

Promiscuous splicing-derived hairpins are dominant substrates of tailing-mediated defense of miRNA biogenesis in mammals

Seungjae Lee, David Jee, Sid Srivastava, Acong Yang, Abhinav Ramidi, Renfu Shang, Diane Bortolamiol-Becet, Sebastien Pfeffer, Shuo Gu, Jiayu Wen, et al.

▶ To cite this version:

Seungjae Lee, David Jee, Sid Srivastava, Acong Yang, Abhinav Ramidi, et al.. Promiscuous splicingderived hairpins are dominant substrates of tailing-mediated defense of miRNA biogenesis in mammals. Cell Reports, 2023, 42 (2), pp.112111. 10.1016/j.celrep.2023.112111. hal-04035525

HAL Id: hal-04035525 https://hal.science/hal-04035525

Submitted on 17 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cell Reports

Promiscuous splicing-derived hairpins are dominant substrates of tailing-mediated defense of miRNA biogenesis in mammals

Authors

Eric C. Lai

Seungjae Lee, David Jee,

Correspondence

Graphical abstract



Highlights

- Mirtrons are splicing-derived hairpins that mimic pre-miRNAs for small RNA biogenesis
- Structured splicing-derived RNAs (ssdRNAs) associate with Dicer and Ago as hairpins
- ssdRNA and mirtron hairpins are dominant substrates for **TENT/TUT** enzymes
- Tailing defends the canonical miRNA pathway from fortuitous splicing-derived hairpins

jiayu.wen@anu.edu.au (J.W.), laie@mskcc.org (E.C.L.)

In brief Lee et al. find hundreds of intermediatelength, structured, splicing-derived RNAs (ssdRNAs) in human cells. ssdRNAs are related to mirtrons but do not generate substantial small RNAs; still, these hairpins can associate with the miRNA factors Dicer and Argonaute. Their collective existence contaminates the canonical miRNA pathway, provoking

3'-tailing defense via TENT2/TUT4/TUT7.

Sid Srivastava, ..., Shuo Gu, Jiayu Wen,



Cell Reports

Article

Promiscuous splicing-derived hairpins are dominant substrates of tailing-mediated defense of miRNA biogenesis in mammals

Seungjae Lee,¹ David Jee,^{1,2} Sid Srivastava,^{1,3} Acong Yang,⁴ Abhinav Ramidi,^{1,3} Renfu Shang,¹

Diane Bortolamiol-Becet,^{1,5} Sébastien Pfeffer,⁵ Shuo Gu,⁴ Jiayu Wen,^{6,*} and Eric C. Lai^{1,2,7,*}

¹Developmental Biology Program, Sloan-Kettering Institute, 1275 York Avenue, Box 252, New York, NY 10065, USA

²Weill Graduate School of Medical Sciences, Weill Cornell Medical College, New York, NY 10065, USA

³High Technology High School, Lincroft, NJ 07738, USA

⁴RNA Biology Laboratory, Center for Cancer Research, 8 National Cancer Institute, Frederick, MD 21702, USA

⁵Université de Strasbourg, Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du CNRS, 2 Allée Konrad Roentgen, 67084 Strasbourg, France

⁶Division of Genome Sciences and Cancer, The John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

7Lead contact

*Correspondence: jiayu.wen@anu.edu.au (J.W.), laie@mskcc.org (E.C.L.) https://doi.org/10.1016/j.celrep.2023.112111

SUMMARY

Canonical microRNA (miRNA) hairpins are processed by the RNase III enzymes Drosha and Dicer into \sim 22 nt RNAs loaded into an Argonaute (Ago) effector. In addition, splicing generates numerous intronic hairpins that bypass Drosha (mirtrons) to yield mature miRNAs. Here, we identify hundreds of previously unannotated, splicing-derived hairpins in intermediate-length (\sim 50–100 nt) but not small (20–30 nt) RNA data. Since we originally defined mirtrons from small RNA duplexes, we term this larger set as structured splicing-derived RNAs (ssdRNAs). These associate with Dicer and/or Ago complexes, but generally accumulate modestly and are poorly conserved. We propose they contaminate the canonical miRNA pathway, which consequently requires defense against the siege of splicing-derived substrates. Accordingly, ssdRNAs/mirtrons comprise dominant hairpin substrates for 3' tailing by multiple terminal nucleotidyltransferases, notably TUT4/7 and TENT2. Overall, the rampant proliferation of young mammalian mirtrons/ssdRNAs, coupled with an inhibitory molecular defense, comprises a Red Queen's race of intragenomic conflict.

INTRODUCTION

MicroRNAs (miRNAs) comprise an extensive class of short regulatory RNAs, which collectively mediate broad impacts on gene regulation. Their biogenesis can proceed via diverse mechanistic strategies. In animals, the canonical pathway generates the vast majority of abundant and functional miRNAs.^{1,2} In this scheme, primary miRNA (pri-miRNA) transcripts that bear appropriate hairpin structures are recognized and cleaved by the Microprocessor complex, composed of the RNase III enzyme Drosha and its double-stranded RNA binding domain (dsRBD) partner DGCR8. Since the number of plausible primiRNA hairpins in the transcriptome exceeds the number of expressed miRNAs by at least 2-3 orders of magnitude, accurate and specific selection of hairpins by Microprocessor serves as a gatekeeper for substrate entry to the miRNA pathway. Following their cleavage into pre-miRNA hairpins, these are exported to the cytoplasm and cleaved near their terminal loops by the RNase III enzyme Dicer. The resultant duplexes are loaded into Argonaute (Ago) effector proteins, and one of the strands is either ejected or cleaved to yield a single-stranded

effector complex programmed with a mature miRNA. Following association with the GW182 (TNRC6) adaptor protein, the Ago complex is guided by the miRNA sequence to appropriate regulatory targets.

In animals, short sequence complementarity, typically Watson-Crick pairing to nt 2-8 of the miRNA, is required for target recognition and functional repression.³⁻⁶ Accordingly, miRNAs can in principle regulate large cohorts of targets. Individual conserved mammalian miRNAs are typically associated with hundreds, and over a thousand in some cases, evolutionarily conserved target sites.⁷ Moreover, genomic assays indicate that individual miRNAs can directly repress numerous conserved and even non-conserved targets, although the magnitude of miRNA-mediated regulation is usually modest.^{6,8} Because of the broad regulatory impact of individual miRNAs, the small RNA contents of Ago proteins must be carefully controlled. Although >1,000 canonical miRNA loci have been annotated in the human genome,⁹ typical cells and tissues express a limited repertoire of miRNAs at levels expected to confer substantial regulation.^{10–14} Although our capacity to visualize the contents of Ago proteins continues to increase, with ever-greater



throughput of next-generation sequencing, the majority of miRNA reads in a given library can often be accounted for by a few dozens of loci. Indeed, it has been proposed that additional *bona fide* human miRNAs are unlikely to remain.¹⁵

In addition to the canonical miRNA biogenesis pathway, many non-canonical routes exist that can populate Ago proteins with small RNAs.¹⁶ The first recognized pathway comprises splicing-dependent "mirtrons," whose biogenesis is consequently independent of Drosha/DGCR8 (Figure 1). In the original description of mirtron biogenesis, splicing and debranching of an intron directly defines both ends of the pre-miRNA hairpin.^{17–19} Subsequently, we recognized variants of mirtron biogenesis, in which only one end of the pre-miRNA hairpin coincides with a splice site.²⁰ In the case of 3'-tailed mirtrons, evidence from Drosophila indicates that the RNA exosome trims the unstructured 3' end to reveal the pre-miRNA hairpin.²¹ 5'-tailed mirtrons have also been documented and are especially abundant in mammals,²² but the nuclease(s) involved in 5'-tail removal has not yet been defined. The dominant class of splicing-derived miRNAs in mammals comprises 5'-tailed mirtrons, followed by conventional mirtrons and then 3'-tailed mirtrons.²³ There are also rare "dual-tailed mirtrons" in which neither pre-miRNA hairpin terminus seems to be defined directly by splicing, but where one end resides too close to the splice site to represent a plausible Microprocessor substrate.^{23,24}

Beyond mirtrons, a wide variety of other non-canonical miRNA substrates bypass either Microprocessor or Dicer.¹⁶ Curiously, very few non-canonical miRNA loci are conserved, with many alternative pathways supporting the biogenesis of just one or a couple of conserved miRNAs. None of these are shared across C. elegans. Drosophila, or vertebrates, in stark contrast to the dozens of pan-metazoan miRNA families that are identical (or nearly so) between invertebrates and vertebrates. This is not to say that non-canonical miRNA loci are rare, guite the contrary. This is most evident for various classes of mirtrons. These collectively total at least 500 in both the mouse and the human genomes, although <10 are conserved among mammals.²³ Thus, there are distinct evolutionary trajectories for different classes of miRNA substrates, even though their ultimate activities to program Ago complexes are seemingly identical. Accordingly, we speculated that there should be a biochemical basis to distinguish canonical pre-miRNAs from mirtron hairpins, leading to preferential suppression and evolutionary extinction of the latter.

We and others indeed provided evidence for this model in *Drosophila*. The observation that mirtron-3p (but not mirtron-5p or most canonical miRNA-5p/3p) species exhibit unusually high frequencies of untemplated uridines permitted the isolation of Tailor, a terminal uridyltransferase (TUTase) that preferentially modifies conventional mirtron hairpins.^{25,26} Tailor exhibits selectivity by recognizing the 3'-terminal AG nucleotides that are a signature of the 3' splice site, and its activity reduces mirtron dicing and regulatory activity. Of note, most *Drosophila* mirtrons are of the conventional variety. Mammals have a distinct issue, since they innovated the 5'-tailed class, which accounts for several hundreds of loci in individual mammalian genomes.^{22,23} Moreover, although mammals have an expanded family of terminal nucleotidyltransferase (TENT) enzymes compared with *Drosophila*, they evidently lack a clear ortholog of Tailor.²⁷

Cell Reports Article

Thus, while several TENTs (and specifically TUTases) regulate diverse aspects of mammalian miRNA biogenesis, ^{1,28,29} it was unclear which mammalian TENTs, if any, control mirtrons.

Here, we show that conventional and 5'-tailed mirtrons are broadly subject to extremely high levels of tailing on mirtron-3p reads (mostly uridylation and some adenylation), at frequencies far higher than for canonical miRNAs. We broaden this scope by showing that hundreds of additional intronic hairpins, which reside preferentially at 3' intron termini, are similarly specifically and highly tailed. We term this larger class as structured splicing-derived RNA (ssdRNAs) and find broad evidence for their capacity to associate with Dicer and Ago proteins as intermediate-sized hairpin species. Based on the inference that tailing represents a defense strategy against adventitious premiRNA mimics, we identify multiple TENT enzymes that participate in ssdRNA/mirtron tailing: TUT4, TUT7, and TENT2 (also known as TUT2 or Gld2). Misexpression of these TENT enzymes antagonizes ssdRNAs/mirtrons, while their combinatorial loss results in biased upregulation of mirtron-derived small RNAs. The functional outcomes of tailing loss are also complex and can in part be attributed to the heterogeneity of ssdRNA structures. However, a clear impact of their action is observed in the depletion of conserved canonical pre-miRNAs that resemble 3' intron termini. Overall, we find that intronic hairpins represent a dominant target of TUT/TENT enzymes in mammalian cells, as part of a quality control strategy to suppress the capacity of genomically abundant splicing by-products from contaminating the canonical miRNA pathway.

RESULTS

Identification of numerous human intronic hairpin RNAs that lack corresponding small RNAs

Since the advent of miRNA annotation using deep sequencing data, there have been efforts to encourage robust and stringent criteria.³⁰ However, as a community effort, different groups utilized varying levels of evidence. Notably, numerous published loci deposited in the miRBase registry (https://www.mirbase. org/) have not been validated in experimental tests.^{15,31,32} One potential caveat regards the lack of clear cutoffs on how efficient biogenesis should be to qualify as a bona fide miRNA. It is evident that numerous guestionable miRNA annotations are simply RNA degradation fragments that fortuitously map to short hairpins and do not derive from a specific biogenesis pathway.³³ On the other hand, it is also not uncommon for miRNA biogenesis to be regulated and, thus, may not necessarily be efficient in a given assay context.^{34–36} If only a small minority of an input primary transcript is eventually converted into Ago-bound small RNAs, does that locus qualify as an miRNA?

The picture is potentially more complicated with non-canonical miRNA substrates, whose biogenesis does not follow typical conventions. We and others previously annotated mirtrons as intron-terminal hairpins that yield confident and specific miRNA/star duplexes, indicative of endogenous Dicer processing, and these exist in multiple invertebrate^{18,19,37} and vertebrate^{17,22,23} species. These annotations relied on specific processing as judged from meta-analysis of publicly available small-RNA data (Table S1), but made no implicit conclusions





Figure 1. Multiple classes of splicingderived intermediate-length and short RNAs in human cells

Splicing can generate a variety of pre-miRNA mimics ("mirtrons") that bypass the Drosha RNase III enzyme, which is obligate for cleavage of canonical pre-miRNAs. Three types of splicingderived pre-miRNAs are clearly a direct consequence of splicing, whereby both hairpin ends (conventional mirtron) or one hairpin end (as in 5'-tailed or 3'-tailed mirtrons) lie at splice sites. For tailed mirtrons, additional nucleases are inferred to reduce the debranched intron into a pre-miRNAlike hairpin. In mammals, 5'-tailed mirtrons are by far the most abundant subtype, followed by conventional mirtrons and rare 3'-tailed mirtrons; rare intronic hairpins are inferred to undergo tail removal from both termini (two-tailed). Historically, mirtrons were annotated based on their capacity

to generate small RNA duplexes with 3' overhangs (indicative of Dicer cleavage) and/or Ago-associated small RNAs (i.e., mature miRNAs). In this study, we now recognize a substantial population of structured, splicing-derived RNAs (ssdRNAs). These lack compelling small RNA evidence, but exhibit other hallmarks of mirtron hairpins. We document ssdRNAs in association with Dicer and/or Ago2, but they may also potentially exist in a free hairpin state.

regarding efficiency of biogenesis. Kjems and colleagues also described "Agotrons," intronic hairpins that can also associate with Ago proteins and guide target silencing.³⁸ We noticed that some of the best-expressed Agotron hairpins were among the first annotated mammalian small-RNA-generating mirtrons (e.g., mir-1225/Pkd1).¹⁷ Further inspection showed that the strong majority of Agotrons were loci we had annotated as conventional or 5'-tailed mirtron hairpins (Table S2) and thus generated at least some duplex small RNAs indicative of dicing.^{22,23} Only a minority of loci were not present in our scans, and many of these were very scarcely detected or did not exhibit positional read enrichment relative to flanks (either the remainder of introns or in flanking exons; Table S2). Thus, it is reasonable to consider the previously described Agotrons as mirtron hairpins that are inefficiently diced, but can still associate with Ago proteins.

Importantly, there is no reason to presume that mirtron hairpins are, as a class, robust Dicer substrates. The vast majority of mirtrons are evolutionarily young and thus not expected to be selected for optimal structural features of canonical premiRNAs that generate abundant small RNAs.39,40 Moreover, the collective hairpin features of mirtrons are explicitly known to be more heterogeneous compared with canonical premiRNAs.^{23,37} Interestingly, some mirtrons with very-well-defined dicing patterns, even with phased terminal loop reads, generate abundant intermediate-length reads in Ago2 (e.g., mir-1226; Figure 2A). Although read abundances between such libraries are not directly comparable, they suggest that at least some annotated mirtrons are not effectively diced. Relevant to this, other hairpins that are explicitly Dicer independent, such as pre-mir-451⁴¹⁻⁴³ or its engineered derivatives,⁴⁴⁻⁴⁶ load directly into Ago proteins. In addition, canonical pre-miRNA hairpins can associate with Ago proteins, especially under certain aberrant conditions.^{47,48} Overall, these considerations made us question if it was overly restrictive to define mirtrons based on Dicer signatures of small-RNA duplexes, as is normally the case when performing miRNA annotation.

We sought to identify additional splicing-derived, pre-miRNAlike hairpins, using deep sequencing data of intermediate-length RNA species. Currently, library construction of intermediatelength RNAs (which should contain pre-miRNA hairpins) is known to be inefficient and highly biased, due to challenges in linker ligation to structured RNAs. Consequently, very few such datasets are extant, in contrast to 10,000s of published small-RNA sequencing datasets. We focused on three such datasets of \sim 50–100 nt species from human samples: (1) HEK293T Ago2-IP samples from the Brian Gregory lab,⁴⁹ (2) total normalized HeLa samples from the Burroughs group,⁵⁰ and (3) a newly generated dataset from K562 cell Ago2-IP material (see STAR Methods). Inspection of the vicinity of intron boundaries revealed numerous loci with intermediate-sized reads, but few or no corresponding small-RNA reads. Accordingly, these escaped our previous annotations of mirtrons from hundreds of small-RNA datasets.²³ For example. the 27th intron of the recently recognized miRNA pathway factor Ankrd52⁵¹ yielded >10,000 hairpin reads, but no mature miRNAs, in the Gregory lab paired Ago2-IP datasets (Figure 2B). The Ankrd52 hairpin resides precisely at the 3' end of the intron, with an inferred 5' tail of several hundred nucleotides. Although corresponding small RNAs were not present in the paired HEK293T Ago2-IP dataset, a low level of Ankrd52 small RNAs from the hairpin arms was detected (Figure 2B) using other aggregated small-RNA datasets (Table S1). Thus, it is subject to a very modest level of dicing, but the predominant species seem to be hairpins.

Loci such as these encouraged systematic analysis of mammalian intermediate-sized reads in the vicinity of splice sites. Applying stringent criteria, we identified 273 newly recognized human intron termini that generate >30 intermediate-sized RNAs with pre-miRNA-like structure (Table S3). We provide additional compelling examples of such loci in Figure S1, emphasizing that these loci generally yielded numerous intermediate RNAs but few or no small RNAs in the paired Ago2-IP datasets. These mirrored the subcategories of known mammalian mirtrons,²³ in that most were 5'-tailed hairpins followed by whole-intron hairpins, with a small subset of 3'-tailed hairpins.





A Previously annotated mirtron (*mir-1226*)

exon		intron	exon	Gregory	HEK AGO-IP
GTCGACCATTAACAGGTGAGGGCATGCAGGCCTGGATGGGC	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	GGAGGCCACACGGTTACGGAGCCGAT	size on	g short
	$(\cdot(((((((())))))))))))))))))))))))))$))))))))))))))))))))))))))))))))))	<u>.</u>	(nt) read	Ís reads
GTGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>t</mark>	76 2310	52 -
GTGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>a</mark>	76 143	77 -
GGTGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>t</mark>	77 24	77 -
GGTGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>a</mark>	77 118	30 -
	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>t</mark>	75 84	19 -
GTGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>e</mark>	76 32	24 -
GTGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>.</mark>	75 29	94 -
TGAGGGCATGCAGGCCTGGATGGGG	CAGCTGGGATGGTCCAAAAGGGTGGCC	TCACCAGCCCTGTGTTCCCTAG	<mark>a</mark>	75 2	79 -
(>120 distinct species >	50 nts in length and >40 distinct spe	cies <30 nt in length)			
selected small RNA reads:					
		TCACCAGCCCTGTGTTCCCTAG	<mark>t</mark>	23 -	373
		TCACCAGCCCTGTGTTCCCTAG	<mark>a</mark>	23 -	140
•••••••••••••••••••••••••••••••••••••••		TCACCAGCCCTGTGTTCCCTAG		23 -	54
		TCACCAGCCCTGTGTTCCCTAG	<mark>e</mark>	23 -	1
GTGAGGGCATGCAGGCCTGGATGGGG			<u>.</u>	26	ı -
GTGAGGGCATGCAGGCCTGGATGGG.				25	I -
GTGAGGGCATGCAGGCCTGGATG				23	I -
· · · · · · · · · · · · · · · · · · ·	CAGCTGGGATGGTCCAAAAGGGTGGCC			27 (other	libs, eq DroshaTN

B Newly annotated structured, splicing-derived RNA (ssdRNA)



Figure 2. Exemplar mirtron and ssdRNA loci

 (A) A previously annotated mirtron (*mir-1226*) that is a definitive Dicer substrate, as evidenced by phased miRNA-5p, miRNA-3p, and terminal loop reads. However, paired intermediate-length and small RNA Ago2-IP data from HEK293T cells show that *mir-1226* is predominantly uncleaved.
(B) Example of a newly annotated ssdRNA derived from an intron of the miRNA factor *Ankrd52*. This locus generated abundant hairpin reads, but no small RNAs were detected in matched small and intermediate-sized Ago2-IP data from HEK293T cells (bottom), although inspection of 756 aggregated small-RNA datasets reveals modest small-RNA production from its hairpin arms (UCSC genome browser tracks).

We emphasize that, beyond these confident annotations, additional intermediate-sized loci were plausible candidates but fell below read cutoffs, had overly heterogeneous termini, or had suboptimal structures. As these current annotations derived from only three cell-line datasets, additional loci may prove confident with broader cell/tissue profiling.

The Gregory HEK293T intermediate RNA library was especially useful for these annotations, and many loci were detected only in this dataset (Table S3). However, we caution against inferring this to reflect cell specificity, as the efficiency of pre-miRNA cloning strategies and their individual biases likely vary widely across datasets. For example, the Burroughs HeLa data are from total RNA and not Ago2-IP, our K562 data may have less efficient hairpin cloning, or perhaps the Gregory data may exhibit PCR skew. Nevertheless, many of these newly recognized intronic hairpin loci were co-detected by long reads in two or more of the datasets, indicating their breadth. For example, of loci confidently annotated from HEK293T, HeLa data share 64 loci with \geq 2 hairpin reads and 21 with \geq 10 hairpin reads, while K562 expresses 139 loci with \geq 2 hairpin reads and 56 with \geq 10 hairpin reads (Table S3). In addition, six known mirtrons were present in HeLa and/or K562 data, but not HEK293, and four ssdRNA loci were annotatable only from the HeLa data (none were exclusive to K562). One notable example of this was an intron of *MFSD3* (uc003zdi.1_intron_3), which had >1,000 hairpin reads in HeLa, but none in HEK293T data (Table S3). Overall, while it may not be appropriate to directly compare read numbers across these libraries, we conclude that long hairpin species emanate from numerous intron termini across multiple cell lines.





(legend on next page)



Mammalian mirtrons belong to a larger class of ssdRNAs

To provide a basis for understanding the properties of these newly recognized loci, we first examined 480 previously annotated mirtrons across a range of informative datasets.²³ Using the paired datasets of 50-100 and 20-25 nt RNAs from HEK293T Ago2-IP samples,⁴⁹ we observed that previously annotated mirtrons at 3' intron boundaries (i.e., conventional and 5'-tailed mirtrons) were supported by substantial intermediatesized- and small-RNA reads. Because we previously annotated mirtrons from numerous small-RNA datasets, we also evaluated their coverage across aggregated data (Table S1). As expected, the vast majority yielded small RNAs at a stringent cutoff (>10 reads per million mapped reads [RPM]) and exhibited paired read pileups that reflect 5p/3p duplexes (Figures 3A and 3B). We also analyzed 3'-tailed mirtrons and observed similar but slightly less robust patterns, owing to their smaller number and generally lower expression (Figure S2).

Our newly annotated intronic RNA hairpins also exhibit similarly characteristic size and distribution. When we anchored on the 3' ends of introns, the sharp peak of structured RNA mappings was observed, reflecting a dominant population of intronic hairpins with 5' tails (Figure 3C). As expected from our annotation pipeline, these produce very few small RNAs. We considered if such local enrichment might be a by-product of anchoring on a defined genomic feature, perhaps reflecting intron degradation. However, we can rule out this artifact, given the sharp "skyscraper" shape of intron-terminal intermediate-length RNAs, with low levels of upstream intronic reads. Moreover, the adjacent exons have a much lower level of mapped reads, even though mRNAs are far more abundant than introns, and these lack specific local enrichment (Figure 3C). Similar to the distribution of annotated mirtrons, a smaller set of introns yielded specific hairpin species from their 5' ends, which are also far more abundant than flanking exonic reads but do not yield comparable short RNAs (Figure S2).

We broadened the analysis of intron-terminal hairpins defined from intermediate-length-RNA data to other aggregated small-RNA data (Table S1). We observed a local peak of small RNAs at intron termini, which coincides with the intermediate-sized-RNA peak but also exhibits a central dip, characteristic of 5p/ 3p miRNA duplexes (Figure 3D). However, the abundance of

Cell Reports Article

short RNAs mapped to intron hairpins was only marginally greater than that of heterogeneous short RNAs mapped to adjacent exons, which were presumably mRNA degradation products. This emphasizes that, while terminal intronic hairpins are a genomically broad source of intermediate-length RNAs, on the whole, they are modest Dicer substrates at best.

The long-standing concept of mirtrons is that they are splicingderived hairpins that are diced into mature miRNA-class regulatory species.^{18,19,23} Since most of the newly annotated loci are not detectably converted into small RNAs, we now propose ssdRNAs as a broader parent class that includes mirtrons, except that there is no overall requirement for them to be Dicer substrates (Figure 1). Overall, ssdRNAs comprise an extensive class of non-coding RNAs detected in mammalian cells that predominantly arise from 3' intron termini but are also detected at 5' intron termini. With this foundation, we turned our attention to their functional properties.

ssdRNAs/mirtrons are broadly and specifically accessible to the core miRNA factors Dicer and Ago

Synthetic mirtrons are a viable strategy to generate functional silencing RNAs in invertebrates⁵² and mammals.^{53–55} Nevertheless, as endogenous mirtrons typically generate only modest amounts of small RNA (Figures 3A and 3B), and are mostly not conserved,^{23,56} this raises questions as to their relevance to gene regulation. In fact, miRNAs must typically achieve substantial abundance to exert meaningful regulatory impacts, in part due to a plethora of low-affinity fortuitous sites throughout the transcriptome.² With this perspective, we assessed the relationship of known mirtrons and our newly recognized ssdRNAs with miRNA factors.

We first analyzed two Dicer photoactivatable ribonucleosideenhanced crosslinking and immunoprecipitation (PAR-CLIP) datasets from HEK293T.^{57,58} Although Dicer and Ago proteins form complexes, the stringency of CLIP protocols is presumed to reduce indirect interactions. Although mirtrons were not previously reported to associate with Dicer, metagene plots show positional enrichment that matches well to the annotated hairpins (Figure 3E). Strikingly, even though ssdRNAs are by definition poorly diced, we observe a comparable local spike of Dicer occupancy precisely on ssdRNA regions (Figure 3G). For both classes

```
Figure 3. Partner proteins and untemplated modifications of ssdRNAs/mirtrons
```

⁽A) Positional properties of 480 previously annotated human mirtrons show they are associated with specific peaks at 3'-intron termini in intermediate-length and small RNA Ago2-IP data from HEK293T cells.

⁽B) Meta-analysis of known mirtrons across hundreds of small-RNA datasets shows a characteristic "double peak" reflecting 5p/3p duplexes.

⁽C) Positional properties of 270 newly recognized ssdRNAs in HEK293T Ago2-IP library show they generate specific 3'-intron terminal intermediate-sized species with few small RNAs or adjacent exonic RNAs (i.e., conventional ssdRNA or 5'-tailed ssdRNA).

⁽D) Meta-analysis of ssdRNAs across hundreds of small-RNA datasets shows a signature of Dicer cleavage. However, these small-RNA reads are marginally above the background of degraded RNAs from flanking exons.

⁽E-H) Both mirtrons (E and F) and ssdRNAs (G and H) are substantially associated with core miRNA factors Dicer (E and G) and Ago2 (F and H).

⁽I–L) Neither mirtrons nor ssdRNAs associate with Microprocessor (Drosha or DGCR8; I and K) or the non-miRNA factor Pum2 (J and L). The average signal for each group (line) and the associated 95% confident interval (CI) are represented in (A–L).

⁽M) CDF plots of minimum free energies of canonical pre-miRNAs, mirtron hairpins, and newly recognized ssdRNAs compared with a background set of intron termini with matched nucleotide content. All three classes of miRNA/miRNA-like species have similar minimum free energy (MFE) profiles and are well separated from control introns.

⁽N and O) Untemplated single-nucleotide additions to 3' ends of canonical miRNAs, mirtrons, and ssdRNAs in paired intermediate-length RNA and short RNA Ago2-IP data from HEK293T cells. Only a subset of mirtrons/ssdRNAs could be analyzed, since we analyzed only those reads that extend to the 3' splice site (-AG) or carry 3' untemplated additions; a number of loci had 3' trimming, and those reads were not relevant for this analysis. Error bar represents SE.

of intronic RNAs, Dicer occupancy level plummets in the flanking downstream exons and is at a background level in the upstream flanking introns, indicating its specificity for ssdRNA/mirtron hairpins. We next aggregated Ago CLIP-seq and PAR-CLIP data from HEK293T cells^{59–61} and observed similar patterns of positional enrichment on mirtrons (Figure 3F) and ssdRNAs (Figure 3H). In addition, the rarer classes of mirtrons and ssdRNAs that reside at 5' intron termini were similarly detectable in both Dicer and Ago2 CLIP-seq data (Figure S2A–S2G).

Overall, about half of known mirtrons and newly recognized ssdRNAs bound Dicer, and 80% of mirtrons/50% of ssdRNAs were associated with Ago, at levels >1 read/10 M. Although this is a modest level, no such positional enrichment on ssdRNAs was observed in CLIP-seq datasets of other core miRNA factors such as Drosha⁶² or DGCR8⁶³ or a distinct RBP regulator, Pum2⁶⁴ (Figures 3I–3L). We note that DGCR8 was substantially enriched precisely at canonical pre-miRNAs, but Drosha was not (Figure S2H). This may reflect that DGCR8 is technically easier to capture in association with pri-miRNA transcripts than the nuclease Drosha. Regardless, this demonstrates that lack of DGCR8 binding at mirtrons and ssdRNAs is a genuine negative result and reciprocally indicates that the presence of Dicer binding at these intronic hairpins is also a specific positive result. We conclude that ssdRNAs/mirtrons frequently adopt sufficient structural similarity to canonical pre-miRNAs to enable their specific association with Dicer and/or Ago proteins. To solidify this conclusion, we plotted the cumulative distribution function (CDF) of minimum free energies predicted for canonical pre-miRNAs, previously annotated mirtrons, and our newly annotated set of ssdRNAs compared with a matched set of control intron termini with comparable sequence content. These analyses showed that all three classes of hairpin RNAs comprise similar ranges of low free energies and are highly separated from control intron termini (Figure 3M). This conclusion is bolstered by the fact that most ssdRNAs exhibit typical premiRNA-like hairpin structures (predicted secondary structures of 42 well-expressed ssdRNAs are depicted in Figure S3).

Of note, we previously reported that aggregate structural features of mirtron hairpins are much more heterogeneous than for canonical pre-miRNAs and often include features that are likely deleterious for Dicer cleavage.^{23,37} Accordingly, the presumably suboptimal pre-miRNA features of mirtron and ssdRNA hairpins may underlie why they can associate with Dicer, but be inefficiently cleaved, and/or associate with Ago effectors that would normally prefer to load with small-RNA duplexes.

ssdRNA/mirtron-3p species are the predominant class of tailed miRNA species in mammals

In *Drosophila*, mirtrons are sensed by the TUTase Tailor, which preferentially uridylates their 3' hairpin ends and opposes their maturation.^{25,26} Individual mammalian genomes such as human and mouse both have more than an order of magnitude more mirtrons than does *Drosophila*.²³ Moreover, we have substantially increased the scope of splicing-derived small RNAs, by now revealing hundreds of ssdRNAs, a parent class of mirtrons that are minimally diced. Altogether, the sheer genomic numbers of ssdRNA/mirtron loci suggest they collectively present a greater threat to the canonical miRNA in mammals compared



with invertebrates. How have mammalian genomes countered this menace? Indeed, one may reasonably conclude that present-day mammals do not cope with mirtrons as well as invertebrates do. This may in part be due to the exuberant proliferation of a biogenesis subclass (5'-tailed mirtron; Figure 1) that has not been detected in invertebrates.^{22,37}

A number of canonical miRNAs are modified by various TENT enzymes,^{1,29} which collectively may induce modifications at hairpin, duplex, or single-stranded stages. Notably, preferential uridylation of miRNA reads deriving from the 3p arm of their pre-miRNAs has been documented,65-67 suggesting a biased pattern of pre-miRNA modifications. Nevertheless, in the vast majority of cases, tailed mature species comprise only a few percent of unmodified reads.⁶⁸ Among archetypal cases, premiRNA tailing can have diverse outcomes. For example, oligouridylation of group I pre-let-7 members inhibits their biogenesis,^{69–71} while monouridylation of group II pre-let-7 members by TENT2/TUT4/TUT7 promotes their dicing⁷²; the position of Dicer cleavage and thus mature strand selection from pre-mir-324 is adjusted by its TUT4/7-mediated uridylation.⁷³ Thus, different TENTs have overlapping activities on certain miRNAs, but opposite activities on different miRNA substrates; they also have overall minor effects on accumulation of most miRNAs.⁷

Our initial inspection of a then-limited set of mammalian mirtrons (11 in human and 20 in mouse) revealed that all mirtron-3p species were heavily tailed, similar to invertebrate mirtron-3p species.⁷⁵ Now, meta-analysis of the tailing properties of our highly expanded catalog of mammalian mirtrons,²³ across numerous public small-RNA datasets (Table S1), broadly reaffirms this perspective. In particular, mirtron-5p reads exhibit similar modest U/A tailing, comparable to aggregate canonical miRNA-5p and miRNA-3p reads, while mirtron-3p species accumulate high amounts of untemplated adenylation and particularly of untemplated uridylation (Figure S4). These are specific patterns, since other intron-3p reads did not exhibit comparable uridylation, although some signal for adenylation was observed among more abundant intron-3p species (Figure S4).

Importantly, as the dominant class of splicing-derived mirtrons is by far 5'-tailed mirtrons, followed by conventional mirtrons, we could define the 3' ends of most mirtron-3p species with respect to the -AG splice acceptor (SA). Since this provides an independent reference for the primary endonucleolytic cleavage position, we are able to infer untemplated additions to mirtron-3p reads that match the genome. This was particularly valuable as some mirtrons yielded dominant populations of modified reads, whose terminal nucleotide(s) fortuitously matches the downstream exon (Figures 2 and S5).

The highly preferential modification patterns of mirtron-3p species implied that their hairpin precursors are the tailing substrate. For example, this was seen in the cases of *pre-mir-1226* and *mirtron-Ankrd52* (Figure 2). Of note, the dominant hairpin reads of *mirtron-Ankrd52* have one or two nucleotides trimmed from the 3'-SA, and thus, tailed reads are a minority of the overall hairpin reads. However, for *pre-mir-1226*, >99.9% of hairpin reads carry 3' untemplated modifications, mostly U tails and a smaller fraction of A tails.

To evaluate this more systematically and uniformly, we compared canonical miRNAs and ssdRNAs/mirtrons using



CelPress

Cell Reports Article

Figure 4. TUT4, TUT7, and TENT2 preferentially modify ssdRNA/mirtron hairpins

(A–C) Northern analysis of ssdRNA/mirtron products from HEK293T cells transfected with the indicated expression constructs. (A) *mir-1225* and *Acadvl* barely yield northern-detectable species, but hairpin-sized RNAs accumulated upon coexpression of Ago2. (B) Levels of *pre-mir-1226* and miR-1226-3p are decreased upon coexpression of TENT2, TUT4, and TUT7, without substantial effects on control *pre-mir-375*. Mature miR-375-3p is selectively tailed by TENT2. (C) Accumulation of *mir-1225/Pkd1* and mirtron-*Acadvl* was similarly selectively suppressed by these TENT/TUT enzymes.

(D) Luciferase sensor assays corroborate selective suppression of *mir-1225* and mirtron-*Acadvl* by TENT2/TUT2/TUT7. Sensor activities were normalized to cells transfected with a non-cognate mirtron expression construct; n = 3 for each experiment and unpaired Student's t test was applied. The error bars represent standard deviation; **p < 0.01, ****p < 0.0001.

showed that ectopic expression of a Drosophila TUTase (Tailor) could modify and/or inhibit mirtrons.²⁵ we screened seven available mammalian TENT constructs⁷⁰ for analogous activities. In testing expression vectors for various ssdRNA/mirtron constructs, we did not detect RNA products from several of these (including mir-6511a, mir-6864/ 6865, and hsa-uc010zyo.2). However, the accumulation of hairpin species from some constructs was potentiated by coexpression of Ago2 (e.g., mouse mirtron mir-1225/Pkd1 and mirtron-Acadvl; Figure 4A), as reported.³⁸ This aligns with the notion of homeostatic control of Argonaute occupancy^{76,77} and further indicates that non-optimal Dicer substrates with Argonaute proteins in Dicer-proficient

can associate w cells.^{38,41,42,78}

We determined mammalian ssdRNA/mirtron constructs that yielded detectable hairpin- and/or small-RNA products and used *mir-375* as a control canonical miRNA that is not expressed in HEK293T cells (Figures 4B and 4C). We co-transfected these small RNA constructs with Ago2 and individual TENT vectors and assayed their responses using northern blotting. None of these enzymes had substantial effects on *pre-mir-375*, although ectopic TENT2 induced tailing of mature miR-375 (Figure 4B). Although miR-375 is an miR-3p species, presumably its modification was induced at a post-dicing step, and TENT2 (also termed TUT2 or Gld2) is known to modify mature small RNAs.^{79,80} In contrast, *pre-mir-1226* hairpin levels were strongly reduced in the presence of ectopic TENT2, TUT4, and TUT7 (Figure 4B). Similarly, the hairpin levels of mirtron *mir-1225/Pkd1* and *mirtron-Acadvl* were robustly reduced by TENT2/TUT4/TUT7,

matched datasets of Ago-bound intermediate-length and small RNAs from HEK293T cells.⁴⁹ As expected, the bulk of canonical pre-miRNAs and mature miRNAs were unmodified (Figures 3N and 3O). In stark contrast, the vast majority of known mirtrons carry untemplated U or A in the hairpin fraction, and a comparably large fraction of ssdRNAs carry untemplated uridylation (Figure 3N). These features carry over into the small-RNA pool, with the caveat that very few ssdRNA-derived small RNAs were available for analysis (Figure 3O). We provide the tailing properties on a per-locus basis in Table S4. Overall, we conclude that ssdRNA/mirtron hairpins are the dominant substrates for 3' tailing among mammalian miRNA-related populations.

TENT2/TUT4/TUT7 specifically modify ssdRNA/mirtron hairpins and antagonize their activity

We sought to identify specific TENTs, which include TUTases, that were responsible for mirtron modification. As we previously



but not TENT1 (Figure 4C). Finally, to assess if these TENTs could inhibit mirtron activity, we performed luciferase sensor assays for mirtron-*Acadvl* and mirtron *mir-1225/Pkd1*. We utilized sensors designed to the 5p ends of these mirtrons, which can detect their activity as Ago-associated hairpins.³⁸ The repression capacities of both mirtrons were little changed in the presence of ectopic TENT1, but strongly reduced upon co-expression of TENT2/TUT4/TUT7 (Figure 4D).

Altogether, these data indicate that specific TENT enzymes antagonize mirtron accumulation and activity in mammals, as with *Drosophila* Tailor^{25,26} and *C. elegans* CID-1.⁸¹ However, we note that (1) TENT2 is homologous to *Drosophila* Gld2, (2) mammalian TUT4/7 and nematode CID-1 are distinctive mammalian TUTases that are absent from *Drosophila*, marked by a duplicated (inactive) nucleotidyltransferase domain, and (3) fly Tailor lacks an apparent direct mammalian homolog.²⁷ Thus, mirtron modification comprises widespread, yet evolution-arily convergent, activities in different species.

Combined activities of endogenous TUT4/7 with TENT2 mediate ssdRNA/mirtron tailing

To assess if these TENTs were responsible for endogenous modifications of mirtrons and ssdRNAs, we utilized a panel of mutant HEK293T cells bearing mutations in TENT2, TUT4, and/ or TUT7. Due to the challenges in detecting endogenous mirtrons using northern blotting, we co-transfected mirtron constructs (*mir-1226*, *mirtron-Acadvl*, and *mir-1225/Pkd1*) with canonical *mir-375* and used the latter to assess specificity of modification changes.

Although TENT2/TUT4/TUT7 each had strong and selective effects on mirtrons in gain-of-function settings, individual



Figure 5. Endogenous TUT4/7 and TENT2 selectively modify mirtrons

(A–C) Wild-type and TENT/TUT knockout HEK293T cells were transfected with mirtron, canonical miRNA (*mir-375*), and hAgo2 expression constructs, and total RNAs were analyzed by northern blotting. The accumulation of a longer intermediate-length species for mirtron-*Acadvl* (A), *pre-mir-1225* (B), and *pre-mir-1226* (C) was reduced in *TUT4/7-dKO* cells and absent in *TUT4/7+TENT2-tKO* cells.

(D) Luciferase sensor assays show that the activity of mir-1225 and mirtron-*Acadvl* is increased in tKO cells. Sensor activities were normalized to those of cells transfected with a non-cognate mirtron expression construct; n = 3 for each experiment and unpaired Student's t test was applied. The error bars represent standard deviation; ***p < 0.001, ****p < 0.0001.

knockouts for these factors did not substantially alter their patterns as hairpins or mature small RNAs (Figures 5A–5C). However, we found that *TUT4*/7-dKO cells⁸² resulted in a noticeable shift of the hairpin forms of all three mirtrons to a shorter species, demonstrating

that the longer isoforms truly bear untemplated additions from these TUTases. Although TUT4/7 are the major miRNA uridylases, there are specific pre-miRNAs (e.g., the "group II" subclass of let-7 precursors) for which adenylase TENT2 has certain overlapping activities with TUT4/7.⁷² This encouraged us to analyze a triple knockout (tKO) of *TUT4/* 7+*TENT2*,⁸³ which nearly eliminated the longer mirtron hairpin species of *mir-1226*, *mirtron-Acadvl*, and *mir-1225/Pkd1* (Figures 5A–5C). By contrast, no changes to length isoforms were observed with hairpin or mature forms of *mir-375* in *dKO* or *tKO* cells.

To test for regulatory effects, we focused on *Acadvl* and *mir-1225/Pkd1*. Due to relatively inefficient transfection especially of *tKO* cells, we took care to normalize these tests against non-cognate mirtrons. The responses of *Acadvl*-5p and miR-1225-5p/Pkd1 sensors were greater in combination mutants of these tailing enzymes (Figure 5D). Overall, these assays demonstrate that multiple TENT enzymes (TUT4/7 and TENT2) are coordinately involved in endogenous tailing and suppression of mirtrons in mammalian cells.

Endogenous TUT4/7 with TENT2 broadly modify splicing-derived small RNAs

Despite early precedents for TENT-mediated regulation of miRNA biogenesis and/or levels,²⁷ analysis of *TENT2* or *TUT4/7* knockouts showed that these enzymes have subtle impacts on overall miRNA biogenesis.^{68,74,82–84} This is the case, despite the fact that individual miRNA substrates can be substantially regulated (e.g., let-7 members). Similarly, in *C. elegans*, depletion of the miRNA-regulating TENT enzyme CID-1 has only minor effects on miRNA accumulation.⁸¹





Figure 6. Genome-wide analysis of mirtrons in TUT4/7 ± TENT2 mutants

(A) Levels of untemplated modifications to canonical miRNAs and mirtron-derived small RNAs in control, *TUT4/7-dKO*, and *TUT4/7+TENT2-tKO* cells. The normally high frequency of untemplated uridylation to mirtron-derived small RNAs is reduced in *dKO* and strongly abrogated in *tKO* cells. Untemplated adenylation to mirtrons is elevated in *dKO* but decreased in *tKO* cells. Error bar represents standard error (SE).

(B) Example behaviors of individual mirtron loci. Although small RNAs are detected from all of these loci, they accumulate substantial or predominant hairpin reads. Small-RNA sequencing indicates that tailing loss correlates with increased levels in *TUT4/7-dKO* cells. However, among these, miR-1229 also had substantially increased small RNAs in *TUT4/7+TENT2-tKO* cells.

(C) Volcano plots of replicate small-RNA data show unidirectional increase in mirtron-derived species in *TUT4/7-dKO* vs. control cells. The modest expression of most mirtron-derived small RNAs hinders quantification; however, there is a bias for increased levels among changed loci that did not meet statistical significance. Small-RNA levels were normalized to a set of small-RNA spike-ins.

(D) Analysis of *TUT4/7+TENT2-tKO* shows a bias toward increased levels of mirtron-derived small RNAs, but this is not as directional as with *TUT4/7-dKO* cells. (E) Northern blotting for ssdRNA-*Ankrd52*, which increases in dKO and tKO cells.

(F) Quantification of ssdRNA-*Ankrd52* levels across three independent experiments; n = 3 for each experiment, and unpaired Student's t test was applied. The error bars represent standard deviation; **p < 0.01, ***p < 0.001.

We analyzed the consequences of individual and combinations of TENT enzyme knockouts on different classes of miRNAs by generating replicate libraries from control, *TUT4/7-dKO*, and *TUT4/7+TENT2-tKO* HEK293T cells. As discussed earlier, while we used intermediate-length RNA data for annotation, this strategy is of uncertain utility for quantification due to efficient and biased cloning of structured RNAs. Instead, we focused these analyses on mirtron-derived small RNAs, since our newly recognized ssdRNAs yield negligible small RNAs in HEK293T.

Although it is well known that a subset of canonical miRNAs is modified by TUT4/7 and TENT2,^{73,82,85} to our knowledge, previous studies did not specifically segregate mirtrons for analysis. By direct comparison of canonical miRNAs and mirtron-derived small RNAs, we can observe that both classes received untemplated uridylation and adenylation via TUT4/7 and TENT2. Importantly, however, the vast majority of canonical miRNA species from either hairpin arm are not modified, so that the respective changes in *TUT4/7* and *TUT4/7*+*TENT2* mutant cells affect a small portion of total miRNA reads (Figure 6A).

By contrast, the vast majority of mirtron-3p reads are 3'-modified, as mentioned, and their untemplated uridylation is substantially dependent on TUT4/7; this double mutant also yields a concomitant increase in adenylation of mirtron-3p reads (Figure 6A). In *TUT4/7+TENT2* mutant cells, the overall levels of untemplated additions to mirtron-3p are markedly reduced, almost to those of canonical miRNA levels (Figure 6A). Nevertheless, while low, the remaining frequencies of untemplated U/A additions to mirtron-3p species in triple-mutant cells are actually somewhat higher than the aggregate untemplated additions to canonical miRNAs in wild-type cells, suggesting that another TENT(s) can sense the difference in biogenesis heritage of miRNA subclasses.

It was earlier reported that small-RNA tailing is linked to downregulation by various mechanisms.^{47,69–71,86} However, it was more recently concluded that there is no overall linkage between miRNA tailing and steady-state accumulation,^{82,87} and tailing-independent mechanisms for miRNA turnover exist.^{88,69} Analysis of canonical miRNAs from these TENT/TUT mutants confirms expectations that multiple members of the let-7 family are distinctly and highly upregulated in *TUT4/7* and *TUT4/* 7+*TENT2* mutant small-RNA data; these clearly comprise the dominant targets (Figure S6). In addition, a small subset of other canonical miRNAs increases in one or both mutants. However, another subset of miRNAs decreases, while the strong majority

is unchanged (Figure S6). This extends previous conclusions that there may be several, and in aggregate non-directional, impacts of TUT4, TUT7, and TENT2 on canonical miRNAs.⁸³

By comparison, inspection of individual mirtrons in TENT/TUT mutants showed multiple loci with clearly increased small RNAs (Figure 6B). With this in mind, we performed systematic analysis of mirtron-derived small RNAs in TENT/TUT mutants. Indeed, mirtron products were directionally upregulated in *TUT4/7-dKO* cells (among statistically significant changes, 16 increased and none decreased; Figure 6C). Moreover, loci whose changes did not reach statistical significance were biased toward increased levels. Because so many mirtron-derived small RNAs have modest biogenesis, it is difficult to measure their levels accurately. Nevertheless, it is evident that uridylation overall inhibits steady-state accumulation of mirtrons.

The effect of tailing loss in TUT4/7+TENT2 mutant cells on the accumulation of mirtron-derived small RNAs was more complex. Inspection of individual mirtrons (Figure 6B) showed that small RNAs from some loci increased further in tKO vs. dKO cells (e.g., mir-1229), but that mature levels were lower in tKO cells for some others (e.g., mir-1226, mir-3620). On one hand, one may have expected more severe effects on mirtron processing in triple mutants, since our biochemical and genomic analyses showed that TUT4/7+TENT2 mutants exhibit far less tailing on mirtrons than the double mutant (Figures 5 and 6). However, we also note other precedents, e.g., group II let-7 members,⁷² for which the combined tailing activity of TUT4/7 + TENT2 promotes the biogenesis of certain pre-miRNAs, even though TUT4/7 inhibit the biogenesis of other let-7 members. On the genome-wide scale, we observe that TUT4/7+TENT2 cells still exhibit directional increase in mirtron-derived small RNAs (Figure 6D). Still, these effects are somewhat tempered with respect to dKO cells.

One interpretation is that general gene expression defects are greater in triple-mutant cells, and thus tKO cells may harbor more indirect effects on mirtron host genes, leading to compromised signals. However, we also note that mammalian mirtrons are known to comprise highly heterogeneous hairpin structures,^{22,23} which fits with the fact that they are not subject to gatekeeping by Microprocessor. Notably, unlike the majority of initially cloned Drosophila and nematode mirtrons, which generate small RNAs from hairpins bearing 2-nt 3' overhangs,^{18,19} many initially discovered mammalian mirtrons lack typical pre-miRNA overhangs.¹⁷ Instead, many of them notably exhibit 1-nt 3' overhangs, or other non-standard (e.g., 5' overhangs) or heterogeneous overhangs that arise from presumably imprecise removal of the 5' tail (Figure S7). For these reasons, and unlike with canonical pre-miRNAs,^{31,32} it is difficult to assess mirtron hairpin termini systematically. However, tailing of ssdRNA/mirtron hairpins with 1-nt 3' overhangs or 5' overhangs may yield structures that are more favorable for dicing (Figure S7). Thus, akin to group II let-7 members, abundant mirtron hairpin tailing by multiple TENT enzymes may paradoxically aid the dicing of a population of ssdRNAs/mirtrons.

The newly recognized class of ssdRNA loci in this study intrinsically does not generate much small RNA, precluding their analysis. Only 12 ssdRNAs had even one small RNA in any of our



newly made datasets, and we did not observe directional increase in TUT4/7 \pm TENT2 mutant cells, except for three paralogs of the known *mir-6511* mirtron family. The fact that tailing loss was insufficient to induce ectopic ssdRNA-derived small RNAs may reflect that they are also largely suboptimal Dicer substrates. As an alternative approach to test the accumulation of endogenous ssdRNA, we assayed several by northern blotting. We did not succeed for several loci, but obtained robust signals for ssdRNA-*Ankrd52* (Figure 2B), allowing us to compare across genotypes. These tests show that endogenous ssdRNA-Ankrd52 is specifically and reproducibly upregulated in *TUT4/7* and *TUT4/7*+*TENT2* cells, compared with control and *TENT2-KO* cells (Figures 6E and 6F).

Overall, key conclusions from these studies are that, unlike with canonical miRNAs, the biogenesis of mirtrons into small RNAs is directionally suppressed by tailing enzymes, potentially at the levels of both hairpin- and small-RNA accumulation. We provide comprehensive expression information for different classes of small RNAs in the panel of TENT/TUT mutants in Table S5.

Impact of 3'-untemplated modification systems on canonical miRNA evolution

We showed that evolutionarily young ssdRNAs/mirtrons can broadly associate with core miRNA machinery, including both Dicer and Argonaute proteins (Figure 3). We hypothesized that such generally low-level interactions may not exert beneficial consequences for gene regulation, but instead collectively contaminate the canonical miRNA pathway (Figure 1). In this view, tailing is a genomic defense against unvetted RNA hairpins that have opportunities to associate with canonical miRNA machinery.

In *Drosophila*, the effective suppression of mirtrons via the Tailor TUTase is correlated not only with rapid evolutionary turnover of mirtrons across fly species, ⁵⁶ but also with the depletion of canonical miRNAs bearing terminal 3'-AG residues.^{25,26} This seems to be accounted for by intrinsic recognition of intronderived hairpins bearing signature 3' splice sites by Tailor.^{90,91}

Mammalian mirtrons similarly evolve extremely rapidly, as only a few percent of ssdRNAs/mirtrons in humans are shared in mouse.²³ However, is there also evidence that the inferred intragenomic conflict between splicing-derived hairpins and tailing pathways has had collateral effects on the modification and/or evolution of mammalian canonical miRNAs? If so, that might be the clearest evidence that the role of ssdRNA/mirtron tailing in mammals is to suppress their incorporation into miRNA regulatory networks.

We first addressed this by performing a meta-analysis of aggregated human small-RNA data (Table S1). Even though the overall level of modifications to canonical miRNAs is very low, by combining numerous datasets, we gained sufficient power to resolve trends for preferred trends of untemplated additions to different individual 3' miRNA termini (Figure 7). Overall, these analyses support the following conclusions (see also Table S4 for per-locus data).

First, 3' untemplated uridylation is biased to occur on premiRNA hairpins compared with mature small RNAs, as reflected by the far lower frequencies of uridylation on miRNA-5p vs.



CelPress

Cell Reports Article

Figure 7. Collateral effect of tailing enzymes on canonical miRNA evolution

(A and B) Reverse CDF plots depicting untemplated mono-uridylation to human miRNA-5p (A) and miRNA-3p (B) species, compared with mirtron-3p species. For these analyses, we considered a single dominant 5p and 3p species of each canonical miRNA and plotted only single nucleotide additions; we considered only mirtron-3p species at splice sites (i.e., bearing -AG termini) and their untemplated uridvlated forms. At least 10 reads were needed to include the locus in analysis; this could differ between 5p and 3p loci of the same hairpin. (A) There are no substantial differences in tailing among mirtron-5p and different terminal nucleotide classes of miRNA-5p species. (B) Mirtron-3p species are dominant recipients of untemplated 3' uridvlation, as seen in other analyses. However, canonical miRNA-3p species ending in 3'-AG exhibit greater tailing than other canonical miRNA-3p species that end in 3'-G (i.e., UG/CG/ GG = BG), which in turn exhibit greater tailing than all other canonical miRNA-3p species (3'-A/C/U = H).

(C) Biased evolutionary distribution of miRNA loci based on 3' terminal nucleotides of miRNA-3p species. We segregated conserved miRNAs (i.e., across Drosophilid phylogeny, at least from *D. melanogaster* to *D. pseudoobscura*, or human miRNAs conserved to rodents) from recently emerged loci (only in *melanogaster* group species or only in primates, respectively). Among recently emerged miRNAs (gray bars) there is a relatively even distribution of miRNAs of different 3' nucleotides. The apparent excess of 3'-U miRNAs may potentially be clouded by the possibility of genome-matching untemplated uridylation. By comparison, conserved 3'-G miRNA-3p loci are depleted in both flies and mammals, as determined by Fisher's exact test (*p < 0.05).

miRNA-3p small RNAs (Figure 7A). This difference is particularly exaggerated for mirtrons, as seen with HEK293T data (Figure 6A). Next, among unambiguous 3'-monouridylation to miRNA-5p species, there were no substantial differences in the overall frequencies of tailing to miRNAs that ended in G vs. other nucleotides (i.e., H = A/U/C) and no substantial difference if we divided the latter into 3'-AG miRNAs and 3'-BG miRNAs (where B = U/C/G). However, the picture was very different for canonical miRNA-3p species. All canonical miRNA-3p loci that end in G exhibit enhanced uridylation compared with A/U/C-terminal miRNAs (Figure 7B), supporting the idea that the tailing mechanism is sensitive to terminal nucleotide identity. In particular, canonical miRNA-3p reads that end in 3'-AG collectively exhibit greater terminal uridylation than do 3'-BG miRNA-3p reads (Figure 7B). Although the frequency of tailing of such canonical miRNA-3p 3'-AG species remains far lower than that of mirtron-3p species, it is consistent with the notion that the uridylation factor(s) is biased to modify SA mimics.

Next, we evaluated the potential relationship of this with canonical miRNA evolution. We segregated canonical human miRNAs into conserved (pan-mammalian) and newly emerged (only in primates or humans) loci and sorted them by annotated miRNA-3p nucleotide. This analysis is partially limited by the fact that many pre-miRNAs are not cleaved uniformly and instead generate alternative products.^{31,32} This analysis attributes each miRNA into one terminal nucleotide group, as it may be more difficult to interpret if individual miRNAs reside in two or more such groups. Our analysis of *Drosophila* miRNAs shows an apparent excess of canonical miRNAs that end in U and A (Figure 7C). This might be the case, but it may potentially reflect a certain degree of ambiguity if some of these are actually untem-

plated modifications that happen to match the genome. In any case, as reported earlier,²⁵ the frequency of 3'-G miRNA-3p loci is near background expectation among newly emerged miRNAs, but substantially depleted among conserved miRNAs. Similarly, among human canonical miRNAs, we again find that 3'-U is overrepresented among all groups of loci, possibly reflecting tailing activity that fortuitously generates reads that match the genome. However, while human non-conserved 3'-G miRNA-3p loci are at the expected frequency, mammalian conserved 3'-G miRNA-3p loci are substantially depleted (Figure 7C). The underlying statistics for these analyses by miRNA evolutionary depth are provided in Table S6.

Altogether, we find evidence for parallel evolutionary behavior of miRNA loci in *Drosophila* and mammals, in which canonical 3'-G miRNA hairpins are specifically underrepresented among miRNA loci that have survived across distant phylogeny. We take this to reflect the fundamental and convergent action of TENT enzymes to suppress this intragenomic conflict, by which unvetted splicing-derived RNA hairpins gain unwanted access to the canonical miRNA machinery. This pathway promotes the evolutionary turnover of ssdRNAs/mirtrons, but also has a collateral impact to deplete similar canonical pre-miRNA hairpins.

DISCUSSION

Tailing defends the canonical miRNA pathway from parasitic splicing-derived substrates

The ribonucleolytic activity of the spliceosome generates numerous non-canonical pre-miRNA substrates in diverse metazoan species. While this in principle expands the regulatory reach of miRNA networks, splicing-derived pre-miRNAs are

rarely selected for effective maturation of small RNAs and are typically poorly conserved. We now reveal a much broader universe of ssdRNAs in mammalian cells, further emphasizing that an abundance of fortuitous pre-miRNA mimics may not necessarily benefit normal gene regulation. On the contrary, their existence may generally be a nuisance to the canonical miRNA pathway, which is normally initiated by careful substrate selection and processing by the gatekeeper Microprocessor complex.^{31,92,93}

Indeed, the modest biogenesis properties of most ssdRNAs/ mirtrons now make ontological sense. The majority of canonical miRNAs with efficient biogenesis are conserved and selected for effective biogenesis features, whereas most mirtrons are recently emerged and appear to have fortuitous access to miRNA biogenesis factors. We conceive that tailing enzymes comprise a genomic defense against the assault of unvetted ssdRNA/mirtron species and take advantage of their dominant signatures (i.e., hairpins that end in SA signals). Remarkably, we find evidence for widespread inhibition of splicing-derived hairpins by tailing, despite the fact that the underlying mirtron pathways, substrate molecules, and even relevant TENT enzymes are not shared between Drosophila^{25,26} and mammals (this study). A further wrinkle is that C. elegans, in which mirtron-3p-3' termini are highly uridylated,⁷⁵ contains a TUT4/7 homolog named CID-1/PUP-1²⁷ that is responsible for tailing mirtrons and canonical miRNAs.⁸¹ Although it is difficult to be certain of their evolutionary history, the distinct major strategies for splicing-derived miRNA biogenesis and defense in invertebrates and mammals^{20,23,56} suggest they may have emerged independently. If so, this could reflect inherent sensitivities of the canonical miRNA pathway to molecular parasitism and thus recurrent cycles of infiltration and suppression.

During revision of this study, Svoboda and colleagues reported a role for the mammalian Dicer helicase domain as a "sensor" that prevents it from processing long dsRNA.⁹⁴ Since mirtrons have heterogeneous structures, owing to bypass of quality control by Microprocessor, a subset of them have atypically long stems.²³ Strikingly, in Dicer∆hel cells, there is increased dicing of a subset of miRNA loci, and these turn out to be substantially enriched in mirtron loci.⁹⁴ Thus, it seems there are indeed two parallel strategies that oppose ssdRNAs/mirtrons: direct recognition and modification by TENT enzymes and also direct inspection of structures by Dicer helicase. We take this as a further indication of the importance of preventing access of fortuitous splicing-derived substrates to the canonical miRNA pathway.

Overall, the evolutionary dynamics of ssdRNAs/mirtrons appear to conform to a Red Queen arms race of intragenomic conflict.⁹⁵ An analogy may be found in the transposable elements (TEs), whose actions are generally deleterious to host genomes and organismal fitness and thus require specialized and efficient mechanisms of host defense. Some of these molecular defenses against TEs can have knock-on effects on endogenous gene regulation (i.e., heterochromatinization and/or piRNA-mediated post-transcriptional silencing). At the same time, a subset of TEs can over time become "tamed" and incorporated into normal and essential gene regulatory pathways, contributing both *cis*-regulatory sites and host transcription factors.⁹⁶



By analogy, we propose that most ssdRNAs/mirtrons have fortuitous association with canonical miRNA machinery and thus have deleterious impacts, by displacing access of bona fide miRNA substrates to limiting miRNA factors⁷⁶ and/or by effectively inducing off-target gene regulation. Even though the regulatory influence of individual low-expressed loci may be modest, it is not difficult to envision that the collective impacts of many hundreds of such loci associating with Dicer and/or Ago necessitates a vigorous defense. Such a tailing defense has had detectable collateral impact on canonical miRNA biogenesis and evolution, in the form of depletion of conserved pre-miRNAs that resemble splicing products. However, this is presumably an acceptable price to pay to enable purification of the canonical miRNA pathway. At the same time, a small handful of splicing-derived miRNA substrates have become incorporated into normal regulatory networks, as evidenced by higher steadystate accumulation and evolutionary persistence.^{17,23} Thus, even enemies of the state can in the end become allies.

Quality control in the miRNA pathway

There is increasing recognition for the importance of quality control and surveillance pathways that function at molecular, organellar, and cellular levels.⁹⁷ These are not always essential for normal operation of a given pathway or process. That is, cell and/or organismal viability can tolerate a certain amount of noise or frank contamination of regulatory pathways. But the fact of numerous documented quality control mechanisms reflects that suboptimal performance will be outcompeted in wild settings or over evolutionary timescales.

Within Argonaute pathways specifically, diverse strategies have been reported. While untemplated tailing was the first strategy implicated,²⁸ many molecular mechanisms have now been described. For example, plants utilize multiple factors to prevent inappropriate small RNA clients of the miRNA effector AGO1. The F-box SCF component FBW2 targets unloaded AGO1 for turnover. This pathway is not intrinsically needed for normal development, but fbw2 mutants can access illegitimate small RNAs for gene silencing.⁹⁸ Similarly, the nucleotidase/phosphatase FIERY1, in concert with exoribonucleases XRN2 and XRN3, suppresses rRNA-derived small interfering RNAs (risiRNAs). In their absence, risiRNAs not only accumulate aberrantly but also can misload into AGO1, thereby interfering with miRNA regulation.99,100 The complex intersections of plant small RNA pathways also necessitate TUTase enzymes (URT1 and HESO1) and decapping factors, ^{101–103} as well as efficient ribosome release and RNA exosome activity,¹⁰⁴ to limit miRNAinduced production of secondary siRNAs. Finally, direct analysis of plant pre-miRNAs reveals both untemplated uridylation and cytidylation, with the former mediated by HESO1 and the latter mediated by HESO1/NTP6/NTP7.¹⁰⁵ A combination of functional impacts was inferred, with some trimmed pre-miRNAs subject to hairpin repair to promote biogenesis, while others were subject to HESO1-mediated oligouridylation for turnover.

The action of tailing enzymes for quality control in small-RNA pathways is presumably ancient or highly recurrent, given its evolutionary breadth. Beyond plants, the yeast TENT enzymes Cid16 (adenylase) and Cid14 (uridylase) are involved in RNAi quality control via the exosome, by tailing and degrading small



RNAs from euchromatic loci that might inappropriately lead to their heterochromatic silencing.¹⁰⁶ In animal cells, targetdirected miRNA tailing and degradation was proposed to purify the contents of Ago proteins.⁸⁶ TUT4/7 lie at the heart of one of the first paradigms of regulated miRNA biogenesis.^{70,71,107} They are recruited by LIN28, a specificity determinant that identifies pre-let-7 target hairpins.^{108–112} Subsequent polyuridylation of pre-let-7 hairpins leads to their degradation by the Dis3L2 nuclease.^{113,114} Similarly, TUT4/7 were found to promote turnover of defective pre-miRNA hairpins.47 An opposite role is proposed for the same enzymes, along with TUT2, in which monouridylation of certain pre-let-7 hairpins with suboptimal overhangs (group II) improves their dicing.⁷² Pre-miRNA uridylation by TUT4/7 has also been described to alter the Dicer processing register, and thus the functional miRNA output, of an individual hairpin.⁷³ In addition, TUT4/7 can also enhance the base-pairing properties of an miRNA, thereby adjusting its targeting capacity.¹¹⁵ Thus, tailing can lead to downregulation, upregulation, or altered function of an miRNA and can target both specific loci and entire classes of miRNA substrates.

Limitations of the study

Methodologies to clone and sequence short RNAs have been optimized for presumably unstructured linear species, such as mature miRNAs. It is known that highly structured species such as pre-miRNA hairpins are typically recalcitrant to adaptor ligation, and they may further be subject to biased amplification during library construction for intermediate-sized RNA species. For example, the levels of known canonical premiRNAs in such libraries may not correspond to their behavior on an independent assay, such as northern blotting. Accordingly, there are very few studies of pre-miRNA analysis from deep sequencing. This study focused on the annotation of previously unrecognized species in intermediate libraries that resemble pre-miRNA hairpins, but it remains unknown how quantitatively reliable the counts from such libraries are. Therefore, at present, until technical improvements are developed, it may be challenging to directly infer, for example, dicing efficiency or relative expression of pre-miRNA and pre-miRNA-like species from intermediate-length RNAsequencing data.

STAR*METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Processing and curation of human small RNA, intermediate-sized RNA, and CLIP-seq data
 - Intermediate-length RNA Ago2-IP library from K562 cells



- Small RNA libraries from HEK293T cells
- ssdRNA identification and annotation
- Analysis of 3' untemplated additions
- Analysis of miRNA evolution
- Small RNA Northern blotting
- Luciferase sensor assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j. celrep.2023.112111.

ACKNOWLEDGMENTS

We thank Fuqu Hu for conducting some initial analyses of small-RNA responses to ectopic TENT enzymes, Erik Ladewig for initial studies of mirtron tailing, and Jorgen Kjems for sharing mirtron plasmids. S.L. was supported by a training award from the NYSTEM, contract C32559GG, and the Center for Stem Cell Biology at MSKCC. Work in S.P. lab was funded by the European Research Council (ERC-CoG-647455 RegulRNA) and was performed under the framework of the LABEX: ANR-10-LABX-0036_NETRNA and ANR-17-EURE-0023. J.W. was supported by an Australian Research Council (ARC) Future Fellowship (FT16010043) and an Australian National University (ANU) Futures Scheme. Work in E.C.L. lab was supported by NIH grants R01-GM083300 and R01-HL135564 and by the MSK Core Grant P30-CA008748.

AUTHOR CONTRIBUTIONS

S.L., D.J., A.Y., R.S., and D.B.-B. conducted biochemical and functional analysis of mirtrons/ssdRNAs. S.S. and A.R. annotated ssdRNAs. S.L. and J.W. performed genomic and bioinformatic analyses. S.P., S.G., J.W., and E.C.L. were responsible for supervision and funding acquisition. S.L. and E.C.L. wrote the manuscript with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 28, 2022 Revised: November 16, 2022 Accepted: January 30, 2023 Published: February 16, 2023

REFERENCES

- Treiber, T., Treiber, N., and Meister, G. (2019). Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat. Rev. Mol. Cell Biol. 20, 5–20. https://doi.org/10.1038/s41580-018-0059-1.
- Bartel, D.P. (2018). Metazoan MicroRNAs. Cell 173, 20–51. https://doi. org/10.1016/j.cell.2018.03.006.
- Lai, E.C. (2002). microRNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. 30, 363–364.
- Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of MicroRNA-target recognition. PLoS Biol. 3, e85.
- Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife 4, e05005. https://doi.org/10.7554/eLife.05005.

- McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., and Bartel, D.P. (2019). The biochemical basis of microRNA targeting efficacy. Science 366, eaav1741. https://doi.org/10.1126/science.aav1741.
- Friedman, R.C., Farh, K.K.H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105.
- Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769–773.
- Fromm, B., Keller, A., Yang, X., Friedlander, M.R., Peterson, K.J., and Griffiths-Jones, S. (2020). Quo vadis microRNAs? Trends Genet. 36, 461–463. https://doi.org/10.1016/j.tig.2020.03.007.
- Keller, A., Gröger, L., Tschernig, T., Solomon, J., Laham, O., Schaum, N., Wagner, V., Kern, F., Schmartz, G.P., Li, Y., et al. (2022). miRNATissueAtlas2: an update to the human miRNA tissue atlas. Nucleic Acids Res. 50, D211–D221. https://doi.org/10.1093/nar/gkab808.
- Chu, A., Robertson, G., Brooks, D., Mungall, A.J., Birol, I., Coope, R., Ma, Y., Jones, S., and Marra, M.A. (2016). Large-scale profiling of microRNAs for the cancer genome atlas. Nucleic Acids Res. 44, e3. https://doi.org/ 10.1093/nar/gkv808.
- Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173, 291–304.e6. https://doi.org/10.1016/j.cell. 2018.03.022.
- Wen, J., Mohammed, J., Bortolamiol-Becet, D., Tsai, H., Robine, N., Westholm, J.O., Ladewig, E., Dai, Q., Okamura, K., Flynt, A.S., et al. (2014). Diversity of miRNAs, siRNAs and piRNAs across 25 Drosophila cell lines. Genome Res. 24, 1236–1250.
- Rahmanian, S., Murad, R., Breschi, A., Zeng, W., Mackiewicz, M., Williams, B., Davis, C.A., Roberts, B., Meadows, S., Moore, D., et al. (2019). Dynamics of microRNA expression during mouse prenatal development. Genome Res. 29, 1900–1909. https://doi.org/10.1101/gr. 248997.119.
- Fromm, B., Zhong, X., Tarbier, M., Friedländer, M.R., and Hackenberg, M. (2022). The limits of human microRNA annotation have been met. RNA 28, 781–785. https://doi.org/10.1261/ma.079098.
- Yang, J.S., and Lai, E.C. (2011). Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol. Cell 43, 892–903.
- Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol. Cell 28, 328–336.
- 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.
- Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature 448, 83–86.
- Westholm, J.O., and Lai, E.C. (2011). mirtrons: microRNA biogenesis via splicing. Biochimie 93, 1897–1904.
- Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D., and Lai, E.C. (2010). microRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol. Cell 38, 900–907.
- Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O., and Lai, E.C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. Genome Res. 22, 1634–1645. https://doi.org/10.1101/gr. 133553.111.
- Wen, J., Ladewig, E., Shenker, S., Mohammed, J., and Lai, E.C. (2015). Analysis of nearly one thousand mammalian mirtrons reveals novel features of dicer substrates. PLoS Comput. Biol. *11*, e1004441. https:// doi.org/10.1371/journal.pcbi.1004441.
- 24. Castellano, L., and Stebbing, J. (2013). Deep sequencing of small RNAs identifies canonical and non-canonical miRNA and endogenous siRNAs



in mammalian somatic tissues. Nucleic Acids Res. *41*, 3339–3351. https://doi.org/10.1093/nar/gks1474.

- Bortolamiol-Becet, D., Hu, F., Jee, D., Wen, J., Okamura, K., Lin, C.J., Ameres, S.L., and Lai, E.C. (2015). Selective suppression of the splicing-mediated MicroRNA pathway by the terminal uridyltransferase tailor. Mol. Cell 59, 217–228. https://doi.org/10.1016/j.molcel.2015. 05.034.
- Reimão-Pinto, M.M., Ignatova, V., Burkard, T.R., Hung, J.H., Manzenreither, R.A., Sowemimo, I., Herzog, V.A., Reichholf, B., Fariña-Lopez, S., and Ameres, S.L. (2015). Uridylation of RNA hairpins by tailor confines the emergence of MicroRNAs in Drosophila. Mol. Cell *59*, 203–216. https://doi.org/10.1016/j.molcel.2015.05.033.
- Liudkovska, V., and Dziembowski, A. (2021). Functions and mechanisms of RNA tailing by metazoan terminal nucleotidyltransferases. Wiley Interdiscip. Rev. RNA 12, e1622. https://doi.org/10.1002/wrna.1622.
- Yu, S., and Kim, V.N. (2020). A tale of non-canonical tails: gene regulation by post-transcriptional RNA tailing. Nat. Rev. Mol. Cell Biol. 21, 542–556. https://doi.org/10.1038/s41580-020-0246-8.
- Michlewski, G., and Cáceres, J.F. (2019). Post-transcriptional control of miRNA biogenesis. RNA 25, 1–16. https://doi.org/10.1261/ rna.068692.118.
- Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. (2003). A uniform system for microRNA annotation. RNA 9, 277–279.
- Kim, K., Baek, S.C., Lee, Y.Y., Bastiaanssen, C., Kim, J., Kim, H., and Kim, V.N. (2021). A quantitative map of human primary microRNA processing sites. Mol. Cell *81*, 3422–3439.e11. https://doi.org/10.1016/j. molcel.2021.07.002.
- Kang, W., Fromm, B., Houben, A.J., Høye, E., Bezdan, D., Arnan, C., Thrane, K., Asp, M., Johnson, R., Biryukova, I., and Friedländer, M.R. (2021). MapToCleave: high-throughput profiling of microRNA biogenesis in living cells. Cell Rep. 37, 110015. https://doi.org/10.1016/j.celrep. 2021.110015.
- Berezikov, E., Liu, N., Flynt, A.S., Hodges, E., Rooks, M., Hannon, G.J., and Lai, E.C. (2010). Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. Nat. Genet. 42, 6–9. https://doi.org/10.1038/ ng0110-6.
- Shang, R., Baek, S.C., Kim, K., Kim, B., Kim, V.N., and Lai, E.C. (2020). Genomic clustering facilitates nuclear processing of suboptimal primiRNA loci. Mol. Cell 78, 303–316.e4. https://doi.org/10.1016/j.molcel. 2020.02.009.
- Fang, W., and Bartel, D.P. (2020). MicroRNA clustering assists processing of suboptimal MicroRNA hairpins through the action of the ERH protein. Mol. Cell 78, 289–302.e6. https://doi.org/10.1016/j.molcel.2020. 01.026.
- Hutter, K., Lohmüller, M., Jukic, A., Eichin, F., Avci, S., Labi, V., Szabo, T.G., Hoser, S.M., Hüttenhofer, A., Villunger, A., and Herzog, S. (2020).
 SAFB2 enables the processing of suboptimal stem-loop structures in clustered primary miRNA transcripts. Mol. Cell 78, 876–889.e6. https:// doi.org/10.1016/j.molcel.2020.05.011.
- Chung, W.J., Agius, P., Westholm, J.O., Chen, M., Okamura, K., Robine, N., Leslie, C.S., and Lai, E.C. (2011). Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. Genome Res. *21*, 286–300. https://doi.org/10.1101/gr. 113050.110.
- Hansen, T.B., Venø, M.T., Jensen, T.I., Schaefer, A., Damgaard, C.K., and Kjems, J. (2016). Argonaute-associated short introns are a novel class of gene regulators. Nat. Commun. 7, 11538. https://doi.org/10. 1038/ncomms11538.
- Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N., and Kay, M.A. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. Cell *151*, 900–911. https://doi. org/10.1016/j.cell.2012.09.042.



- 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. https://doi.org/10.1038/ nature10198.
- 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. https://doi.org/10.1126/science.1190809.
- Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature 465, 584–589. https://doi.org/10.1038/nature09092.
- 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. https://doi.org/10.1073/pnas. 1006432107.
- Yang, J.S., Maurin, T., and Lai, E.C. (2012). Functional parameters of Dicer-independent microRNA biogenesis. RNA *18*, 945–957. https:// doi.org/10.1261/rna.032938.112.
- Maurin, T., Cazalla, D., Yang, S., Jr., Bortolamiol-Becet, D., and Lai, E.C. (2012). RNase III-independent microRNA biogenesis in mammalian cells. RNA 18, 2166–2173. https://doi.org/10.1261/rna.036194.112.
- Shang, R., Zhang, F., Xu, B., Xi, H., Zhang, X., Wang, W., and Wu, L. (2015). Ribozyme-enhanced single-stranded Ago2-processed interfering RNA triggers efficient gene silencing with fewer off-target effects. Nat. Commun. 6, 8430. https://doi.org/10.1038/ncomms9430.
- Liu, X., Zheng, Q., Vrettos, N., Maragkakis, M., Alexiou, P., Gregory, B.D., and Mourelatos, Z. (2014). A MicroRNA precursor surveillance system in quality control of MicroRNA synthesis. Mol. Cell 55, 868–879. https://doi. org/10.1016/j.molcel.2014.07.017.
- Kim, Y.K., Kim, B., and Kim, V.N. (2016). Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. Proc. Natl. Acad. Sci. USA *113*, E1881–E1889. https://doi.org/10.1073/pnas. 1602532113.
- Silverman, I., Gosai, S.J., Vrettos, N., Foley, S.W., Berkowitz, N.D., Mourelatos, Z., and Gregory, B.D. (2018). Isolation and sequencing of AGObound RNAs reveals characteristics of mammalian stem-loop processing in vivo. Preprint at bioRxiv. https://www.biorxiv.org/content/10. 1101/294488v294481.
- Burroughs, A.M., Kawano, M., Ando, Y., Daub, C.O., and Hayashizaki, Y. (2012). 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. Nucleic Acids Res. 40, 1424–1437. https://doi.org/10.1093/nar/gkr903.
- Golden, R.J., Chen, B., Li, T., Braun, J., Manjunath, H., Chen, X., Wu, J., Schmid, V., Chang, T.C., Kopp, F., et al. (2017). An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. Nature 542, 197–202. https://doi.org/10.1038/nature21025.
- Dexheimer, P.J., Wang, J., and Cochella, L. (2020). Two MicroRNAs are sufficient for embryonic patterning in C. elegans. Curr. Biol. 30, 5058– 5065.e5. https://doi.org/10.1016/j.cub.2020.09.066.
- Orlans, H.O., McClements, M.E., Barnard, A.R., Martinez-Fernandez de la Camara, C., and MacLaren, R.E. (2021). Mirtron-mediated RNA knockdown/replacement therapy for the treatment of dominant retinitis pigmentosa. Nat. Commun. *12*, 4934. https://doi.org/10.1038/s41467-021-25204-3.
- Seow, Y., Sibley, C.R., and Wood, M.J.A. (2012). Artificial mirtron-mediated gene knockdown: functional DMPK silencing in mammalian cells. RNA 18, 1328–1337. https://doi.org/10.1261/rna.030601.111.
- 55. Sibley, C.R., Seow, Y., Saayman, S., Dijkstra, K.K., El Andaloussi, S., Weinberg, M.S., and Wood, M.J.A. (2012). The biogenesis and charac-



terization of mammalian microRNAs of mirtron origin. Nucleic Acids Res. 40, 438–448. https://doi.org/10.1093/nar/gkr722.

- Mohammed, J., Flynt, A.S., Panzarino, A.M., Mondal, M.M.H., DeCruz, M., Siepel, A., and Lai, E.C. (2018). Deep experimental profiling of micro-RNA diversity, deployment, and evolution across the Drosophila genus. Genome Res. 28, 52–65. https://doi.org/10.1101/gr.226068.117.
- Hasler, D., Lehmann, G., Murakawa, Y., Klironomos, F., Jakob, L., Grässer, F.A., Rajewsky, N., Landthaler, M., and Meister, G. (2016). The lupus autoantigen La prevents Mis-channeling of tRNA fragments into the human MicroRNA pathway. Mol. Cell 63, 110–124. https://doi. org/10.1016/j.molcel.2016.05.026.
- Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A variety of dicer substrates in human and C. elegans. Cell *159*, 1153–1167. https://doi.org/10.1016/j.cell.2014.10.040.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141. https://doi.org/10.1016/j.cell.2010.03.009.
- Karginov, F.V., and Hannon, G.J. (2013). Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates. Genes Dev. 27, 1624–1632. https:// doi.org/10.1101/gad.215939.113.
- Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat. Methods 8, 559–564. https://doi.org/10.1038/nmeth.1608.
- Kim, B., Jeong, K., and Kim, V.N. (2017). Genome-wide mapping of DROSHA cleavage sites on primary MicroRNAs and Noncanonical substrates. Mol. Cell 66, 258–269.e5. https://doi.org/10.1016/j.molcel.2017. 03.013.
- Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Cáceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nat. Struct. Mol. Biol. 19, 760–766. https://doi.org/10. 1038/nsmb.2344.
- Sternburg, E.L., Estep, J.A., Nguyen, D.K., Li, Y., and Karginov, F.V. (2018). Antagonistic and cooperative AGO2-PUM interactions in regulating mRNAs. Sci. Rep. 8, 15316. https://doi.org/10.1038/s41598-018-33596-4.
- 65. Burroughs, A.M., Ando, Y., de Hoon, M.J.L., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2010). A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res. 20, 1398–1410. https:// doi.org/10.1101/gr.106054.110.
- 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. https://doi.org/10.1101/ gad.1884710.
- 67. 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. https://doi.org/10.1101/gr.116657.110.
- Thornton, J.E., Du, P., Jing, L., Sjekloca, L., Lin, S., Grossi, E., Sliz, P., Zon, L.I., and Gregory, R.I. (2014). Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). Nucleic Acids Res. 42, 11777– 11791. https://doi.org/10.1093/nar/gku805.
- 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. https://doi.org/10.1038/ nsmb.1676.

- 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 micro-RNA biogenesis through pre-microRNA uridylation. Cell *138*, 696–708.
- Thornton, J.E., Chang, H.M., Piskounova, E., and Gregory, R.I. (2012). Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). RNA *18*, 1875–1885. https://doi.org/10.1261/rna.034538.112.
- Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H., and Kim, V.N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell *151*, 521–532. https://doi. org/10.1016/j.cell.2012.09.022.
- Kim, H., Kim, J., Yu, S., Lee, Y.Y., Park, J., Choi, R.J., Yoon, S.J., Kang, S.G., and Kim, V.N. (2020). A mechanism for microRNA arm switching regulated by uridylation. Mol. Cell 78, 1224–1236.e5. https://doi.org/10. 1016/j.molcel.2020.04.030.
- Morgan, M., Much, C., DiGiacomo, M., Azzi, C., Ivanova, I., Vitsios, D.M., Pistolic, J., Collier, P., Moreira, P.N., Benes, V., et al. (2017). mRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. Nature 548, 347–351. https://doi.org/10.1038/nature23318.
- Westholm, J.O., Ladewig, E., Okamura, K., Robine, N., and Lai, E.C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. RNA *18*, 177–192. https://doi.org/10. 1261/rna.030627.111.
- Smibert, P., Yang, J.S., Azzam, G., Liu, J.L., and Lai, E.C. (2013). Homeostatic control of Argonaute stability by microRNA availability. Nat. Struct. Mol. Biol. 20, 789–795. https://doi.org/10.1038/nsmb.2606.
- Diederichs, S., and Haber, D.A. (2007). Dual role for argonautes in MicroRNA processing and posttranscriptional regulation of MicroRNA expression. Cell 131, 1097–1108.
- Yang, J.S., and Lai, E.C. (2010). Dicer-independent, Ago2-mediated microRNA biogenesis in vertebrates. Cell Cycle 9, 4455–4460.
- D'Ambrogio, A., Gu, W., Udagawa, T., Mello, C.C., and Richter, J.D. (2012). Specific miRNA stabilization by Gld2-catalyzed monoadenylation. Cell Rep. 2, 1537–1545. https://doi.org/10.1016/j.celrep.2012. 10.023.
- Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S.I., Baba, T., and Suzuki, T. (2009). Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes Dev. 23, 433–438. https://doi.org/10.1101/gad. 1761509.
- Vieux, K.F., Prothro, K.P., Kelley, L.H., Palmer, C., Maine, E.M., Veksler-Lublinsky, I., and McJunkin, K. (2021). Screening by deep sequencing reveals mediators of microRNA tailing in C. elegans. Nucleic Acids Res. 49, 11167–11180. https://doi.org/10.1093/nar/gkab840.
- Yang, A., Shao, T.J., Bofill-De Ros, X., Lian, C., Villanueva, P., Dai, L., and Gu, S. (2020). AGO-bound mature miRNAs are oligouridylated by TUTs and subsequently degraded by DIS3L2. Nat. Commun. *11*, 2765. https://doi.org/10.1038/s41467-020-16533-w.
- Yang, A., Bofill-De Ros, X., Stanton, R., Shao, T.J., Villanueva, P., and Gu, S. (2022). TENT2, TUT4, and TUT7 selectively regulate miRNA sequence and abundance. Nat. Commun. 13, 5260. https://doi.org/10. 1038/s41467-022-32969-8.
- Mansur, F., Ivshina, M., Gu, W., Schaevitz, L., Stackpole, E., Gujja, S., Edwards, Y.J.K., and Richter, J.D. (2016). Gld2-catalyzed 3' monoadenylation of miRNAs in the hippocampus has no detectable effect on their stability or on animal behavior. RNA 22, 1492–1499. https://doi.org/10. 1261/rna.056937.116.
- Medhi, R., Price, J., Furlan, G., Gorges, B., Sapetschnig, A., and Miska, E.A. (2022). RNA uridyl transferases TUT4/7 differentially regulate miRNA variants depending on the cancer cell type. RNA 28, 353–370. https://doi. org/10.1261/rna.078976.121.
- Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of



small silencing RNAs. Science 328, 1534–1539. https://doi.org/10. 1126/science.1187058.

- Kingston, E.R., and Bartel, D.P. (2019). Global analyses of the dynamics of mammalian microRNA metabolism. Genome Res. 29, 1777–1790. https://doi.org/10.1101/gr.251421.119.
- Han, J., LaVigne, C.A., Jones, B.T., Zhang, H., Gillett, F., and Mendell, J.T. (2020). A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. Science 370, eabc9546. https:// doi.org/10.1126/science.abc9546.
- Shi, C.Y., Kingston, E.R., Kleaveland, B., Lin, D.H., Stubna, M.W., and Bartel, D.P. (2020). The ZSWIM8 ubiquitin ligase mediates targetdirected microRNA degradation. Science 370, eabc9359. https://doi. org/10.1126/science.abc9359.
- Kroupova, A., Ivascu, A., Reimão-Pinto, M.M., Ameres, S.L., and Jinek, M. (2019). Structural basis for acceptor RNA substrate selectivity of the 3' terminal uridylyl transferase Tailor. Nucleic Acids Res. 47, 1030– 1042. https://doi.org/10.1093/nar/gky1164.
- Cheng, L., Li, F., Jiang, Y., Yu, H., Xie, C., Shi, Y., and Gong, Q. (2019). Structural insights into a unique preference for 3' terminal guanine of mirtron in Drosophila TUTase tailor. Nucleic Acids Res. 47, 495–508. https:// doi.org/10.1093/nar/gky1116.
- Fang, W., and Bartel, D.P. (2015). The menu of features that define primary MicroRNAs and enable De Novo design of MicroRNA genes. Mol. Cell 60, 131–145. https://doi.org/10.1016/j.molcel.2015.08.015.
- Kwon, S.C., Baek, S.C., Choi, Y.G., Yang, J., Lee, Y.S., Woo, J.S., and Kim, V.N. (2019). Molecular basis for the single-nucleotide precision of primary microRNA processing. Mol. Cell 73, 505–518.e5. https://doi. org/10.1016/j.molcel.2018.11.005.
- Zapletal, D., Taborska, E., Pasulka, J., Malik, R., Kubicek, K., Zanova, M., Much, C., Sebesta, M., Buccheri, V., Horvat, F., et al. (2022). Structural and functional basis of mammalian microRNA biogenesis by Dicer. Mol. Cell 82, 4064–4079.e13. https://doi.org/10.1016/j.molcel.2022. 10.010.
- 95. Van Valen, L. (1973). A new evolutionary law. Evol. Theor. 1, 1-30.
- Fueyo, R., Judd, J., Feschotte, C., and Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. Nat. Rev. Mol. Cell Biol. 23, 481–497. https://doi.org/10.1038/s41580-022-00457-y.
- Neal, S., Ohtake, F., Cuervo, A.M., Hegde, R.S., Jakob, U., Lazarou, M., Gilbert, W.V., Chen, Z.J., Tooze, S.A., Haber, J.E., et al. (2022). Quality control: maintaining molecular order and preventing cellular chaos. Mol. Cell 82, 1390–1397. https://doi.org/10.1016/j.molcel.2022.04.002.
- Hacquard, T., Clavel, M., Baldrich, P., Lechner, E., Pérez-Salamó, I., Schepetilnikov, M., Derrien, B., Dubois, M., Hammann, P., Kuhn, L., et al. (2022). The Arabidopsis F-box protein FBW2 targets AGO1 for degradation to prevent spurious loading of illegitimate small RNA. Cell Rep. 39, 110671. https://doi.org/10.1016/j.celrep.2022.110671.
- Gy, I., Gasciolli, V., Lauressergues, D., Morel, J.B., Gombert, J., Proux, F., Proux, C., Vaucheret, H., and Mallory, A.C. (2007). Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. Plant Cell 19, 3451–3461. https://doi.org/10.1105/tpc.107.055319.
- 100. You, C., He, W., Hang, R., Zhang, C., Cao, X., Guo, H., Chen, X., Cui, J., and Mo, B. (2019). FIERY1 promotes microRNA accumulation by suppressing rRNA-derived small interfering RNAs in Arabidopsis. Nat. Commun. 10, 4424. https://doi.org/10.1038/s41467-019-12379-z.
- 101. Tu, B., Liu, L., Xu, C., Zhai, J., Li, S., Lopez, M.A., Zhao, Y., Yu, Y., Ramachandran, V., Ren, G., et al. (2015). Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in Arabidopsis. PLoS Genet. *11*, e1005119. https://doi.org/10.1371/journal. pgen.1005119.
- Wang, X., Zhang, S., Dou, Y., Zhang, C., Chen, X., Yu, B., and Ren, G. (2015). Synergistic and independent actions of multiple terminal



CelPress

- Scheer, H., de Almeida, C., Ferrier, E., Simonnot, Q., Poirier, L., Pflieger, D., Sement, F.M., Koechler, S., Piermaria, C., Krawczyk, P., et al. (2021). The TUTase URT1 connects decapping activators and prevents the accumulation of excessively deadenylated mRNAs to avoid siRNA biogenesis. Nat. Commun. *12*, 1298. https://doi.org/10.1038/s41467-021-21382-2.
- Vigh, M.L., Bressendorff, S., Thieffry, A., Arribas-Hernández, L., and Brodersen, P. (2022). Nuclear and cytoplasmic RNA exosomes and PELOTA1 prevent miRNA-induced secondary siRNA production in Arabidopsis. Nucleic Acids Res. 50, 1396–1415. https://doi.org/10.1093/nar/ gkab1289.
- 105. Song, J., Wang, X., Song, B., Gao, L., Mo, X., Yue, L., Yang, H., Lu, J., Ren, G., Mo, B., and Chen, X. (2019). Prevalent cytidylation and uridylation of precursor miRNAs in Arabidopsis. Nat. Plants 5, 1260–1272. https://doi.org/10.1038/s41477-019-0562-1.
- Pisacane, P., and Halic, M. (2017). Tailing and degradation of Argonautebound small RNAs protect the genome from uncontrolled RNAi. Nat. Commun. 8, 15332. https://doi.org/10.1038/ncomms15332.
- 107. Kim, B., Ha, M., Loeff, L., Chang, H., Simanshu, D.K., Li, S., Fareh, M., Patel, D.J., Joo, C., and Kim, V.N. (2015). TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. EMBO J. 34, 1801–1815. https://doi.org/10.15252/embj.201590931.
- 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.
- 109. Piskounova, E., Viswanathan, S.R., Janas, M., Lapierre, R.J., Daley, G.Q., Sliz, P., and Gregory, R.I. (2008). Determinants of microRNA pro-

cessing inhibition by the developmentally regulated RNA-binding protein Lin28. J. Biol. Chem. 283, 21310–21314.

- 110. Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97–100.
- 111. 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.
- 112. 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.
- 113. Chang, H.M., Triboulet, R., Thornton, J.E., and Gregory, R.I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature 497, 244–248. https://doi.org/10.1038/nature12119.
- Ustianenko, D., Hrossova, D., Potesil, D., Chalupnikova, K., Hrazdilova, K., Pachernik, J., Cetkovska, K., Uldrijan, S., Zdrahal, Z., and Vanacova, S. (2013). Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. RNA *19*, 1632–1638. https://doi.org/10. 1261/rna.040055.113.
- 115. Yang, A., Bofill-De Ros, X., Shao, T.J., Jiang, M., Li, K., Villanueva, P., Dai, L., and Gu, S. (2019). 3' uridylation confers miRNAs with non-canonical target repertoires. Mol. Cell 75, 511–522.e4. https://doi.org/10.1016/ j.molcel.2019.05.014.
- 116. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.
- 117. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNAsequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007.







STAR*METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
EIF2C2 monoclonal antibody (M01), clone 2E12-1C9	Abnova	Cat#H00027161-M01; RRID: AB_565459
Chemicals, peptides, and recombinant proteins		
EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride)	Thermo Fisher Scientific	Cat#22980
Critical commercial assays		
Lipofectamine 2000	Thermo Fisher Scientific	Cat#15338030
ATP, [gamma-P32]	PerkinElmer	Cat#BLU502Z250UC
T4 PNK	NEB	Cat#M0201L
SequaGel UreaGel System	National Diagnostics	Cat#EC-833
Novex [™] TBE-Urea Gels, 6%, 10 well	Invitrogen	Cat#EC6865BOX
Dual Glo luciferase assay system	Promega	Cat#E2940
Trizol	Life Technologies	Cat#15596-018
GeneScreen Plus membrane	PerkinElmer	Cat#NEF1017001PK
T4 RNA ligase 2 (1-249; K227Q)	NEB	Cat#M0351L
T4 RNA ligase 1	NEB	Cat#M0204L
Superscript III RT kit	Invitrogen	Cat#18080-051
Dynabeads Protein G	Invitrogen	Cat#10004D
CircLigase II ssDNA Ligase	Epicentre	Cat#CL9021K
Deposited data		
Intermediate-sized RNA data from K562 cells and small RNA-seq data from control, TUT4/7-dKO and TENT2/TUT4/7-tKO HEK293T cells.	This study	GSE206186
Experimental models: Cell lines		
TENT2-KO	Shuo Gu	N/A
TUT4-KO	Shuo Gu	N/A
TUT7-KO	Shuo Gu	N/A
TUT4/7-dKO	Shuo Gu	N/A
TUT4/7+TENT2-tKO	Shuo Gu	N/A
Oligonucleotides		
Refer to Table S7	N/A	N/A
Recombinant DNA		
pcDNA6.2-N-terminal EmGFP-TOPO	Invitrogen	Cat#360-20
pcDNA3-Myc-Ago2	Jidong Liu	N/A
pcDNA6.2-hsa-miR-1226	This study	N/A
pcDNA6.2-hsa-miR-375	This study	N/A
pcDNA3-mmu-miR-1225/PKD1	Jørgen Kjems ³⁸	N/A
pcDNA3-mmu-ACADVL	Jørgen Kjems ³⁸	N/A
psiCHECK-miR1225/PKD1-5p	Jørgen Kjems ³⁸	N/A
psiCHECK-ACADVL-5p	This study	N/A
psiCHECK-ACADVL-3p	This study	N/A
pCK-FLAG-TENT1	V. Narry Kim	N/A
pCK-FLAG-TENT2	V. Narry Kim	N/A
pCK-FLAG-TENT4B	V. Narry Kim	N/A

(Continued on next page)

CellPress OPEN ACCESS

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
pCK-FLAG-TENT4	V. Narry Kim	N/A
pCK-FLAG-TENT4A	V. Narry Kim	N/A
pCK-FLAG-TENT6	V. Narry Kim	N/A
pCK-FLAG-TUT7	V. Narry Kim	N/A
Software and algorithms		
Prism 9	GraphPad	N/A
BBTools	JGI	N/A
FASTX-Toolkit	Hannon Lab	N/A
Bowtie	Salzberg Lab Langmead et al. ¹¹⁶	N/A
Limma	Smyth Lab Ritchie et al.117	N/A

RESOURCE AVAILABILITY

Lead contact

Requests for materials and bioinformatic methods should be directed the lead contact, Eric Lai (laie@mskcc.org).

Materials availability

All plasmids generated for this study are available upon request.

Data and code availability

- All of the raw sequencing data for K562 Ago2-IP intermediate data, and replicate libraries of HEK293T, TUT4/7-dKO and TENT2/TUT4/7-tKO small RNA data, were deposited in NCBI Gene Expression Omnibus under accession GSE206186.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The HEK293T and K562 cell lines have been used in this study with standard culture condition (DMEM culture medium supplemented with 10% fetal bovine serum, FBS).

METHOD DETAILS

Processing and curation of human small RNA, intermediate-sized RNA, and CLIP-seq data

We collected 756 small RNA, 8 intermediate-length RNA, and 255 CLIP-seq datasets from various tissues/cell lines from NCBI GEO/ SRA. The accession numbers of datasets analyzed in this study are summarized in Table S1. Amongst these, 180 human small RNA datasets were added since our previous meta-analysis.²² Of particular importance to the annotation of splicing-derived RNAs in this study were datasets of human intermediate-length RNAs, generated by sequencing of Ago2-associated RNAs from HEK293T cells⁴⁹ or from HeLa total RNAs depleted of abundant ncRNA species using antisense LNAs.⁵⁰

After removing 3' adaptor sequences, we mapped small RNA reads to human (hg38) genome assembly. Unmapped reads were iteratively trimmed one nucleotide each iteration retaining a read length of \geq 15 nt, and then mapped to the genome using Bowtie with no mismatches, up to 5 iterations. After removing both 5' and 3' adaptor sequences, we mapped intermediate-length RNA reads and CLIP-seq reads to human (hg38) genome assembly using Biowtie2 with the default parameters. The uniquely mapped reads were extracted by "samtools view -bq 20" and the coverage normalized by library sizes were generated for the downstream analysis.

Intermediate-length RNA Ago2-IP library from K562 cells

Lentiviral particles for control shRNA were produced in HEK293T cells and 5 ml cell culture supernatants containing lentivirus were added to K562 cells grown in 60 mm dishes. Four days after virus infection and puromycin selection, cells were collected for Ago2 IP using Ago2 antibody (Abnova) and the RNA products were purified using Trizol and PAGE-size selected for the intermediate fragments between 40 nt and 80 nt. These intermediates are ligated to 3' adapter using T4 RNA Ligase 2 (truncated KQ) and then purified for reverse transcription by barcoded primer containing complementary sequences to the 3' adaptor on the RNA. Following, the cDNA products are purified on PAGE and circularized using CircLigase II, annealed to the Cut Oligo and digested using BamHI. Finally, the library is amplified from the linear cDNAs using Phusion High-Fidelity PCR master mix with P5/P3 Solexa primers for



20 cycles. The PCR products are purified on Novex 8% TBE gel and send for sequencing with HiSeq 4000 in PE100 mode. Oligo sequences used for shRNA transduction and library preparation are provided in Table S7.

Small RNA libraries from HEK293T cells

For small RNA analysis, we extracted biologically independent total RNA samples from HEK293T, TUT4/7-dKO and TENT2/TUT4/7-tKO cells using Trizol (Invitrogen) according to the manufacturer's instructions. 1 μ g of total RNA were used to generate small RNA libraries for each sample. After adding 1 μ L of QIAseqTM miRNA Library QC Spike-Ins (QIAGEN) to total RNA, 3' linker was ligated to total RNA in the reaction containing 100 μ M 3' linker, T4 RNA Ligase 2, truncated KQ (NEB), 10% PEG8000, 1X RNA ligation buffer, and 20 U RNase OUT overnight at 17°C. Using radio-labeled oligonucleotides as size-markers, 48-57 nucleotide sized 3' linker-ligated small RNAs were purified by 12% Urea-PAGE gel. The small RNA-linker hybrid was then subjected to 5' ligation reaction overnight at 17°C containing 100 μ M 5' adaptor, 20% PEG8000, 1X RNA ligase buffer, 1 mM ATP, 20U RNaseOUT and 5U T4 RNA ligase 1. The linker ligated RNAs were reverse transcribed, and PCR amplified using 12 cycles of PCR with forward and Illumina index reverse primers and the amplified libraries were purified by 8% non-denaturing acrylamide gel. Sequencing was performed with HiSeq 4000 in PE50 mode. Oligo sequences used for library preparation are provided in Table S7.

Each of the small RNA sequencing libraries were processed by merging two overlapping paired reads into a single read using BBMerge, and the adaptor sequences were subsequently removed. The reads were then collapsed to remove PCR artifacts using fastx_collapser, and the 5' and 3' end linker sequences were clipped. We then mapped small RNA-seq reads from these datasets to the hg38 genome assembly. Unmapped reads were iteratively trimmed one nucleotide each iteration retaining a read length of 15 nt, and then mapped to the genome using Bowtie with no mismatches, up to 30 iterations. The reads were required to match the micro-RNA mature sequences of both canonical miRNAs and mirtrons.²³ with at least 15 nt overlap and within 2 nt of the 5' end. The reads mapped to the spike-in sequences were counted. We normalized small RNA-seq by total spike-in reads in each library. We performed small RNA differential expression analysis between TUTase double and triple mutants against wild type using Bioconductor R package Limma.

ssdRNA identification and annotation

The sequencing coverage bigwig files from intermediate-length RNA libraries were merged. We applied a custom method to call peaks genome-wide as follows. We first called peaks genomewide with different coverage cutoffs: ≥ 2 RPM, ≥ 5 RPM, and ≥ 10 RPM. The peaks were combined from different cutoffs and merged to keep the unique peaks by filtering out the peaks with lower cutoffs if they overlapped with those with higher cutoffs. The called peaks were annotated using Ensembl gene annotation GRCh38 (v100), and the peaks with a distance of ≤ 10 nt to the intron ends and ≥ 30 intermediate-length RNA reads were retained. The known RNAs and mirtrons, as well as the reads from sRNA, intermediate-size, AGO CLIP-seq and DICER CLIP-seq datasets, were further used to annotate these peaks.

Since we sought to annotate a new RNA class without previously known concrete attributes, we inspected the read patterns at of the initially 1000s of intronic loci by hand to derive the confident set of ssdRNAs. These were regrouped and their attributes were set to exceed these cutoffs: Minimum total intermediate reads \geq 30 reads; MFE /base \leq -0.08 (MFE \leq -9); Hairpin distance to intron ends (not including tailed reads): \leq 4 nt. Many loci fall short of these cutoffs but appear similar to annotated ssdRNAs; it is likely that additional ssdRNA would be confident with further intermediate-sized RNA data or other cell/tissue profiling.

Analysis of 3' untemplated additions

Mirtron and ssdRNA hairpin boundaries were defined on the basis of intron splice donor (GT) or acceptor (AG) sequences. The derivation of mirtron-3p and ssdRNA-3p reads from splicing permitted us to utilize the "AG" splice acceptor as an external reference for the primary-processed species. Thus, reads that extended past the "AG" likely bear untemplated nucleotides, and were called as such regardless of whether they match the genome or not. The boundaries of canonical pre-miRNA hairpins were based on miRBase v20 with sequences defined by the most abundant matching read from all aggregated small RNA libraries analyzed in this study.

The reads were required to match to the defined mirtron, ssdRNA and canonical miRNA sequences with at least 15 nt overlap, and within 2 nt of the 5' end. Trimmed reads were required to match within 4 nt of the defined 3' end, with no restriction on 3' end of tailed reads in order to analyze tailing events. The trimmed reads were not considered for the tailing analysis.

For comparison, we used bulk intron ends as controls, for which we removed those containing identified mirtrons and having a minimum free energy (MFE) of the 150nt window from intron 3p ends less than 90% percentile of mirtron MFEs. To analyze bulk intron-3p reads, we selected human and mouse introns with \geq 1 AG terminal read, for which any longer reads terminated 1 or 2 nt downstream. This filter therefore discarding all loci for which intron annotations may be spurious, including loci that exhibit a probable signature of degradation reads across an annotated site; reads from these loci are expected to actually be genome-derived as opposed to bearing 3' untemplated additions beyond the AG splice acceptor.

To quantify mono-nucleotide tailings, we calculated the percentage of reads that extended exactly one nucleotide beyond the defined miRNA sequence or bulk intron end, relative to the total number of reads (mono-nucleotide tailed reads and unmodified reads). t -test was used to measure statistical significance of mean frequency of mono-nucleotide tailings among mirtrons, canonical miRNAs, and bulk intron ends.

Analysis of miRNA evolution

For canonical miRNA-3p species, we searched the sequence conservation of a conserved 7-mer seed region to be conserved across UCSC human or *Drosophila* multiple alignments. For human, we required a conservation PhyloP score of a miRNA greater than 1.3 to be conserved and less than 0 to be non-conserved. For *Drosophila*, we required conserved miRNAs from D. melanogaster to D. pseudoobscura, and non-conserved miRNAs to be only in melanogaster group species.

Small RNA Northern blotting

10–50 μ g of total RNA were resolved on 15% Urea acrylamide gel (SequaGel UreaGel System, National Diagnostics) in 1X TBE buffer, and transferred to GeneScreen Plus membrane (PerkinElmer). After UV-crosslinking or EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide)-mediated chemical crosslinking, the membrane was hybridized with 5' end labeled oligonucleotides with (γ -32P) ATP for overnight at 42°C. The membrane was exposed to phosphorimaging plate (Fujifilm) and read by BAS-3000 system (Fujifilm). Sequences of Northern probes are listed in Table S7.

Luciferase sensor assays

1-40 and 46–85 position of Acadvl hairpin sequences were PCR amplified from pcDNA3-mmu-ACADVL and cloned into psiCHECK2 luciferase sensors, to generate Acadvl-5p and Acadvl-3p sensor, respectively. Primers used for cloning are listed in Table S7. Mirtron overexpression vectors for mouse Acadvl and mir-1225/Pkd1, and miR-1225/Pkd1 sensor, were published previously.³⁸

To perform sensor assay, 1×10^5 HEK293T cells were seeded per well of 96-well plate and transfected with 50ng mirtron-overexpression constructs, 50ng psiCHECK2 sensors and 15ng Ago2-overexpression construct. Luciferase were measured 2 days after transfection using Dual Glo luciferase assay system (Promega) and Cytation5 (BioTek) according to the manufacturer's instructions. The fold repression values were normalized to luciferase activity measured by expression of the sensor co-transfected with noncognate target mirtron. Individual tests were done in quadruplicate and the averaged values from three biological replicate samples were subjected to statistical analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

For sensor assays presented in Figures 4D and 5D, unpaired Student's t test was used to evaluate significance of comparisons between repression activity of mirtrons against their sensors in different genetic backgrounds. The error bars represent standard deviation and p-values are denoted above bars in the figures above.

For quantification of Northern blot signals represented in Figure 6F, mean ratios of either the ssdRNA-ANKRD52 or miR-16 from triplicate experiments were plotted and unpaired Student's t test was performed to evaluate significance in accumulation patterns of ssdRNAs under the denoted mutant backgrounds.



