

# Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs

CHRISTIAN COLE,<sup>1,6</sup> ANDREW SOBALA,<sup>2,6</sup> CHENG LU,<sup>3</sup> SHAWN R. THATCHER,<sup>3</sup> ANDREW BOWMAN,<sup>2</sup> JOHN W.S. BROWN,<sup>4,5</sup> PAMELA J. GREEN,<sup>3</sup> GEOFFREY J. BARTON,<sup>1</sup> and GYORGY HUTVAGNER<sup>2</sup>

<sup>1</sup>Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom

<sup>2</sup>Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom

<sup>3</sup>Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA

<sup>4</sup>Plant Sciences Division, Scottish Crop Research Institute, University of Dundee, Invergowrie, Dundee DD2 5DA, United Kingdom

<sup>5</sup>Genetics Programme, Scottish Crop Research Institute, University of Dundee, Invergowrie, Dundee DD2 5DA, United Kingdom

## ABSTRACT

Deep sequencing technologies such as Illumina, SOLiD, and 454 platforms have become very powerful tools in discovering and quantifying small RNAs in diverse organisms. Sequencing small RNA fractions always identifies RNAs derived from abundant RNA species such as rRNAs, tRNAs, snRNA, and snoRNA, and they are widely considered to be random degradation products. We carried out bioinformatic analysis of deep sequenced HeLa RNA and after quality filtering, identified highly abundant small RNA fragments, derived from mature tRNAs that are likely produced by specific processing rather than from random degradation. Moreover, we showed that the processing of small RNAs derived from tRNA<sup>Gln</sup> is dependent on Dicer in vivo and that Dicer cleaves the tRNA in vitro.

**Keywords:** small RNA; miRNA; deep sequencing; Dicer; tRNA

## INTRODUCTION

The recent development of high-throughput sequencing technology has accelerated the identification of small regulatory RNAs. Deep sequencing of small RNA fractions from cell culture and animal tissues in which key small RNA processing factors have been impaired, or by cataloging small RNAs bound to different Argonaute proteins, has started to provide valuable data showing how different species of small RNAs are processed and sorted into different regulatory complexes (Calabrese et al. 2007; Azuma-Mukai et al. 2008; Babiarczyk et al. 2008; Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Okamura et al.

2008b; Tam et al. 2008; Watanabe et al. 2008b; for review, see Okamura and Lai 2008).

Animal cells encode a wide arsenal of small RNAs that regulate the expression of protein coding genes and control selfish genetic elements. miRNAs are 21–23-nucleotides (nt) long, fast evolving regulatory RNAs that are mainly coded by short hairpin structures in intronic sequences. They are processed cotranscriptionally into pre-miRNAs by the microprocessor that contains an RNase III enzyme called Drosha, and an RNA binding protein, DGCR8 (Lee et al. 2003; Han et al. 2004). However, mirtrons, miRNAs located in small introns, bypass this process and are generated by splicing (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007). Pre-miRNAs are transported into the cytoplasm where they are further processed into miRNAs by a complex containing Dicer (Grishok et al. 2001; Hutvagner et al. 2001; Yi et al. 2003). Finally, mature miRNAs are loaded into an Argonaute complex that regulates gene expression by altering the protein synthesis of mRNAs or interfering with transcription (for review, see Hock and Meister 2008; Hutvagner and Simard 2008). siRNAs are another source of small regulatory RNAs

<sup>6</sup>These authors contributed equally to this work.

**Reprint requests to:** Gyorgy Hutvagner, Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom; e-mail: g.hutvagner@dundee.ac.uk; fax: 44-1382-386310; or Geoffrey J. Barton, Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom; e-mail: g.j.barton@dundee.ac.uk; fax: 44-1382-385764.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1738409>.

that are processed from diverse naturally formed double-stranded RNA precursors. The majority of endo-siRNAs are derived from transposable and repetitive elements, and they are believed to regulate the expression of transposons (Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008). Hairpin-derived endogenous siRNAs were shown to be processed to their mature form in *Drosophila* and mouse ES cells (Babiarz et al. 2008; Okamura et al. 2008b). *Cis*- and *trans*-derived endogenous siRNA loci have also been discovered in flies and mammalian cells (Okamura et al. 2008a; Tam et al. 2008; Watanabe et al. 2008a,b). The common feature of these siRNAs is that their processing is not dependent on Drosha and DGCR8 (Babiarz et al. 2008; Ghildiyal et al. 2008). Some classes of endo-siRNAs were shown or predicted to target mRNAs (Okamura et al. 2008b; Tam et al. 2008; Watanabe et al. 2008b). Piwi-associated small RNAs (piRNAs) are longer (27–31-nt long) small RNAs that exclusively regulate the expression of transposable elements in the germline. The maturation of these types of regulatory RNAs are Drosha and Dicer independent, and they are generated by consecutive cleavage activity of different Piwi proteins (Brennecke et al. 2007; Gunawardane et al. 2007).

Cloning and sequencing of small RNA libraries always generates fragments that are derived from abundant non-coding RNAs such as rRNAs, tRNAs, snoRNAs, and snRNAs. These RNAs are generally considered to be random degradation products, and in some cases it has been shown that their abundance does not change in a mutant background that impairs canonical regulatory small RNA production (Calabrese et al. 2007). Furthermore, in yeast, a mechanism has been discovered that prevents rRNA and tRNA species from entering the small RNA pathway (Buhler et al. 2008). However, recent data suggest that abundant noncoding RNAs can produce stable small siRNA-sized RNA fragments that are unlikely to be random degradation products, and one tRNA has been shown to produce a Dicer-dependent small RNA that is expressed with low abundance in mouse ES cells (Babiarz et al. 2008; Kawaji et al. 2008). A similar tRNA fragment was also detected by Northern hybridization in immortalized human cell lines; however, its processing properties were not investigated (Kawaji et al. 2008). In addition, it has been recently demonstrated that snoRNAs can produce small RNAs that are processed in a Drosha-independent and Dicer-dependent manner and incorporate into functional Argonaute complexes (Ender et al. 2008). In this paper we show that quality filtering of deep sequencing data reveals a wide range of small RNAs that are likely products of processing events. Interestingly, we have found that an abundant small RNA species in HeLa cells is derived from the 5' ends of a series of mature tRNAs, and have shown that the production of small RNAs derived from tRNA<sup>Gln</sup> is Dicer dependent.

## RESULTS

### Purification of small RNAs

Since miRNAs (Politz et al. 2006) and small RNAs processed from snoRNAs (AI Lamond, pers comm.) could be detected in the nucleolus, we initiated deep sequencing to identify the small RNA populations of this cellular compartment. We isolated total RNA from purchased purified HeLa nucleoli (Supplemental Fig. 1A) and deep sequenced the small RNA fraction by the Solexa method following a published protocol (Lu et al. 2007). However, post-sequencing testing of the nucleoli and other fractions derived from the purification steps have revealed that our original sample was contaminated with nucleoplasmic and, to a lesser extent, cytoplasmic materials (Supplemental Fig. 1B).

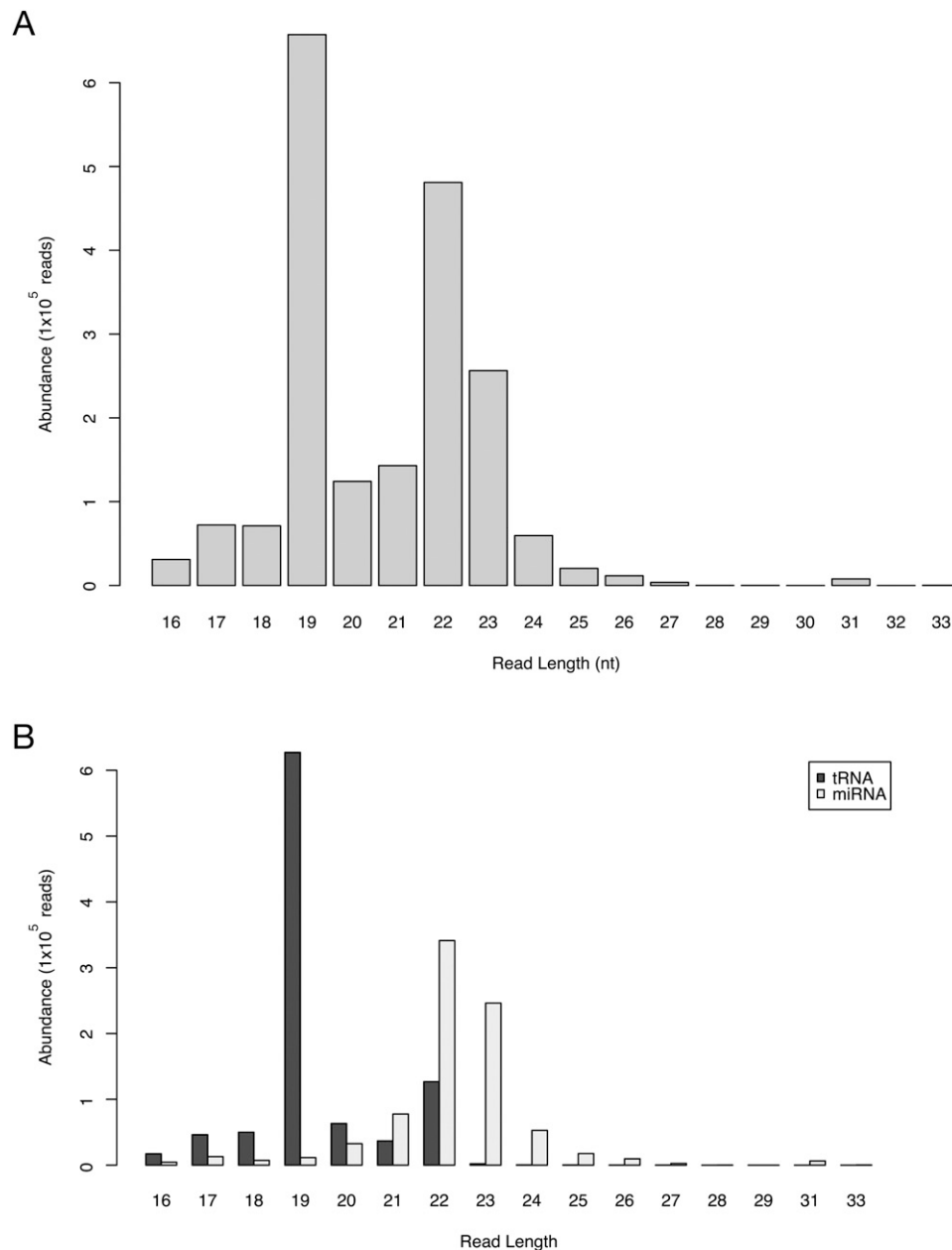
### Small RNA sequence analysis

A total of 2.3 million sequence reads were obtained from the sequencing experiment and following filtering and collation (see Materials and Methods), 22,386 unique reads remained. The sequences were size-selected small RNAs, with a distribution of lengths up to 33 nt, as shown in Figure 1A. The length distributions show peaks at around 19 and 22 nt, which are discussed below.

The sequence reads were compared by Vmatch (Abouelhoda et al. 2004) to the miRNA database miRBase (Griffiths-Jones 2004) and the human complement of noncoding RNA (ncRNA) sequences available from Ensembl (Birney et al. 2004). Table 1 details matches to known ncRNAs and Supplemental Table 1 contains all noncoding matches ranked by abundance. The vast majority of matches found were to known tRNA and miRNA sequences, but other RNA species were also represented. For example, the recently published (Ender et al. 2008) snoRNA (ACA45, ENST00000390972) was found in our data set with an abundance of 22 reads and the matching region overlaps that found by Ender and coworkers. Figure 1B details the length distribution of the tRNA and miRNA read sequences. Comparing Figure 1, A and B, it is clear that the two peaks in the overall length distribution correspond to the tRNA (Fig. 1B, dark gray bars) and the miRNA (Fig. 1B, light gray bars) reads, revealing that the tRNA sequences have a preferred length of 19 nt and the miRNAs have a preferred length of 22 nt. This is consistent with the understood maturation of miRNAs following processing by the Dicer enzyme. The reads match to 190 known human miRNA mature sequences as detailed in Supplemental Table 2.

### Specific processing of miRNAs and tRNAs

The tRNA fragments observed might be the result of random degradation of tRNA, or some specific processing event that leads to 19-mer reads. Visualization of the aligned



**FIGURE 1.** Length distribution of all sequence reads. (A) Plot shows the length distribution by abundance of each of the 22,836 unique sequences obtained from deep sequencing of HeLa cell extracts. (B) Length distribution of tRNA and miRNA matching reads. The length distribution of the sequence reads of classified by matching to either tRNA (dark gray bars) or miRNA (light gray bars).

tRNA sequence reads and comparing to other noncoding RNAs (ncRNAs) shows two types of matching. Ladder-type matching, as illustrated in Supplemental Figure 2A, suggests the reads are the result of random degradation of RNA. In contrast, reads that are “stacked” as shown in Supplemental Figure 2B, with the majority of reads matching to one region of the RNA suggest that the RNA sequences observed are the result of specific processing by cleavage enzymes, such as Dicer in the case of miRNAs.

In order to quantify the amount of specific processing observed for the ncRNA matches, a simple metric  $S_p$  was

defined as the number of reads matching to an RNA region divided by the length of the matching region. A larger  $S_p$  indicates a higher propensity for specific processing. Values of  $S_p$  for the different ncRNA matches are detailed in Table 2 and box plots shown in Figure 2. miRNAs were found to have a mean  $S_p$  of 139, but interestingly, tRNA sequences showed a higher  $S_p$  than the miRNAs with a mean  $S_p$  of 490 and includes very high scoring regions with  $S_p$  of up to 15,333. This is significantly different to the other ncRNAs ( $P \leq 6.2 \times 10^{-4}$  except for the tRNA pseudogenes, scRNA and snoRNA pseudogenes) whose highest  $S_p$  values are in

**TABLE 1.** Details of reads matching to known ncRNAs determined with Vmatch

ncRNA type <sup>a</sup>	Known ncRNAs	Known matched (up to two mismatches)	Matched (%)	Total abundance	Abundance (per matching read)
tRNA	523	485	92.7	970,836	2001.7
miRNA	692	403	58.2	823,836	2044.3
snoRNA	758	384	50.7	9868	25.7
rRNA pseudogene	341	85	24.9	7222	85
Mt tRNA	22	22	100	2796	127.1
rRNA	333	230	69.1	2247	9.8
Misc RNA	934	505	54.1	1359	2.7
snRNA	1288	589	45.7	1068	1.8
Mt rRNA	2	2	100	750	375
tRNA pseudogene	129	14	10.9	433	30.9
snRNA pseudogene	501	15	3	363	24.2
scRNA pseudogene	843	142	16.8	314	2.2
snoRNA pseudogene	486	12	2.5	230	19.2
scRNA	1	1	100	179	179
Mt tRNA pseudogene	603	4	0.7	55	13.8
Misc RNA pseudogene	7	0	0	0	0

Identified noncoding RNA (ncRNA) types (as defined in Ensembl or, miRNAs only, miRBase) matching sequence reads as determined with Vmatch with up to two mismatches were allowed. Known ncRNAs: number of ncRNA present in database; Known matched (up to two mismatches): number of ncRNAs with matching reads; Matched (%): number of ncRNAs with matching reads as a percentage of all present in database; Total abundance: number of reads which match to each ncRNA type; and Abundance (per matching read): total abundance/known matched.

<sup>a</sup>As annotated by Ensembl and miRBase (miRNAs only).

the region of 9.3 (e.g., rRNAs) (Fig. 2). Overall, the tRNAs appear to be more precisely processed than the miRNA population as measured by  $S_p$  (Table 2), with the exception of miR-21, which had an  $S_p$  of 14,893.

miRNAs have a well-characterized processing pathway, involving the enzyme Dicer, which cleaves the pre-miRNA hairpin to the mature miRNA at a length of ~21 nt. The high  $S_p$  for the matched miRNAs reflects this processing. Similarly, the high  $S_p$  for the matched tRNAs suggests the presence of specific processing for these species.

A number of matches to pseudogenes were also found (Table 2). In particular, matches to 85 rRNA pseudogenes were found with a total abundance of 7222, the fourth most abundant ncRNA in the data set and also more abundant than “real” rRNAs (Table 2). It should be noted that ncRNAs found in Ensembl are annotated by sequence matching to known RNAs of similar function (Griffiths-Jones et al. 2005) and by prediction algorithms such as tRNAscan-SE (Lowe and Eddy 1997). Thus, the matched pseudogene ncRNAs may not be genuine ncRNAs or, as the RNA extracts are from HeLa cells, it is possible that they are a product of uncontrolled transcription known to happen in cancerous cells and may not have a biological function.

### miRNA

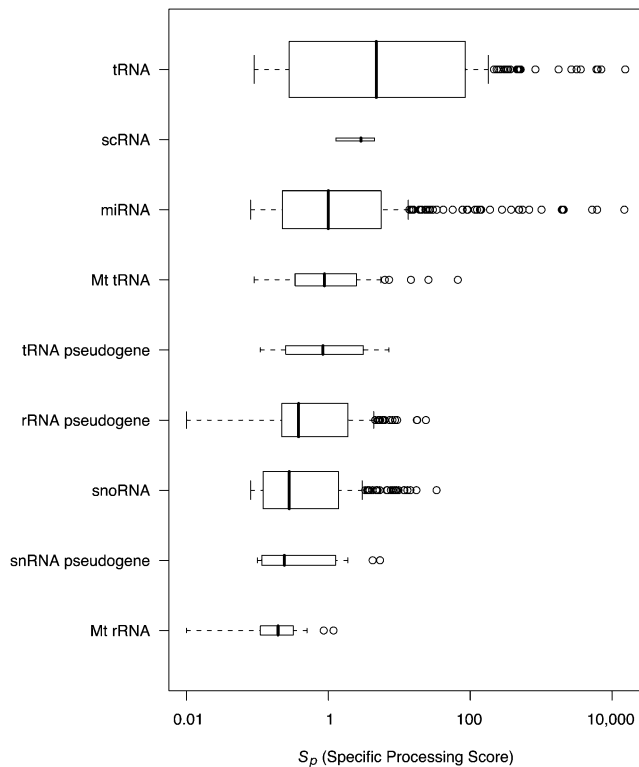
miRNAs undergo several processing steps before becoming “mature” transcripts. The final step is the cleavage by Dicer of the stem–loop from the hairpin structure formed by the

single-stranded RNA molecule. The cleavage produces a short (~21 nt in length) section of double-stranded RNA where one strand is retained as the mature miRNA molecule and the other strand is degraded. The mature miRNA in association with the Argonaute complex then

**TABLE 2.** Mean processing scores for known ncRNAs

ncRNA type	Mean processing scores, $S_p$ (SD)
tRNA	490 (1955)
miRNA	139 (1039)
tRNA pseudogene	1.7 (2.1)
rRNA pseudogene	0.8 (1.7)
snoRNA	1.8 (3.7)
Mt tRNA pseudogene	0.8 (1.4)
Mt tRNA	5.1 (13.3)
rRNA	2.4 (2.3)
scRNA pseudogene	0.5 (0.8)
misc RNA	0.4 (0.5)
snRNA	0.8 (0.8)
snRNA pseudogene	1.0 (1.5)
Mt rRNA	0.3 (0.3)
snoRNA pseudogene	1.5 (1.7)
scRNA	2.9 (2.3)

Specific processing score ( $S_p$ ) averages for each of the known, Ensembl determined, noncoding RNA (ncRNA) types. All ncRNA regions that align with sequence reads with an abundance >1 were averaged.



**FIGURE 2.** Box plot showing the distribution of specific processing scores,  $S_p$ . Only noncoding RNA types with an abundance per matching read of  $>20$  (see Table 1) are plotted. The *left* and *right* of the boxes represent the first and third quartiles of the data (the interquartile range), and the bold line within the box is the median. The dotted lines represent 1.5 times the interquartile range and data points outside deemed to be outliers (open circles). The height of the boxes represent the amount of data there is with taller boxes having more data.

silences its target messenger RNA as previously described. However, sometimes the complementary strand of the stem is detected, but at much lower levels than the main product, and is referred to as the “star” sequence. Supplemental Figure 4 shows the mature and star sequence reads aligning to the human *let-7a-1* hairpin sequence. Reads aligning to the major product can be seen at position 5 on the hairpin sequence with a total abundance of 58,490 reads, whereas the star sequence aligns to position 56 with a total abundance of four reads. A total of 22 miRNAs matching both the major and star regions of the hairpin sequence were found and are detailed in Supplemental Table 3.

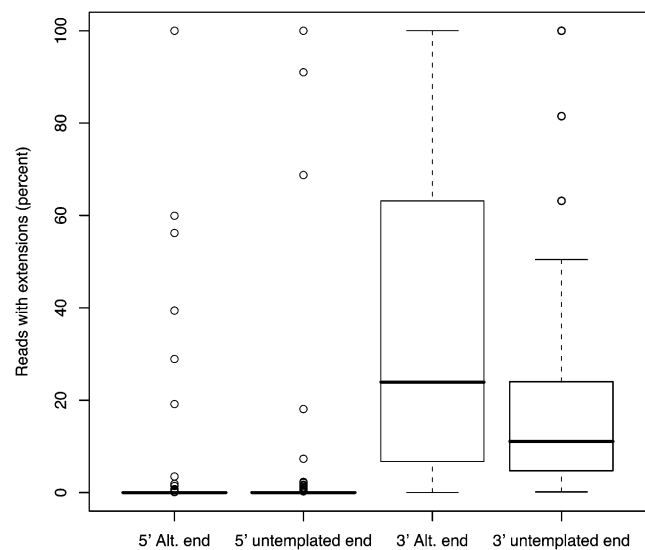
It has been previously noted that in vitro miRNA processing of the 5' and 3' ends is not perfect (Azuma-Mukai et al. 2008). In particular, at the 3' end, reads were found to have additional bases compared with the mature miRNA sequence annotated in miRBase. Figure 3 details the 5' and 3' end processing for miRNAs found in vivo in this experiment and shows that the 5' end of reads are only rarely extended, whereas the 3' ends show a much greater prevalence of extension ( $P \leq 3.8 \times 10^{-10}$ ). Extensions can

be either “alternative” (i.e., identical to the native hairpin sequence) or “untemplated” (i.e., differing from the native hairpin sequence). As shown in Figure 3 the 3' ends of miRNA matching reads show a preference for “alternative” reads with an average of 37.6% reads containing this extension ( $P < 0.05$ , when compared with 3' “untemplated” ends), confirming previously published in vitro data (Azuma-Mukai et al. 2008).

## tRNA

The tRNA sequence matches also have low abundance reads at the 3' end that are complementary to the high abundance 5' end. This suggests that they may be being cleaved from a hairpin intermediate via an equivalent process to the miRNAs which produces a mature and a star (\*) strand. Figure 4A shows an example tRNA with the major product aligning to position 1 of the whole tRNA sequence with a total abundance of 567,275 reads and a minor product (the star sequence) aligning to position 50 with an abundance of 18. There is some evidence of degradation products from the tRNA near the bottom of the alignment, but overall the alignment is dominated by reads aligning to the 5' end of the tRNA.

The 3' end of the tRNA matching 19mer reads is located in the D-loop of the tRNA structure. The D-loop forms a stabilizing contact with the T-loop at the conserved nucleotides  $G^{18}G^{19}$  and in *Drosophila* has been shown to



**FIGURE 3.** 5' and 3' processing variability of miRNA matching reads. The box plots detail the distributions of the percentage of reads that show extensions at the 5' and 3' ends for each of the miRBase annotated mature sequences. To avoid matching to degradation products, extensions at the 5' end were limited to a maximum of two bases. Two types of extensions are determined for each end; when at least one of the extension bases match the hairpin miRNA sequence (Alt. end), and instances when one or more of the extended bases do not match the hairpin miRNA sequence (untemplated end).

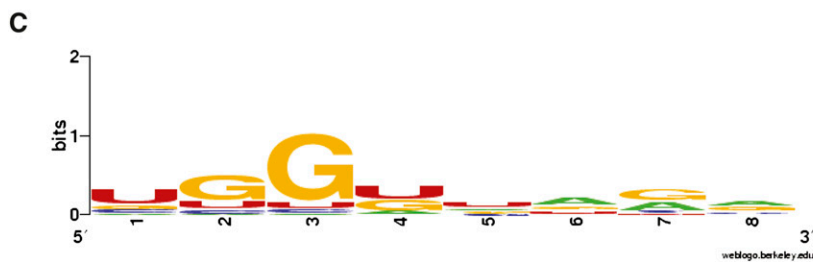
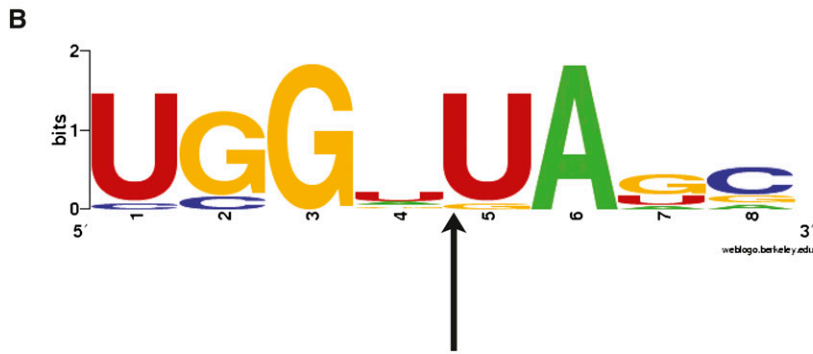
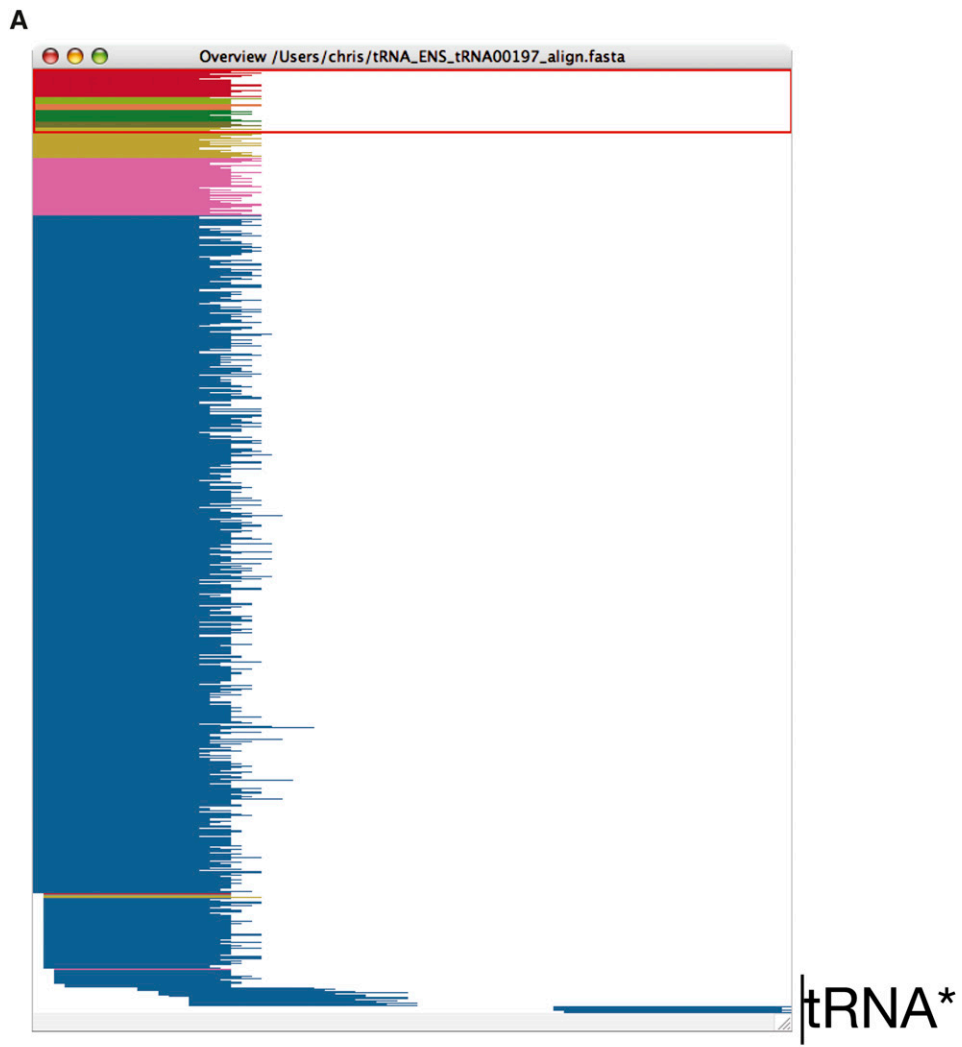


FIGURE 4. (Legend on next page)

be a required contact for RNase P and 3'-tRNA processing (Levinger et al. 1998). In humans, the GG motif is found at bases 17 and 18. Interestingly, the processing observed here occurs at the base immediately following this motif. Analyzing the sequence motif four bases either side of the tRNA cleavage position shows that the bases 3' to the cleavage site have a distinct preference for nucleotides U<sup>20</sup>A<sup>21</sup> (Fig. 4B) when compared with a nonredundant set defined from all tRNA sequences (Fig. 4C). The nonredundant set includes the unique 8mer sequences centered around position 19 for all the known human tRNA sequences. Dicer shows a preference for cleavage at positions 5' to uridine (Aravin et al. 2003) and, thus, is a possible candidate for the selective processing of the tRNAs.

The abundance of tRNA genes is nonuniform across the genome, with the number of tRNA genes varying between 3 (Selenocysteine) and 43 (Alanine), giving 523 tRNA genes in total. Many of the genes are identical at the transcript level, and removing identical representatives leaves 353 unique examples. Figure 5A shows a plot of the number of reads for each tRNA type against the number of unique transcripts and reveals that four types dominate in terms of abundance: lysine, valine, glutamine, and arginine. There does appear to be a general trend where the more abundant tRNAs are those with the most transcripts, but leucine and alanine seem to be underrepresented in abundance. It could mean also, that the tRNAs exist at a basal or steady-state level of ~10,000 reads or fewer and the high abundance tRNAs show an overexpression over the basal levels.

Owing to the degenerate nature of mRNA codons required during translation there are several different codons possible for each amino acid. Differences in the abundance of each tRNA anticodon could reveal selection for specific anticodons within tRNA types. Of the high abundance tRNAs mentioned previously, tRNA<sup>Lys</sup> and tRNA<sup>Gln</sup> have two anticodons, tRNA<sup>Val</sup> has three anticodons and tRNA<sup>Arg</sup> has five. Figure 5B shows that four anticodons are present at significantly higher abundance than the others: tRNA<sup>Val</sup> AAC and CAC, tRNA<sup>Lys</sup> TTT, and tRNA<sup>Gln</sup> CTG, each having around 100,000 reads or more, where the remainder have around 10,000 or fewer. The tRNA<sup>Arg</sup> abundance appears to be spread across all five anticodons with none dominating over the majority of the tRNAs. The underrepresented anticodons tRNA<sup>Ala</sup> AGC, tRNA<sup>Asn</sup> GTT, and tRNA<sup>Cys</sup> GCA show a low abundance despite their high number of transcripts.

This suggests that the high abundance tRNAs found in HeLa cells are found at higher levels than expected based on

the number of known tRNA genes in the genome, and is even more marked when looking at anticodon abundance. HeLa cells are fast growing cells and have a high requirement for processes involved in translation, such as the availability of tRNAs. The specific requirement for high abundance of tRNA<sup>Gln</sup>, tRNA<sup>Lys</sup>, and tRNA<sup>Val</sup> is not clear in a search for proteins with large requirements for these amino acids (data not shown), but may still be of biological relevance.

### Detection of small RNA derived from a tRNA

A tRNA that showed high specific processing was chosen from the sequence data analysis for further investigation. tRNA<sup>Gln</sup> (Fig. 6A, total abundance of 82,739) shows a preference for 19mer fragments (Fig. 6B) and has an  $S_p$  of 2,669. The abundance and processing score are comparable with miRNA *let-7a*. An RNA probe, complementary to the tRNA<sup>Gln</sup> 5' fragment, was used to detect a discrete band at around 20 nt (Fig. 6C) via Northern blotting. This fragment showed exclusive cytoplasmic localization in contrast to the miRNA *let-7* that was also detected in the nuclear fraction (Fig. 6C).

### Processing of small RNA derived from tRNA<sup>Gln</sup> is Dicer dependent

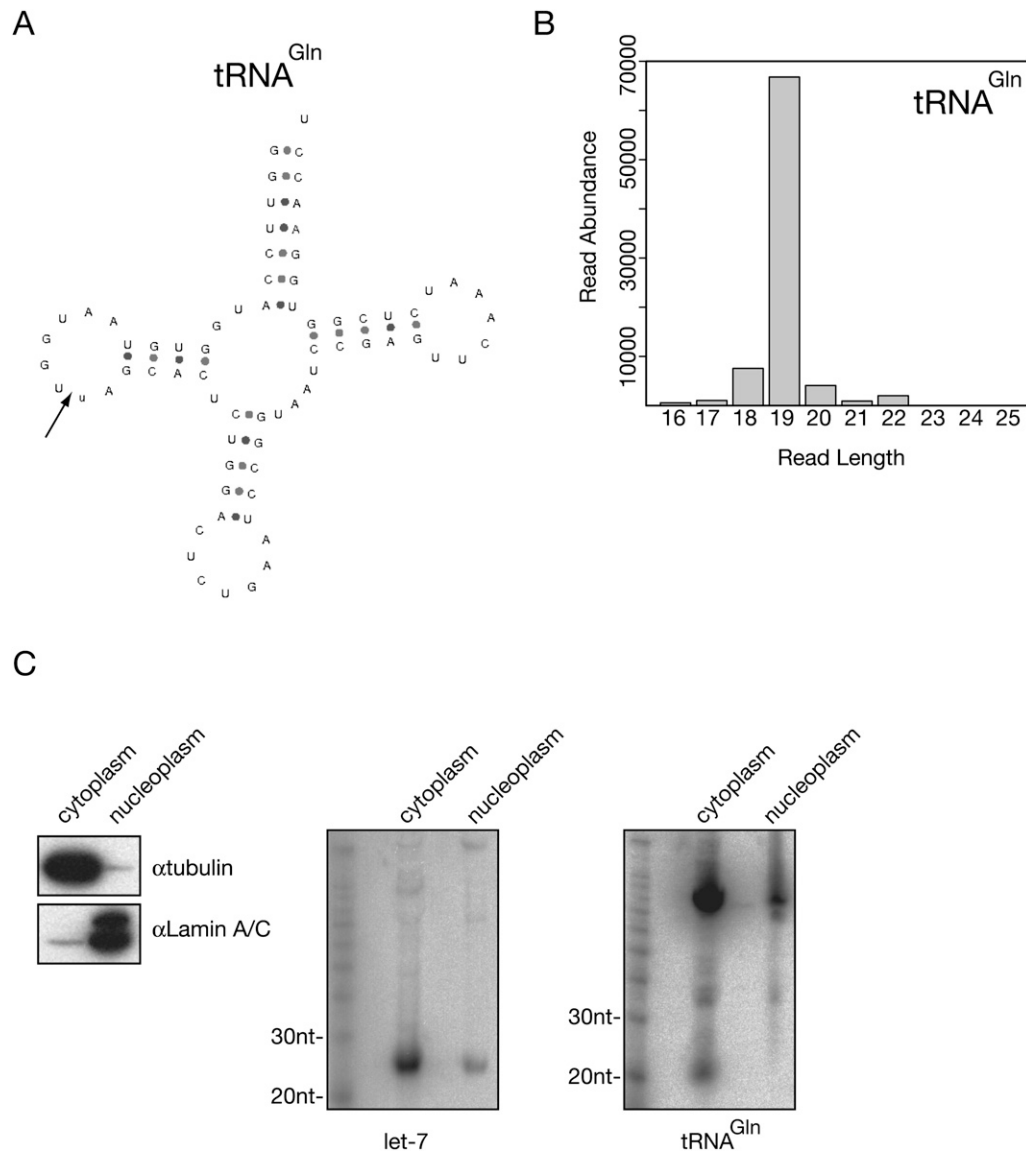
Dicer activity is Mg<sup>2+</sup> dependent (Zhang et al. 2002), thus incubating cell extract with varying amounts of MgCl<sub>2</sub> may give evidence of Dicer-dependent processing of tRNA<sup>Gln</sup>. Figure 7A shows the effect of incubating HeLa S100 extract with 0, 2, and 4 mM concentration of MgCl<sub>2</sub> where the amount of tRNA<sup>Gln</sup> cleavage product at ~20 nt increases with MgCl<sub>2</sub> concentration.

Suppression of Dicer expression with siRNA as shown in Figure 7B reveals a marked decrease in the small RNA fraction derived from tRNA<sup>Gln</sup>, compared with the non-targeting siRNA, further suggesting that Dicer is involved in the generation of this small RNA species. Since this approach resulted in a very modest effect on the steady state level of miR-21 the experiment was repeated in HEK 293 cells that encode doxycycline inducible siRNA against Dicer (Schmitter et al. 2006). A more efficient knock down of Dicer was achieved that resulted in a significant decrease of endogenous miR-21, and again, it showed that the abrogation of Dicer causes a strong decrease of the steady-state level of tRNA<sup>Gln</sup> small RNAs (Fig. 7C).

**FIGURE 4.** (A) Overview window from Jalview (Clamp et al. 2004; Waterhouse et al. 2009) of sequence reads aligning to a known tRNA sequence as determined by Vmatch. Sequence reads are color coded by abundance with red >500 reads and blue <10 reads. The majority of the reads align to the 5' end of the tRNA. A small number of complementary sequences can be seen aligning to the 3' end: tRNA\*. (B) Sequence motif logo for tRNA 8mer sequence around the 19–20 position of selective cleavage. Nonredundant set of tRNAs with matching read abundances of at least 2000 (B), and a nonredundant set of all human tRNAs (C). Arrow indicates the tRNA cleavage site. Images were created with WebLogo (Crooks et al. 2004).







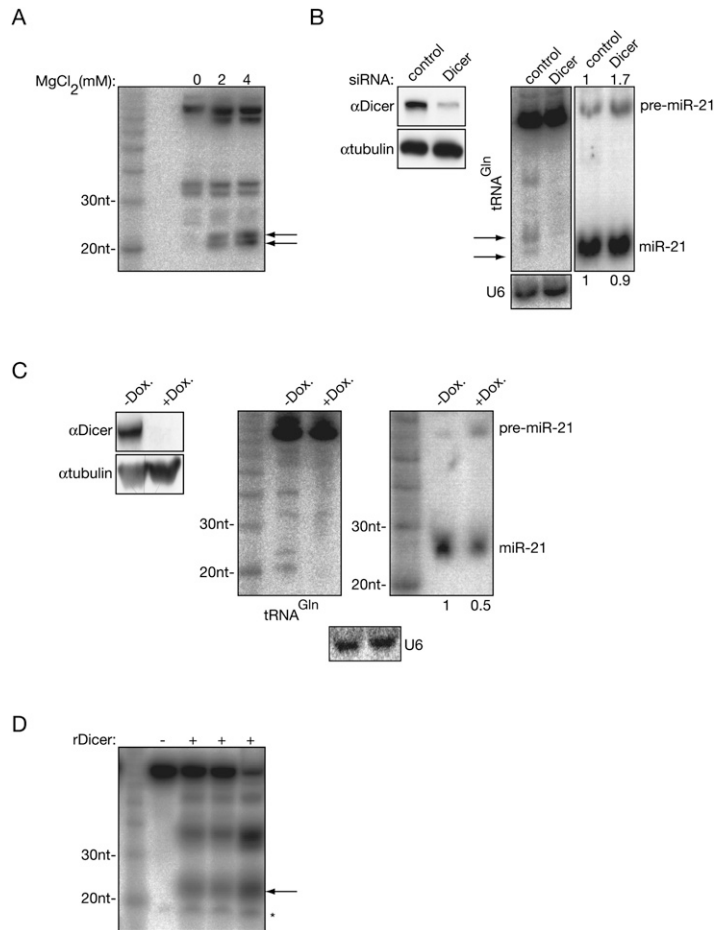
**FIGURE 6.** Small RNA derived from tRNA<sup>Gln</sup>. (A) Putative folding of tRNA<sup>Gln</sup>, arrow shows the predicted cleavage site. (B) Size distribution of small RNAs derived from tRNA<sup>Gln</sup>. (C) Northern hybridization shows that a ~20-nt-sized fragment derived from tRNA<sup>Gln</sup> exclusively localized in the cytoplasm (*right panel*) while the *let-7* miRNA also could be detected in the nucleoplasmic fraction (*middle panel*). The efficiency of cell fractionations was checked with Western blotting for cell fraction specific proteins (*left panel*; tubulin for cytoplasm, Lamin A/C for nucleoplasm).

The modification of the 2' and 3' OH of the last nucleotide of a miRNA and siRNA could be characteristic of their association with specific Argonaute proteins (Forstemann et al. 2007; Tomari et al. 2007) and the presence of the modification in tRNA-derived small RNA would suggest it may have miRNA or siRNA function. Performing  $\beta$ -elimination reactions on miR-21 and the tRNA<sup>Gln</sup> fragment reveals that tRNA<sup>Gln</sup> is resistant to the treatment in contrast to miR-21 where mobility is increased (Fig. 8C). This suggests that the 3' terminal ribose of tRNA-derived small RNA is modified. This finding could potentially explain why tRNA fragments are not efficiently incorpo-

rated into Ago complexes since human miRNAs are unmodified on the 3' terminal ribose.

## DISCUSSION

Deep sequencing of mixed HeLa cell extracts revealed that miRNAs and small RNAs derived from tRNAs are the two major sources of small RNAs in this cell type. The four most abundant tRNAs (tRNA<sup>Lys</sup>, tRNA<sup>Val</sup>, tRNA<sup>Gln</sup>, and tRNA<sup>Arg</sup>) in HeLa cells produce small RNA fragments with an abundance of >20,000 reads each, comparable to the five most abundant miRNAs (miR-21, let-7f-1, let-7a-1, let-7c,



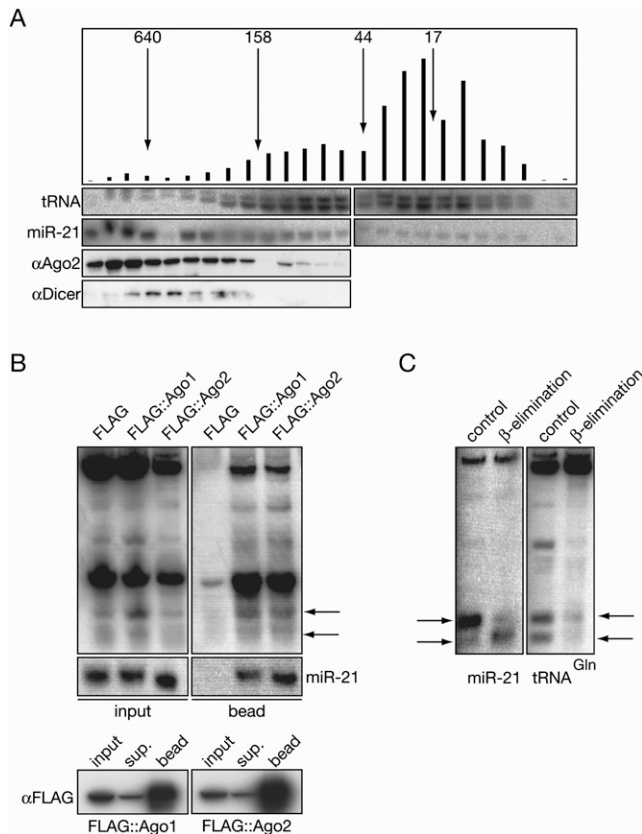
**FIGURE 7.** The processing of small RNAs derived from tRNA<sup>Gln</sup> is Dicer dependent. (A) The processing of endogenous tRNA<sup>Gln</sup> into small RNA could be stimulated with MgCl<sub>2</sub>. S100 extract was supplemented with increasing concentration of MgCl<sub>2</sub> and incubated for 1 h at 37°C. The tRNA fragments were detected with Northern hybridization. (B–D) Dicer is required for the generation of the small RNA derived from tRNA<sup>Gln</sup>. (B) Dicer was knocked down with siRNA in HeLa cells. The efficiency of the knock down was tested with Western blotting and tubulin was used as a loading control (*left* panel). Northern blots show the marked decrease of the tRNA-derived small fragment and the accumulation of pre-miRNA in the Dicer knocked down cells (*right* panel). U6 was used as a loading control. The relative abundance of pre-miR-21 and mature miR-21 to the U6 loading control was normalized to the RNA level of the control siRNA transfected cells. The result of the quantification is indicated on the *top* and the *bottom* of the Northern blot (*right* panel). (C) Similar experiment described in B, only Dicer was knocked down in 293 cells with doxycycline (Dox.) inducible shRNA. (D) Recombinant Dicer was incubated with uniformly labeled tRNA<sup>Gln</sup>, and the products were separated in 15% denaturing gel. “–,” RNA incubated with reaction buffer only; “+,” independent processing experiments; “arrows,” tRNA-derived small RNAs; and “\*,” a non-specific fragment present in the mock Dicer processing assay.

and let-7f-2). The tRNAs are almost exclusively processed from the 5' end with cleavage by Dicer at the D-loop that results in the generation of ~19-nt-long RNA fragments. We have shown evidence that the tRNA fragments we identified are not random degradation products. First, their pattern shows a very precise processing and the  $S_p$  value of these tRNA-derived small RNAs exceed even the  $S_p$  of the specifically processed miRNAs. Second, we demonstrated through tRNA<sup>Gln</sup> that the small RNA produced from tRNAs requires the endonuclease Dicer.

Although these small tRNA fragments are generated using Dicer, the same enzyme as miRNAs, they differ from miRNAs in several ways. First, they are more sensitive to the ablation of Dicer than miRNAs. miRNAs are efficiently incorporated into Argonautes, forming a stable complex; therefore, Dicer knock down with siRNAs for a short period of time sometimes does not result in a significant change in the miRNA steady-state level. In contrast, we observed that the tRNA fragment disappeared after a short siRNA treatment against Dicer. Our fractionation and immunoprecipitation experiments suggest that the tRNA<sup>Gln</sup>-derived small RNAs did not efficiently associate with human Argonautes, and the majority of these small RNAs are seen in fractions smaller than the minimal RISC on a size-exclusion column. Therefore, it is likely that small RNAs from tRNAs either do not associate with the proteins at all, or that these small RNAs incorporate into non-Argonaute complexes that have a faster turnover rate.

A second feature that separates tRNA-derived small RNAs from normal mammalian miRNAs is that they are modified at the terminal 3' ribose as demonstrated by resistance to β-elimination. Plant miRNAs, piRNAs, and siRNAs and fly siRNAs are all methylated at the terminal ribose and miRNAs in these organisms that are sorted into Ago2 complexes are also 3' modified (Matranga and Zamore 2007). Mammalian miRNAs are free from terminal ribose modifications. We do not know whether the small RNA processed from tRNA<sup>Gln</sup> is modified after Dicer processing or if it is modified prior to processing. The latter would be unsurprising since tRNA bases are heavily modified and ribose 2'-O-methylation is one very common modification.

Specific cleavage of tRNAs has been reported in various organisms. Ago associated tRNA fragments were identified in *Schizosaccharomyces pombe* in a mutant in which the TRAMP complex, a complex that is responsible for preventing abundant RNAs such as rRNAs and tRNAs to enter the siRNA pathways, was disrupted. This small tRNA fragment is also generated from the 5' end of the tRNAs; however, it is 23-nt long and its processing is Dicer



**FIGURE 8.** Small RNAs derived from tRNA<sup>Gln</sup> poorly associate with Argonaute proteins. (A) small tRNA<sup>Gln</sup> RNAs partially cofractionate with Ago2. HeLa S100 extract was fractionated through a Superdex-200 column. Every second fraction was tested for Ago2 and Dicer with Western blotting, for miR-21 and small tRNA<sup>Gln</sup> (labeled as tRNA) species with Northern hybridizations. Fractions containing the corresponding size markers (kDa) are indicated with arrows. Vertical bars are representing the level of small tRNA in the corresponding fractions in arbitrary units. (B) RNAs derived from tRNA<sup>Gln</sup> incorporate into Ago1 and Ago2 complexes. FLAG immunoprecipitations were carried out from cells that were transfected with empty FLAG, FLAG::Ago1, and FLAG::Ago2 fusion plasmids. The total lysate and the bound fractions were assayed with Northern hybridization using complementary RNA to the sequenced small tRNA<sup>Gln</sup> fragment as a probe. Black arrows indicate the small RNAs derived from the tRNA. (C) small RNAs derived from tRNA<sup>Gln</sup> are resistant to  $\beta$ -elimination. Northern hybridizations show the results of the effect of  $\beta$ -elimination on miR-21 and small RNA processed from tRNA. Black arrows indicate miRNA and small tRNA fragments.

independent (Buhler et al. 2008). A Dicer-dependent small tRNA fragment processed from the 3' terminus has been annotated in mouse ES cells from a tRNA, which could alternatively fold into miRNA-like secondary structure (Babiarz et al. 2008). Short, 22-nt-long tRNA fragments corresponding to the 3' end of certain tRNAs have been sequenced, and in the case of tRNA<sup>Glu</sup> this fragment was detected with Northern hybridization in several human cells; however, the enzyme that generates such fragments had not been identified (Kawaji et al. 2008). In

addition to these short RNA molecules, tRNA cleavage to produce stable half-tRNAs has been reported in *Aspergillus fumigatus*, where a developmentally regulated endonuclease activity was detected that cleaves tRNAs at the anticodon (Jochl et al. 2008). Similar activity has been recently demonstrated in yeast, plants, and mammalian cells to be a conserved response to oxidative stress (Thompson et al. 2008).

The discovery of Dicer-dependent small tRNA fragments in human cells automatically raises the question of their biological function. It cannot be ignored that a minor fraction of tRNA-derived small RNAs cofractionate and coimmunoprecipitate with Argonautes; therefore, it is plausible that these tRNA-programmed RISCs have messenger RNA targets. However, another possibility—supported by the relatively high turnover rate of these RNA molecules—is that the act of tRNA cleavage itself has a biological effect. It has been shown that improper folding of tRNAs in *E. coli* produces a hairpin structure that could conceivably cause the tRNA to be a Dicer target (Madore et al. 1999). Why does tRNA become a substrate of Dicer in HeLa cells, while other deep sequenced human cells do not show the accumulation of this species of small RNA? If we exclude the possibility that Dicer-dependent tRNA-derived small RNAs were overlooked in other cells so far (we could not identify similar species of small RNAs in deep sequenced primary neurons) (Remenyi J, Hunter JC, Cole C, Impey S, Mouk EC, Martin K, Barton GJ, Hutvagner G, and Arthur JS, unpubl.) then one explanation is that the accumulation of these RNAs represents an aberrant state of the cells.

One can envision that tRNAs, as an abundant RNA species, compete with pre-miRNAs for Dicer resulting in impaired miRNA homeostasis. Changes in Dicer activity and miRNA level have been proposed multiple times to be a significant factor in cancer development. tRNAs may be accessible by Dicer due to a lack of RNA quality control in HeLa cells, analogous to the mechanism to one that has been described in *S. pombe* (Buhler et al. 2008). Alternatively, increased tRNA levels might be responsible for the Dicer-dependent processing of tRNAs by forcing Dicer to accept tRNA as a substrate. It has recently been shown that overexpression of tRNA can induce cell transformation, although the exact mechanism is not known yet (Marshall et al. 2008). It is tempting to speculate that there is a connection between the accumulation of the Dicer dependent small tRNAs and the aberrantly elevated tRNA expression in transformed cells and the transformation is caused by the competition between tRNAs and pre-miRNAs for Dicer accessibility, resulting in impaired miRNA homeostasis, particularly as only certain tRNA anticodon types are expressed at high levels and processed. Alternatively, the cytoplasmic miRNA processing machinery may be actively processing misfolded inactive tRNAs as a defense response to tRNA overexpression.

## MATERIALS AND METHODS

### Antibodies and oligonucleotides

Antibodies and dilutions used for Western blotting were as follows: monoclonal mouse anti- $\alpha$ -tubulin (Sigma), 1:5000; monoclonal mouse antilamin A/C (Santa Cruz), 1:1000; mouse antidicer (Abcam), 1:250; and HRP-conjugated goat-antimouse IgG (Jackson ImmunoResearch), 1:10,000. For Northern blot hybridization, the following RNA oligonucleotides were synthesized (Eurofins MWG):

let-7 (5'-UAUACAACCUACUACCUCAUU-3');  
mir-21 (5'-UCAACAUCAGUCUGAUAGCUA-3'); and  
tRNA<sup>Gln</sup> (5'-ACCAUUAACCAUGGAACC-3').

For U6 hybridization the following DNA oligonucleotides were used:

U6-fwd (5'-GGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGG-3'); and  
U6-rev (5'-CCTTGCGCAG-3').

Oligonucleotides used for siRNA knockdown of Dicer-1 were RNA duplexes (Dharmacon) with the following sense sequences:

5'-UAAAGUAGCUGGAAUGAUGUU-3';  
5'-GAAUAUGGUUGUUUGAAGAUU-3';  
5'-ACACAGCAGUUGUCUUAAAUU-3'; and  
5'-GAAUAUCGAUCCUAUGUUCUU-3'.

The nontargeting duplex (MWG) had the following sense sequence:

5'-AGGUAGUGUAAUCGCCUUGdTdT-3'.

For making Gln<sup>CTG</sup> tRNA, the following DNA oligonucleotides were used:

Transcript: 5'-AGGTCCAGTGAGACTTGAACCTGATCACTGTATTCAGAGTCCAAAGTGCTCACCATTACACCATTGGAACCTATAGTGAGTCGTATTACC-3';  
Forward primer: 5'-AGGTCCAGTGAGACTTGAAC-3'; and  
Reverse primer/T7 promoter: 5'-GGTAATACGACTCACTATAGG-3'.

### Subcellular fractionation

Cytoplasm and nuclei prepared as per Dignam et al. (1983) were purchased from Biovest International. Nuclei were resuspended in 250 mM sucrose, 10 mM MgCl<sub>2</sub>, and centrifuged through a sucrose cushion (880 mM sucrose, 0.5 mM MgCl<sub>2</sub>). Protein for Western blotting was obtained by resuspending equal cell equivalents of material in Laemmli's sample buffer and boiling (5 min, 95°C). For RNA isolation, equal cell equivalents of material were resuspended in 2× proteinase K buffer (200 mM Tris-Cl at pH 7.5; 300 mM NaCl; 25 mM EDTA; 2% (w/v) SDS), and 200  $\mu$ g proteinase K (VWR) were added. The samples were incubated at 65°C for 1 h and RNA isolated by extraction with phenol:chloroform:isoamyl alcohol (24:25:1) (pH 6.6). RNA was precipitated with 3 vol EtOH (abs) in the presence of 35  $\mu$ g glycogen as a carrier.

In order to produce cytoplasmic S-100 extract, cytoplasm was obtained as above and mixed with 0.11 vol buffer B (300 mM HEPES at pH 7.9, 1.4 M KCl, 30 mM MgCl<sub>2</sub>) and then centrifuged at 28,400 rpm (1 h, 4°C, SW41 Ti Rotor). The supernatant was dialyzed against an excess of buffer D (20 mM HEPES at pH 7.9, 20% [v/v] glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT), aliquots flash frozen in liquid nitrogen, and stored at -80°C until required.

### Gel fractionation

Twenty milligrams (1 mL) of S-100 extract were injected onto a Superdex-200 XK16/60 gel filtration column (GE Healthcare) and eluted with 0.65 mL/min buffer D. A total of 1.5 mL fractions were collected, of which RNA was precipitated from 500  $\mu$ L.

### Northern blotting

Northern blots used the enhanced sensitivity method described by Pall et al. (2007), and were probed with the RNA oligonucleotides described above after 5' end labeling with polynucleotide kinase (NEB). Hybridization was done at 37°C overnight and the blots were washed twice for at least 1 h at 37°C in 2× SSC/0.1% (w/v) SDS.

For U6, a DNA probe was synthesized from a single-stranded template using Klenow reagent (Stratagene) on the oligonucleotides described above in the presence of labeled dATP. After synthesis, the duplex was denatured at 95°C and hybridization and washing were performed at 55°C.

### Cell culture, transfection, and siRNA knockdown

Four siRNA duplexes targeting the Dicer-1 ORF (sequences as above) were mixed in a 1:1:1:1 ratio to make a final siRNA concentration of 20  $\mu$ M. A nontargeting duplex was transfected in parallel into control cells.

HeLa cells were transfected in a six-well plate format using oligofectamine (Invitrogen) following the manufacturer's instructions. Cells were harvested 48 h post-transfection, and RNA isolated using Trizol reagent (Invitrogen).

To knock down Dicer-1 in HEK 293 cells, we used HEK 293 cells stably expressing a short hairpin targeting Dicer-1 as described in Schmitter et al. (2006). Briefly, cells were cultured in the presence or absence of 1  $\mu$ g/mL doxycycline for 6 d with the media and doxycycline replaced on 2, 3, and 4 d post-induction. At 6 d post-induction, RNA was isolated using Trizol reagent (Invitrogen).

### S-100 processing assay

HeLa RNA corresponding to 70–80 nt was gel extracted and incubated in a 10  $\mu$ L final reaction volume with cytoplasmic S-100 extract (40  $\mu$ g total protein) in buffer D supplemented with 1 mM ATP and varying amounts of MgCl<sub>2</sub> at 37°C for 2 h. The cleavage products were run on a gel that was subsequently Northern blotted and probed for tRNA<sup>Gln</sup>.

### Dicer assay

A DNA template oligo containing the tRNA<sup>Gln</sup> sequence and T7 RNA polymerase promoter was amplified and made double

stranded in a standard PCR reaction (15 cycles). The PCR product was gel purified and used for T7 in vitro transcription in the presence of  $^{32}\text{P}$ -radiolabeled UTP. The RNA product was gel purified.

This RNA was incubated in the presence or absence of 1 U recombinant Dicer (Genlantis) in a 10  $\mu\text{L}$  reaction containing 4  $\mu\text{L}$  2.5 $\times$  reaction buffer, 1 mM ATP, and 2.5 mM  $\text{MgCl}_2$  at 37°C for 2 h.

### $\beta$ -Elimination

A total of 10  $\mu\text{g}$  HeLa RNA (2.5  $\mu\text{g}/\mu\text{L}$ ), isolated using Trizol reagent (Invitrogen), was resuspended in borate buffer (pH 8.6) (60 mM borax, 60 mM boric acid) and 25 mM  $\text{NaIO}_4$ . The reaction was set up in duplicate. After incubation at 22°C (10 min), the reaction was quenched by addition of 2  $\mu\text{L}$  glycerol. After a further incubation at 22°C (10 min), the RNA pellet was dried in a SpeedVac (45 min, 30°C).

One pellet was resuspended in 50  $\mu\text{L}$   $\text{H}_2\text{O}$  and immediately precipitated as below. The other was subjected to  $\beta$ -elimination by resuspension in borate buffer (pH 9.5, adjusted by addition of NaOH) and incubation (90 min, 45°C). After  $\beta$ -elimination, 150  $\mu\text{L}$  EtOH (abs) and 35  $\mu\text{g}$  glycogen were added and RNA precipitated at  $-80^\circ\text{C}$  for 1 h. The pellet was recovered by centrifugation, washed in 70% (v/v) EtOH, and resuspended in 15  $\mu\text{L}$   $\text{H}_2\text{O}$ .

### Small RNA isolation and sequencing

RNA was isolated with Trizol reagent (Invitrogen) according to the manufacturer's instructions. Small RNA libraries and sequencing was carried out as it was described (Lu et al. 2007)

### Database and genome matching

A total of 2,253,266 raw sequence reads were obtained from the instrument. The native RNA sequences were retrieved by removing the 5' "AG" and 3' "TCGTATGCCGCTCTCTGCTTGT" adaptor tags. The 3' adaptor tags were removed by exact matching to the 5' end of the adaptor sequence with a minimum matching length of four bases. Sequences containing poly-A regions tend to match very frequently to reference genomes, so the 10 reads containing  $(\text{A})_{13}$  were also removed from the data set. In addition, since errors tend to occur toward the 3' end of reads, the raw reads were clipped in order to retain only the bases which satisfied the following quality protocol. Starting from the 5' end, a minimum mean quality score of 20 was required over a sliding window of four bases with the reads clipped at the first position where the mean score was below this minimum. The clipped reads were then detagged and collated to include only reads longer than 16nt and read abundance of  $>1$ , giving 22,386 unique reads.

The collated reads were matched to annotated noncoding RNAs (ncRNAs) (Ensembl release 50) (Birney et al. 2004) and annotated mature/hairpin miRNA sequences (miRBase release 12) (Griffiths-Jones 2004). tRNA sequences are not present in the ncRNA data set available from Ensembl; thus, they were retrieved as "Simple Features" with a perl script accessing the Ensembl API. Any miRNA annotations found in the Ensembl data were removed in preference for the miRBase annotations on the basis that the miRBase data were more recent. Any reads without matches to either the miRNA or ncRNA datasets were searched against the Human genome and annotations retrieved via the Ensembl perl

API. Matching of reads to the RNA sets and the Human genome was performed with the Vmatch algorithm (Abouelhoda et al. 2004), allowing for up to two mismatches per match in all cases.

### Specific processing sites

In order to differentiate between read sequences arising from random degradation versus specific processing the following metric was applied to the ncRNA aligned data.

$$S_p = n/l \quad (1)$$

where  $S_p$  is the specific processing score,  $l$  is the maximal length of the aligned region, and  $n$  is the total abundance of sequence reads matching to the aligned region. A higher score indicates a higher presence of regions which appear to be the result of specific processing.

### SUPPLEMENTAL MATERIAL

Supplemental material can be found at <http://www.rnajournal.org>.

### ACKNOWLEDGMENTS

The work has been funded by the Wellcome Trust and European Framework 6 SIROCCO consortium fund (G.H.), and grant nos. 548569 and 0445638 from the National Science Foundation (P.J.G.). C.C. is funded by the Scottish Funding Council under the Scottish Bioinformatics Research Network (SBRN) grant. G.H. is a Wellcome Trust CD fellow. A.S. is funded by a BBSRC Doctoral Training Grant. S.R.T. was supported in part by National Institutes of Health Training Grant T32 GM-08550. We acknowledge Dr. Thomas Walsh for the technical support of the College of Life Sciences High Performance Computing service, and Dr. David Martin for constructive discussions. We are also very grateful to Dr. Witold Filipowicz for providing the inducible Dicer knock down cells.

Received May 18, 2009; accepted September 3, 2009.

### REFERENCES

- Abouelhoda MI, Kurtz S, Ohlebusch E. 2004. Replacing suffix trees with enhanced suffix trees. *J Discrete Algorithms* 2: 53–86.
- Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 5: 337–350.
- Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC. 2008. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci* 105: 7964–7969.
- Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs. *Genes & Dev* 22: 2773–2785.
- Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. 2007. Mammalian mirtron genes. *Mol Cell* 28: 328–336.
- Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V, Cutts T, et al. 2004. An overview of Ensembl. *Genome Res* 14: 925–928.
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089–1103.

- Buhler M, Spies N, Bartel DP, Moazed D. 2008. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. *Nat Struct Mol Biol* **15**: 1015–1023.
- Calabrese JM, Seila AC, Yeo GW, Sharp PA. 2007. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci* **104**: 18097–18102.
- Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. *Bioinformatics* **20**: 426–427.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14**: 1188–1190.
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**: 798–802.
- Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**: 1475–1489.
- Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. 2008. A human snoRNA with microRNA-like functions. *Mol Cell* **32**: 519–528.
- Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. 2007. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**: 287–297.
- Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320**: 1077–1081.
- Griffiths-Jones S. 2004. The microRNA registry. *Nucleic Acids Res* **32**: D109–D111.
- Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: Annotating noncoding RNAs in complete genomes. *Nucleic Acids Res* **33**: D121–D124.
- Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**: 23–34.
- Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**: 1587–1590.
- Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Dev* **18**: 3016–3027.
- Hock J, Meister G. 2008. The Argonaute protein family. *Genome Biol* **9**: 210. doi: 10.1186/gb-2008-9-2-210.
- Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.
- Hutvagner G, Simard MJ. 2008. Argonaute proteins: Key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**: 22–32.
- Jochl C, Rederstorff M, Hertel J, Stadler PF, Hofacker IL, Schrettl M, Haas H, Huttenhofer A. 2008. Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Res* **36**: 2677–2689.
- Kawaji H, Nakamura M, Takahashi Y, Sandelin A, Katayama S, Fukuda S, Daub CO, Kai C, Kawai J, Yasuda J, et al. 2008. Hidden layers of human small RNAs. *BMC Genomics* **9**: 157. doi: 10.1186/1471-12164-9-157.
- Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793–797.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415–419.
- Levinger L, Bourne R, Kolla S, Cylind E, Russell K, Wang X, Mohan A. 1998. Matrices of paired substitutions show the effects of tRNA D/T loop sequence on *Drosophila* RNase P and 3'-tRNase processing. *J Biol Chem* **273**: 1015–1025.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.
- Lu C, Meyers BC, Green PJ. 2007. Construction of small RNA cDNA libraries for deep sequencing. *Methods* **43**: 110–117.
- Madore E, Florentz C, Giege R, Lapointe J. 1999. Magnesium-dependent alternative foldings of active and inactive *Escherichia coli* tRNA<sup>Glu</sup> revealed by chemical probing. *Nucleic Acids Res* **27**: 3583–3588.
- Marshall L, Kenneth NS, White RJ. 2008. Elevated tRNA<sup>Met</sup> synthesis can drive cell proliferation and oncogenic transformation. *Cell* **133**: 78–89.
- Martinez J, Patkaniowska A, Urlaub H, Luhrmann R, Tuschl T. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**: 563–574.
- Matranga C, Zamore PD. 2007. Small silencing RNAs. *Curr Biol* **17**: R789–R793.
- Okamura K, Balla S, Martin R, Liu N, Lai EC. 2008a. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* **15**: 998. doi: 10.1038/nsmb0908-998.
- Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. 2008b. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**: 803–806.
- Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**: 89–100.
- Okamura K, Lai EC. 2008. Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* **9**: 673–678.
- Pall GS, Codony-Servat C, Byrne J, Ritchie L, Hamilton A. 2007. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by Northern blot. *Nucleic Acids Res* **35**: e60. doi: 10.1093/nar/gkm112.
- Politz JC, Zhang F, Pederson T. 2006. MicroRNA-206 colocalizes with ribosome-rich regions in both the nucleolus and cytoplasm of rat myogenic cells. *Proc Natl Acad Sci* **103**: 18957–18962.
- Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**: 83–86.
- Schmitter D, Filkowski J, Sewer A, Pillai RS, Oakeley EJ, Zavolan M, Svoboda P, Filipowicz W. 2006. Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res* **34**: 4801–4815.
- Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534–538.
- Thompson DM, Lu C, Green PJ, Parker R. 2008. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14**: 2095–2103.
- Tomari Y, Du T, Zamore PD. 2007. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**: 299–308.
- Watanabe T, Imai H, Minami N. 2008a. Identification and expression analysis of small RNAs during development. *Methods Mol Biol* **442**: 173–185.
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008b. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539–543.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.
- Yi R, Qin Y, Macara IG, Cullen BR. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Dev* **17**: 3011–3016.
- Zhang H, Kolb FA, Brondani V, Billy E, Filipowicz W. 2002. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J* **21**: 5875–5885.



# RNA

A PUBLICATION OF THE RNA SOCIETY

## Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs

Christian Cole, Andrew Sobala, Cheng Lu, et al.

RNA 2009 15: 2147-2160 originally published online October 22, 2009  
Access the most recent version at doi:[10.1261/rna.1738409](https://doi.org/10.1261/rna.1738409)

---

### Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2009/09/29/rna.1738409.DC1>

### References

This article cites 51 articles, 14 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/15/12/2147.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>