are as follows: (i) polyuridylation renders the pre-miRNA transcripts extremely unstable and/or (ii) the sequencing depth of our libraries remains too shallow to detect 3′ pre-miRNA poly(U) tailing. This second reason is particularly attractive given similar difficulties in observing single, 3′ U addition events and given that these two processes could be interrelated (33,74) (Supplementary Discussion). Additionally, it is important to note the sequences with poly(U) tails which are detected in large quantities in these libraries could be stabilized through AGO2 interaction.

In summary, this data and analysis open new avenues of research into understanding pre-miRNA regulatory processes. The lengths and sequence composition of many of the processed pre-miRNAs outlined in Figure 6 would preclude detection via traditional deep sequencing and miRNA amplification methods, which could have important ramifications for common laboratory techniques probing siRNA pathways through DICER knockdown. Of additional specific interest are potential relationships of the findings to let-7 family-mediated tumor suppression and miR-21-mediated tumorigenesis, disruption of an alternative hairpin processing pathway (Figure 6) or an introduced imbalance between this pathway and the standard processing pathway could dramatically alter let-7 and miR-21 roles in maintenance of cellular stability. Future improvements to the pre-miRNA yield of this experimental technique will assist in investigation of other areas of pre-miRNA regulation requiring deeper coverage including editing, single nucleotide addition (Supplementary Discussion), and extreme 3′ poly(U)-tailing.

ACCESSION NUMBER

DRA000455 (DDBJ).

NOTE ADDED IN PROOF

Since submitting this manuscript, Newman and colleagues (75) published a paper reporting a 3′ primer-based amplification method for pre-miRNA sequencing. While this method precludes detection of 5′ pre-miRNA variation and the surrounding sequences presented here, the reported 3′ variation largely agrees with results presented here. Two exceptions are 1) their sequencing is capable of detecting greater 3′ end poly(U) tailing; likely related to differences in sequencing depth discussed above. 2) They detect high levels of single/double U addition at several loci which we do not observe. This may point to cell-specific variation, possibly influenced by relative expression of LIN28A or functionally related factors. It also supports a mechanistic demarcation between initial uridylation and processive formation of 3′ poly(U) tails; again likely dependent on expression of processivity factors (see Supplementary Discussion).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables S1–9, Supplementary Figures S1–20, Supplementary Discussion, Supplementary Methods, Supplementary Data sets S1 and S2, and Supplementary References (19,21,25,27–28,33,36,42,53–54,75–85).

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
2. Gregory,R.I., Yan,K.P., Amuthan,G., Chendrimada,T., Doratotaj,B., Cooch,N. and Shiekhattar,R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
3. Han,J., Lee,Y., Yeom,K.H., Nam,J.W., Heo,I., Rhee,J.K., Sohn,S.Y., Cho,Y., Zhang,B.T. and Kim,V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
4. Lund,E., Guttinger,S., Calado,A., Dahlberg,J.E. and Kutay,U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95–98.
5. 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.
6. Hutvagner,G., McLachlan,J., Pasquinelli,A.E., Balint,E., Tuschl,T. and Zamore,P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838.
7. Hutvagner,G. and Simard,M.J. (2008) Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell. Biol.*, **9**, 22–32.
8. Leuschner,P.J., Ameres,S.L., Kueng,S. and Martinez,J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.*, **7**, 314–320.
9. Starega-Roslan,J., Koscianska,E., Kozlowski,P. and Krzyzosiak,W.J. (2011) The role of the precursor structure in the biogenesis of microRNA. *Cell Mol. Life Sci*, **68**, 2859–2871.
10. Winter,J., Jung,S., Keller,S., Gregory,R.I. and Diederichs,S. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.*, **11**, 228–234.
11. Kai,Z.S. and Pasquinelli,A.E. (2010) MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nat. Struct. Mol. Biol.*, **17**, 5–10.
12. Kim,Y.K., Heo,I. and Kim,V.N. (2010) Modifications of small RNAs and their associated proteins. *Cell*, **143**, 703–709.
13. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
14. Heo,I., Joo,C., Cho,J., Ha,M., Han,J. and Kim,V.N. (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell*, **32**, 276–284.
15. Newman,M.A., Thomson,J.M. and Hammond,S.M. (2008) Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA*, **14**, 1539–1549.
16. Rybak,A., Fuchs,H., Smirnova,L., Brandt,C., Pohl,E.E., Nitsch,R. and Wulczyn,F.G. (2008) A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.*, **10**, 987–993.
17. Viswanathan,S.R., Daley,G.Q. and Gregory,R.I. (2008) Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97–100.
18. Yeom,K.H., Heo,I., Lee,J., Hohng,S., Kim,V.N. and Joo,C. (2011) Single-molecule approach to immunoprecipitated protein complexes: insights into miRNA uridylation. *EMBO Rep.*, **12**, 690–696.
19. Aravind,L. and Koonin,E.V. (1999) DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res.*, **27**, 1609–1618.
20. Martin,G. and Keller,W. (2007) RNA-specific ribonucleotidyl transferases. *RNA*, **13**, 1834–1849.
21. Heo,I., Joo,C., Kim,Y.K., Ha,M., Yoon,M.J., Cho,J., Yeom,K.H., Han,J. and Kim,V.N. (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell*, **138**, 696–708.
22. Hagan,J.P., Piskounova,E. and Gregory,R.I. (2009) Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.*, **16**, 1021–1025.
23. Lehrbach,N.J., Armisen,J., Lightfoot,H.L., Murfitt,K.J., Bugaut,A., Balasubramanian,S. and Miska,E.A. (2009) LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. *Nat. Struct. Mol. Biol.*, **16**, 1016–1020.
24. Rau,F., Freyermuth,F., Fugier,C., Villemin,J.P., Fischer,M.C., Jost,B., Dembele,D., Gourdon,G., Nicole,A., Duboc,D. *et al.* (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.*, **18**, 840–845.
25. Diederichs,S. and Haber,D.A. (2007) Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell*, **131**, 1097–1108.
26. 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.
27. 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.
28. 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.
29. Kawano,M., Kawazu,C., Lizio,M., Kawaji,H., Carninci,P., Suzuki,H. and Hayashizaki,Y. (2010) Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing. *Biotechniques*, **49**, 751–755.
30. Agarwal,A., Koppstein,D., Rozowsky,J., Sboner,A., Habegger,L., Hillier,L.W., Sasidharan,R., Reinke,V., Waterston,R.H. and Gerstein,M. (2010) Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. *BMC Genomics*, **11**, 383.
31. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
32. Carninci,P., Nakamura,M., Sato,K., Hayashizaki,Y. and Brownstein,M.J. (2002) Cytoplasmic RNA extraction from fresh and frozen mammalian tissues. *Biotechniques*, **33**, 306–309.
33. Burroughs,A.M., Ando,Y., de Hoon,M.J., Tomaru,Y., Nishibu,T., Ukekawa,R., Funakoshi,T., Kurokawa,T., Suzuki,H., Hayashizaki,Y. *et al.* (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.
34. Lassmann,T., Hayashizaki,Y. and Daub,C.O. (2009) TagDust—a program to eliminate artifacts from next generation sequencing data. *Bioinformatics*, **25**, 2839–2840.
35. Lassmann,T., Hayashizaki,Y. and Daub,C.O. (2011) SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics*, **27**, 130–131.
36. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–157.
37. Hansen,T.B., Kjems,J. and Bramsen,J.B. (2011) Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis. *RNA Biol.*, **8**, 378–383.
38. Bravo,H.C. and Irizarry,R.A. (2010) Model-based quality assessment and base-calling for second-generation sequencing data. *Biometrics*, **66**, 665–674.
39. Ghildiyal,M., Xu,J., Seitz,H., Weng,Z. and Zamore,P.D. (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA*, **16**, 43–56.
40. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
41. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
42. Mayr,C. and Bartel,D.P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**, 673–684.

43. Warf,M.B., Johnson,W.E. and Bass,B.L. (2011) Improved annotation of C. elegans microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. *RNA*, **17**, 563–577.

44. Ruby,J.G., Jan,C., Player,C., Axtell,M.J., Lee,W., Nusbaum,C., Ge,H. and Bartel,D.P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell*, **127**, 1193–1207.

45. Starega-Roslan,J., Krol,J., Koscianska,E., Kozlowski,P., Szlachcic,W.J., Sobczak,K. and Krzyzosiak,W.J. (2011) Structural basis of microRNA length variety. *Nucleic Acids Res.*, **39**, 257–268.

46. Koscianska,E., Starega-Roslan,J., Sznajder,L.J., Olejniczak,M., Galka-Marciniak,P. and Krzyzosiak,W.J. (2011) Northern blotting analysis of microRNAs, their precursors and RNA interference triggers. *BMC Mol. Biol.*, **12**, 14.

47. 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.

48. Katoh,T., Sakaguchi,Y., Miyauchi,K., Suzuki,T., Kashiwabara,S. and Baba,T. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev.*, **23**, 433–438.

49. Wyman,S.K., Knouf,E.C., Parkin,R.K., Fritz,B.R., Lin,D.W., Dennis,L.M., Krouse,M.A., Webster,P.J. and Tewari,M. (2011) post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res.*, **21**, 1450–1461.

50. Berezikov,E., Chung,W.J., Willis,J., Cuppen,E. and Lai,E.C. (2007) Mammalian mirtron genes. *Mol Cell*, **28**, 328–336.

51. 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.

52. Westholm,J.O. and Lai,E.C. (2011) Mirtrons: microRNA biogenesis via splicing. *Biochimie*, **93**, 1897–1904.

53. Flores-Jasso,C.F., Arenas-Huertero,C., Reyes,J.L., Contreras-Cubas,C., Covarrubias,A. and Vaca,L. (2009) First step in pre-miRNAs processing by human Dicer. *Acta Pharmacol Sin*, **30**, 1177–1185.

54. Ando,Y., Maida,Y., Morinaga,A., Burroughs,A.M., Kimura,R., Chiba,J., Suzuki,H., Masutomi,K. and Hayashizaki,Y. (2011) Two-step cleavage of hairpin RNA with 5' overhangs by human DICER. *BMC Mol Biol*, **12**, 6.

55. Ribas,J. and Lupold,S.E. (2010) The transcriptional regulation of miR-21, its multiple transcripts, and their implication in prostate cancer. *Cell Cycle*, **9**, 923–929.

56. Shi,W., Hendrix,D., Levine,M. and Haley,B. (2009) A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nat. Struct. Mol. Biol.*, **16**, 183–189.

57. Taft,R.J., Simons,C., Nahkuri,S., Oey,H., Korbie,D.J., Mercer,T.R., Holst,J., Ritchie,W., Wong,J.J., Rasko,J.E. *et al.* (2010) Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nat. Struct. Mol. Biol.*, **17**, 1030–1034.

58. Ruby,J.G., Stark,A., Johnston,W.K., Kellis,M., Bartel,D.P. and Lai,E.C. (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res.*, **17**, 1850–1864.

59. 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.

60. Langenberger,D., Bermudez-Santana,C., Hertel,J., Hoffmann,S., Khaitovich,P. and Stadler,P.F. (2009) Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics*, **25**, 2298–2301.

61. Berezikov,E., Robine,N., Samsonova,A., Westholm,J.O., Naqvi,A., Hung,J.H., Okamura,K., Dai,Q., Bortolamiol-Becet,D., Martin,R. *et al.* (2011) Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. *Genome Res.*, **21**, 203–215.

62. Muljo,S.A., Kanellopoulou,C. and Aravind,L. (2010) MicroRNA targeting in mammalian genomes: genes and mechanisms. *Wiley Interdisc Rev. Syst. Biol. Med.*, **2**, 148–161.

63. Tuck,A.C. and Tollervey,D. (2011) RNA in pieces. *Trends Genet*, **27**, 422–432.

64. Haussecker,D., Huang,Y., Lau,A., Parameswaran,P., Fire,A.Z. and Kay,M.A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA*, **16**, 673–695.

65. 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.

66. Persson,H., Kvist,A., Vallon-Christersson,J., Medstrand,P., Borg,A. and Rovira,C. (2009) The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat. Cell Biol.*, **11**, 1268–1271.

67. Burroughs,A.M., Ando,Y., de Hoon,M.J., Tomaru,Y., Suzuki,H., Hayashizaki,Y. and Daub,C.O. (2011) Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol.*, **8**, 158–177.

68. Ender,C., Krek,A., Friedlander,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.

69. Smalheiser,N.R., Lugli,G., Thimmapuram,J., Cook,E.H. and Larson,J. (2011) Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. *RNA*, **17**, 166–181.

70. Wang,Y., Juranek,S., Li,H., Sheng,G., Tuschl,T. and Patel,D.J. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, **456**, 921–926.

71. Yuan,Y.R., Pei,Y., Ma,J.B., Kuryavyi,V., Zhadina,M., Meister,G., Chen,H.Y., Dauter,Z., Tuschl,T. and Patel,D.J. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell*, **19**, 405–419.

72. Hock,J. and Meister,G. (2008) The Argonaute protein family. *Genome Biol.*, **9**, 210.

73. Calabrese,J.M., Seila,A.C., Yeo,G.W. and Sharp,P.A. (2007) RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **104**, 18097–18102.

74. 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.

75. Newman,M.A., Mani,V. and Hammond,S.M. (2011) Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA*, **17**, 1795–1803.

76. Kuchta,K., Knizewski,L., Wyrwicz,L.S., Rychlewski,L. and Ginalski,K. (2009) Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res.*, **37**, 7701–7714.

77. Ohrt,T., Mutze,J., Staroske,W., Weinmann,L., Hock,J., Crell,K., Meister,G. and Schwille,P. (2008) Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC *in vivo* in human cells. *Nucleic Acids Res.*, **36**, 6439–6449.

78. Weinmann,L., Hock,J., Ivacevic,T., Ohrt,T., Mutze,J., Schwille,P., Kremmer,E., Benes,V., Urlaub,H. and Meister,G. (2009) Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, **136**, 496–507.

79. Zhang,H., Kolb,F.A., Jaskiewicz,L., Westhof,E. and Filipowicz,W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell*, **118**, 57–68.

80. Zhang,H., Kolb,F.A., Brondani,V., Billy,E. and Filipowicz,W. (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.*, **21**, 5875–5885.

81. 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.

82. Sato,K., Hamada,M., Asai,K. and Mituyama,T. (2009) CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic Acids Res.*, **37**, W277–280.

83. Hamada,M., Yamada,K., Sato,K., Frith,M.C. and Asai,K. (2011) CentroidHomfold-LAST: accurate prediction of RNA secondary structure using automatically collected homologous sequences. *Nucleic Acids Res.*, **39**, W100–106.

84. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

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