Antisense Transcription of Retrotransposons in *Drosophila*: An Origin of Endogenous Small Interfering RNA Precursors

Joseph Russo,[1,2] Andrew W. Harrington,[1] and Mindy Steiniger[3]
Department of Biology, University of Missouri, St. Louis, Missouri 63121

ABSTRACT Movement of transposons causes insertions, deletions, and chromosomal rearrangements potentially leading to premature lethality in *Drosophila melanogaster*. To repress these elements and combat genomic instability, eukaryotes have evolved several small RNA-mediated defense mechanisms. Specifically, in *Drosophila* somatic cells, endogenous small interfering (esi)RNAs suppress retrotransposon mobility. EsiRNAs are produced by Dicer-2 processing of double-stranded RNA precursors, yet the origins of these precursors are unknown. We show that most transposon families are transcribed in both the sense (S) and antisense (AS) direction in Dmel-2 cells. LTR retrotransposons Dm297, mdg1, and blood, and non-LTR retrotransposons juan and jockey transcripts, are generated from intraelement transcription start sites with canonical RNA polymerase II promoters. We also determined that retrotransposon antisense transcripts are less polyadenylated than sense. RNA-seq and small RNA-seq revealed that Dicer-2 RNA interference (RNAi) depletion causes a decrease in the number of esiRNAs mapping to retrotransposons and an increase in expression of both S and AS retrotransposon transcripts. These data support a model in which double-stranded RNA precursors are derived from convergent transcription and processed by Dicer-2 into esiRNAs that silence both sense and antisense retrotransposon transcripts. Reduction of sense retrotransposon transcripts potentially lowers element-specific protein levels to prevent transposition. This mechanism preserves genomic integrity and is especially important for *Drosophila* fitness because mobile genetic elements are highly active.

KEYWORDS antisense; double-stranded RNA; *Drosophila*; convergent transcription; Dicer-2

MOBILE genetic elements are one source of genetic alterations that drive evolution, but can also lead to catastrophic genomic instability. Thus, maintaining an appropriate balance between the potential harm and benefit of transposons (Tns) is vital. If active Tns are not adequately controlled by their hosts, mutations produced by their movement can be detrimental (Lee and Marx 2013). Specifically in *Drosophila*, genetic rearrangements that cause hybrid digenesis syndrome (Kidwell *et al.* 1977; Picard *et al.* 1978) are linked to transposon movement (Bingham *et al.* 1982; Rubin *et al.* 1982).

doi: 10.1534/genetics.115.177196
Manuscript received April 9, 2015; accepted for publication October 23, 2015; published Early Online October 29, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177196/-/DC1.
[1]These authors contributed equally to this work.
[2]Present address: Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523
[3]Corresponding author: 1 University Blvd., University of Missouri, St. Louis, MO 63121. E-mail: steinigerm@umsl.edu

Since the discovery of Tns by Barbara McClintock more than 60 years ago (McClintock 1950), researchers have elucidated key mechanisms describing how Tns incorporate into genomes and how hosts combat these potentially toxic genomic perturbations. However, many aspects of Tn biology remain elusive. While ~44% of the human genome is composed of Tns (Cordaux and Batzer 2009), there is little diversity in active transposons (Mills *et al.* 2007); only autonomous LINE-1 and nonautonomous Alu and SVA retrotransposons are currently mobile (Brouha *et al.* 2003; Cordaux and Batzer 2009; Deininger 2011). While the *Drosophila* genome is only ~22% transposons, many (~30%) of these elements are full length and thought to be active (Kaminker *et al.* 2002; Lerat *et al.* 2003; Kofler *et al.* 2015). Having active transposons from all three major classes of mobile elements to investigate offers a unique opportunity to understand silencing mechanisms in eukaryotic organisms.

Tns are defined by their approach to mobility. Terminal inverted repeat (TIR) Tns encode a Transposase that binds Tn

inverted repeats (in most cases), creates double-strand breaks at the ends of the Tn, and integrates the Tn into a new genomic location. This mechanism can create genomic rearrangements such as insertions, deletions, and inversions. Unlike TIR Tns, retrotransposons (retroTns) include an RNA intermediate in their movement mechanism and therefore encode a reverse transcriptase (RT). RetroTns are divided into long terminal repeat (LTR) and non-LTR retroTns. LTR retroTns are similar to retroviruses and contain several hundred nucleotide terminal repeats at both the 5′ and 3′ ends (Figure 1A). While some *Drosophila* LTR retroTns have *gag* and *env* genes homologous to retroviruses, others have more divergent ORFs that function in retroTn mobility (Figure 1). Non-LTR retroTns lack these terminal repeats and sequences homologous to the *env* gene (Figure 2A), but have conserved RTs (Figure 2). Both LTR and non-LTR retroTns often have an internal promoter located in the 5′ untranslated region (UTR) and a 3′ UTR containing a polyadenylation signal (Gogvadze and Buzdin 2009 and this work). The initial transposition step for all retroTns is RNA polymerase II (RNAPII)-dependent transcription of the entire element followed by translation of each independent ORF in different reading frames from this single, polygenic messenger RNA (mRNA).

Eukaryotic cells have evolved several noncoding RNA-mediated mechanisms to control further genomic spread of retroTns. In humans, mobility of the LINE-1 (L1) retroTn is regulated by both canonical RNA interference (RNAi) (Yang and Kazazian 2006) and endogenous small interfering (esi)RNA-mediated chromatin modifications (Chen *et al.* 2012). Similarly, in *Drosophila*, two distinct RNAi-like processes for silencing Tns have been elucidated. In the germline, the Piwi-interacting RNA (piRNA) pathway generates small RNAs that suppress Tns by inducing heterochromatin formation (Vagin *et al.* 2006; Aravin *et al.* 2007; Brennecke *et al.* 2007; Sentmanat and Elgin 2012; Le Thomas *et al.* 2013). In somatic cells, esiRNAs silence retroTns via a Dicer 2 (Dcr-2)/Argonaute 2 (Ago2)-dependent mechanism (Chung *et al.* 2008; Czech *et al.* 2008; Ghildiyal *et al.* 2008; Saito and Siomi 2010; Xie *et al.* 2013). Global analysis of small RNA libraries generated from embryo-derived *Drosophila* somatic cells (S2) (Schneider 1972) showed that 86% of esiRNAs mapped to Tns; esiRNAs mapping to LTR retroTns were highly enriched (Ghildiyal *et al.* 2008). Dcr-2 is required for generation of esiRNAs (Czech *et al.* 2008; Okamura *et al.* 2008a,b) and retroTn expression increases following RNAi depletion of Dcr-2 (Ghildiyal *et al.* 2008; Marques *et al.* 2010).

The production of esiRNAs by Dcr-2 requires a double-stranded RNA (dsRNA) precursor (Tomari *et al.* 2007; Ghildiyal *et al.* 2008; Marques *et al.* 2010). While dsRNAs generated by hybridization of natural antisense transcripts and their sense counterparts are substrates for Dcr-2 in *Drosophila* (Czech *et al.* 2008; Okamura *et al.* 2008a), retroTn dsRNA precursors have not been systematically investigated. As *Drosophila* does not encode an RNA-dependent RNA polymerase to generate a complementary strand, the origin of the antisense (AS) transcript necessary to form the dsRNA retroTn precursor is unknown. Here, we provide evidence that both non-LTR and LTR retroTns produce sense (S) and AS transcripts from intraelement transcription start sites (tss) with canonical *Drosophila* promoters. We then use a novel polyA+/− fractionation followed by strand-specific RT-qPCR technique to show that most S and AS retroTn transcripts are not enriched for polyadenylation. Finally, increases in AS retroTn transcript levels in Dmel-2 cells RNAi depleted of Dcr-2 indicate that AS and S transcripts are substrates for Dcr-2.

Materials and Methods

Large and small RNA preparation/PolyA+ RNA selection

Total RNA from 8×10^6 *Drosophila* Dmel-2 tissue culture cells was isolated using QIAzol Lysis Reagent (Qiagen). Total RNA was fractionated into large (>200 nt) and small (<200 nt) fractions using RNeasy Mini spin columns and RNeasy MinElute spin columns, respectively (Qiagen). DNA was removed from the large fraction by on-column DNase digestion (Qiagen). Fractionation and DNA removal were verified by RT-qPCR. RNA integrity and size fractionation were confirmed using small RNA and RNA 6000 Pico Bioanalyzer chips (Agilent). Total RNA was then fractionated into polyA+ and polyA− fractions using the MicroPoly(A) Purist Kit (Ambion AM1919). Fractionation was verified by RT-qPCR.

Ribosomal RNA depletion

The 28S, 18S, and 5S ribosomal RNAs (rRNAs) were depleted from 5 μg of each large RNA fraction using the Ribo-Zero Magnetic Kit (Epicentre). While this kit was designed for human/mouse/rat, it performs adequately for *Drosophila*. rRNA depletion was confirmed by RT-qPCR and validated using RNA 6000 Pico Bioanalyzer chips (Agilent).

The 2S rRNA was depleted from the small RNA fraction according to Seitz *et al.* (2008) with the following modifications: 0.1 nM 2S rRNA complementary oligo was bound to 500 μg streptavidin beads in 1 ml 0.5× SSC for 1 hr at 4°. The beads were then washed five times in 0.5× SSC followed by a 5-min incubation at 65° to remove secondary structure. A total of 2 μg of the small RNA fraction was diluted to 12.5 ng/μl and 160 μl was added to the bead slurry. The remaining steps of the protocol were as described (Seitz *et al.* 2008). Following rRNA depletion from both small and large RNA fractions, RNA integrity, and rRNA depletion were validated on a Bioanalyzer.

Library preparation/next-generation sequencing

RNA-seq libraries were prepared in triplicate from 35 ng of the rRNA-depleted large RNA fraction using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). Small RNA sequencing (smRNA-seq) libraries were prepared in triplicate from ~475 ng of the 2S rRNA-depleted small RNA fraction using the NEBNext Small RNA Library Prep Set for Illumina (NEB). Each small interfering RNA (siRNA)- and RNA-seq library was amplified with a primer having a

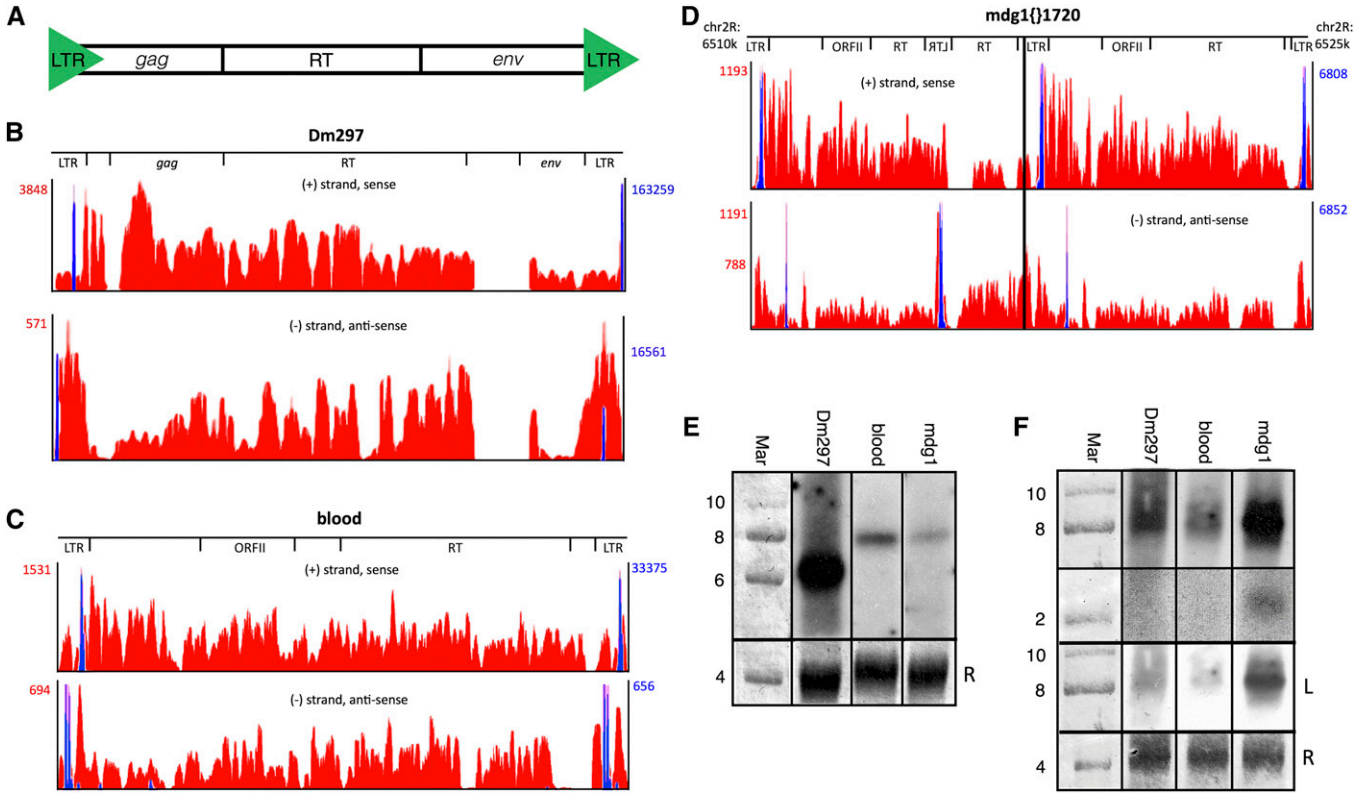


Figure 1 LTR retroTns Dm297, blood, and mdg1{}1720 produce AS transcripts from intraelement tss in or near the LTRs. (A) Schematic of *Drosophila* LTR retroTns. (B–D) Bedgraphs representing S (top) and AS (bottom) nonunique RNA-seq reads mapping to each LTR retroTn are shown in red. Peak reads per million (RPM) are listed to the left (red numbers). For mdg1, two AS RPM values are listed; the top is the RPM for mdg1{}1720 and the bottom is the RPM for only the downstream canonical mdg1 element (right of the black line). Only the chromosome location of mdg1{}1720 is shown as Dm297 and blood bedgraphs are representative examples. Relative locations of specific ORFs are shown above the bedgraphs. Nonunique small-capped RNA-seq reads representing tss are overlaid in blue and RPM values are listed to the right (blue numbers). (E) Representative Northern blots of S LTR retroTn transcripts. The probe used for each blot is indicated above. The first lane is methylene blue-stained RNA marker; the sizes of bands are shown to the left of the blots. Methylene blue-stained 28S rRNA is used as a loading control (bottom) and is marked with an "R." (F) Representative Northern blots of AS LTR retroTn transcripts. The top two panels are from the same longer exposure film while the third panel ("L") is a lighter exposure. Other details are as in E.

unique barcode. The appropriate size of each library was validated on a Bioanalyzer using a high sensitivity DNA chip (Agilent) and quantitated using the Qubit dsDNA BR Assay Kit (Molecular Probes) according to the manufacturer's instructions. All siRNA-seq libraries were multiplexed and sequenced in one flow cell using a MiSeq and MiSeq Reagent Kit v2 (50-cycle) (Illumina). RNA-seq libraries were multiplexed and sequenced in two HiSeq lanes by the Genome Access Technology Center (GATC) at Washington University.

Next-generation sequencing data analyses

All adapter sequences were trimmed and the libraries cleaned using Cutadapt (Martin 2011). We aggressively trimmed the siRNA reads to 25 nt from the 5′ end following adapter removal to filter remaining rRNAs, small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and transfer RNAs (tRNAs) out of the dataset before mapping. All datasets were mapped to the *Drosophila melanogaster* genome and transcriptome using the RNA-seq Unified Mapper (RUM) (Grant *et al.* 2011). The NEB kit used to prepare the RNA-seq samples produces libraries with high directionality and RUM utilized this feature to strand specifically map the RNA-seq reads. RUM separated unique and nonuniquely mapping sequences into separate output files that could be further analyzed. Detailed mapping statistics can be found in Supporting Information, Table S1.

The University of California Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu, Dm6 assembly, August 2014) was used to visualize nonunique and unique bedgraph output files (Kent *et al.* 2002; dos Santos *et al.* 2015). The genome browser displays a peak normalized read count (reads per million, RPM) on the *y*-axis for the visualized genomic location. Nonunique RPM were used to calculate the average S and AS reads for Tn families shown in Table 1, and S and AS transcription of individual Tns having three or more full-length elements (Table S3). For individual Tns, a "+" system was devised to represent relative transcription among the Tns (Table S3). Unique RPM were used to measure S and AS expression levels of protein coding genes (Table S2).

Small-capped RNA-seq datasets (SRA: SRP001584, SRR032457, and SRR032458) were obtained from the Gene Expression Omnibus (GEO) accession number GSE18643

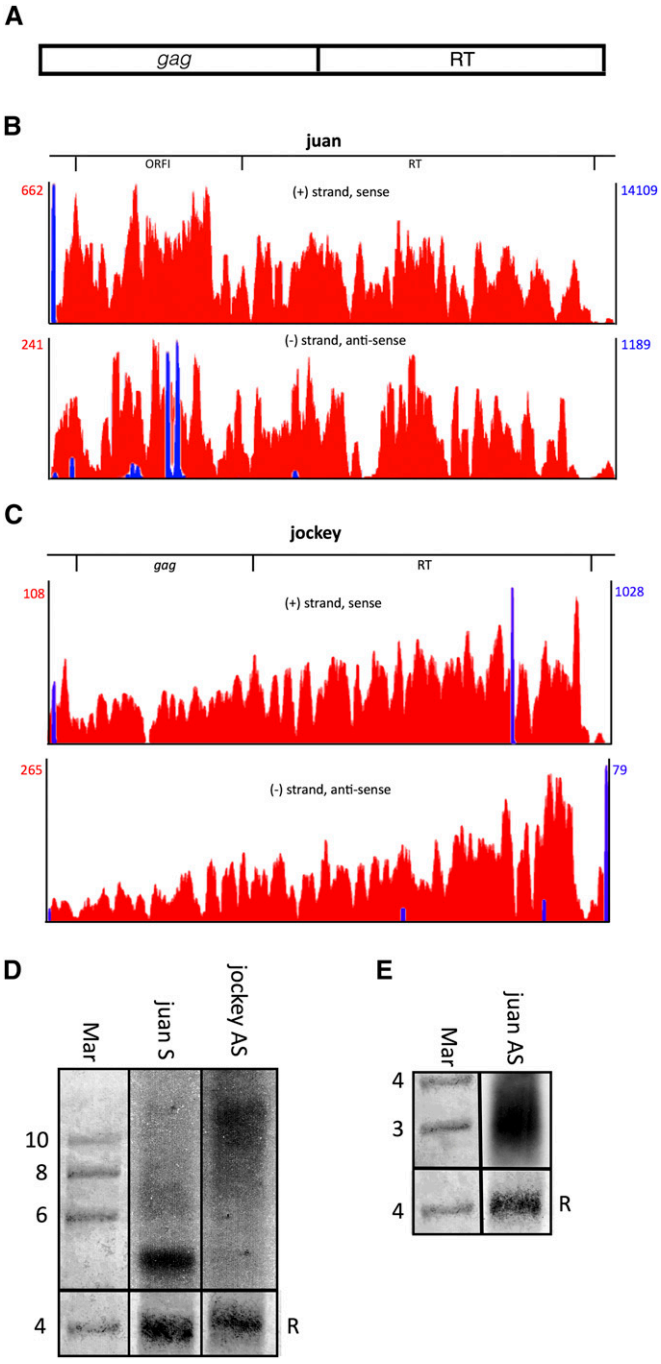


Figure 2 Non-LTR retroTns juan and jockey produce AS transcripts from intraelement tss. (A) Schematic of non-LTR retroTns in *Drosophila*. (B and C) Bedgraphs representing S (top) and AS (bottom) nonunique RNA-seq reads mapping to each non-LTR retroTn are shown in red. Other details are as described in Figure 1. (D and E) Representative Northern blots of juan S and jockey AS (D), and juan AS (E) transcripts are shown. Details are as in Figure 1E.

(Nechaev *et al.* 2010). FASTQ files were mapped strand specifically using RUM to obtain nonuniquely mapping reads. The UCSC genome browser (www.genome.ucsc.edu) was used to visualize the bedgraph output files. For presentation in Figure 1 and Figure 2, screen captures of nonunique S and AS reads and tss mapping to full-length, representative (for Dm297, blood, juan, and jockey) or individual (for mdg1{}1720) Tns were taken and overlaid to scale.

Table 1 *Drosophila* Tn transcripts

Class	Family	No. Tns	%S	%AS	Avg S RPM	Avg AS RPM	S/AS
Non-LTR	Jockey	5	80.0	80.0	302.0	104.8	2.88
	I	1	0.0	0.0	0.0	0.0	—
	R1	2	0.0	50.0	0.0	32.0	—
LTR	Gypsy	18	66.7	72.2	871.3	261.1	3.34
	Pao	3	100.0	100.0	1046.7	153.7	6.81
	Copia	1	100.0	100.0	64258.0	486.0	132.22
TIR	ProtoP	1	100.0	100.0	80.0	31.0	2.58
	Tc1	5	0.0	40.0	0.0	24.5	—
	Transib	1	100.0	100.0	48.0	21.0	2.29
	Pogo	1	100.0	100.0	785.0	28.0	28.0

Transposons sorted by class and family were analyzed. No. Tns, the number of individual Tns within a family having three or more full-length elements; %S or % AS, the percentage of Tns included in column 3 with S or AS nonuniquely mapping RNA-seq reads; AVG S or AVG AS, the average normalized nonunique S or AS read count (RPM) for each Tn family; and S/AS, the ratio of S RPM to AS RPM.

cDNA synthesis

A total of 50 ng of total Dmel-2 RNA isolated using the RNeasy MinElute Cleanup Kit (Qiagen) were reverse transcribed with RevertAid reverse transcriptase (Thermo Scientific) and a strand-specific, gene-specific reverse transcription primer (RT sense or antisense primer). This primer contains a unique nucleic acid tag 5′ of the complementary sequence that does not map to the *Drosophila* transcriptome (Table S5). The RT reaction contained 5× reaction buffer (no random hexamers or oligo dT), 1 μl Ribolock (40 units/μl), 1 mM dNTPs, 100 nM RT primer, and 2 μl of RevertAid (200 units/μl) in a total volume of 20 μl. The reaction incubated at 50° for 1 hr was heat inactivated at 85° for 5 min and was then diluted 1:10 with nuclease free water.

Quantitative PCR

Quantitative PCR (qPCR) was optimized and performed on a Bio-Rad CFX96 Real-Time system using SYBR Green detection chemistry (Bio-Rad SsoAdvanced Universal SYBR green). Briefly, 4 μl of diluted cDNA (10-fold) was mixed with 5 μl 2× SYBR green (Bio-Rad) and 0.5 μl of forward and 0.5 μl of reverse primer (500 nM final concentration). Initial denaturation was carried out at 95° for 3 min followed by a 30-sec denature step and a 30-sec annealing step (40×). Gene-specific primers for strand-specific qPCR are provided in Table S5. All RT-qPCR experiments were conducted in technical triplicates.

Northern blot analysis

A total of 50 μg of total RNA was separated on a 1% agarose denaturing gel at 100 V for ~4.5 hr. RNA was transferred to hybond+ nitrocellulose membrane, UV cross-linked, and rRNA was stained with methylene blue. Following prehybridization, blots were probed with ^{32}P-end labeled ~50-nt probes (Table S7). Blots were grouped based on predicted

transcript levels (group no. 1-S Dm297, S blood and S mdg1; group no. 2-AS Dm297, AS blood and AS mdg1; group no. 3-S and -AS juan, and S and AS jockey) and exposed to film. This treatment ensured that qualitative levels of transcripts within a group could be assessed.

RNAi, nuclear extract preparation, and antibodies

RNAi was performed as described (Sullivan *et al.* 2009). For the LacZ control, dsRNA targeting *E. coli* β-galactosidase was added to Dmel-2 cells to initiate the RNAi pathway without targeting *Drosophila* mRNAs. Comparing experimental knockdowns to this control ensures that molecular phenotypes observed upon RNAi depletion result from knockdown of specific mRNAs and not simply induction of RNAi. On day 5, the cells were counted with a Nexcelom cell counter and harvested. Nuclear extracts (NEs) were prepared as described (Sullivan *et al.* 2009) from 1×10^8 cells and then flash frozen in liquid N_2 and stored at $-80°$. Commercial α-Actin (Abcam ab8224) and α-Dcr-2 (Abcam ab4732) antibodies were used at 1:5000 and 1:1000, respectively, for Western blot.

Data availability

Gene expression data are available at GEO with the accession number: GSE67725

Results

To investigate Tn AS transcription, we performed strand-specific high throughput sequencing (HTS) of rRNA-depleted RNA (>200 nt) from control Dmel-2 cells (LacZ). These libraries were prepared in triplicate and sequenced on an Illumina HiSeq, resulting in an average read depth of 101.5× and >98% of reads mapping to the *Drosophila* genome (Table S1). Surprisingly, 41.9% of the reads mapped nonuniquely (Table S1), indicating that a large percentage of transcripts are derived from non-rRNA repetitive sequences.

To compare abundance of S and AS Tn transcripts, we visualized nonunique RNA-seq reads using the UCSC genome browser. Because Tns are highly conserved and multicopy, RNA-seq reads corresponding to S and AS Tn transcripts map to more than one genomic location. Therefore, the normalized reads per million value identified for each Tn generally represents total cellular S or AS transcription of all copies of that element. Only Tns having more than three full-length annotated elements in the *Drosophila* genome were investigated.

LTR retroTns generate the majority of AS Tn transcripts

This analysis revealed S and AS nonunique reads for the majority of Tns examined (Table 1), while little to no AS transcription of non-Tn genes is evident (Table S2). These observations are consistent with previous data that *Drosophila* does not exhibit AS transcription upstream of mRNA genes (Lapidot and Pilpel 2006; Nechaev *et al.* 2010; Core *et al.* 2012). Tn nonunique RPM show that the most abundant and active *Drosophila* Tn class, LTR retroTns, is highly expressed in the S and AS direction (Kaminker *et al.* 2002; Kofler *et al.* 2015) (Table 1). Non-LTR retroTn and TIR DNA Tns are generally transcribed at lower levels (Table 1). These data are consistent with previous analyses showing that LTR retroTn-derived esiRNAs are more prevalent in S2 cells than esiRNAs originating from non-LTR retroTn or TIR DNA Tns (Ghildiyal *et al.* 2008).

While AS transcription is observed for all but the "I" family of Tns, the ratio of S to AS nonunique reads differs dramatically when S and/or AS RPM are >100. Non-LTR retroTns in the Jockey family and LTR retroTns in the Gypsy family have generally low S/AS ratios (2.88 and 3.34, respectively) while LTR retroTns in the Pao and copia families and TIR DNA Tns in the Pogo family have much higher S/AS ratios (6.81, 132.22, and 28.04, respectively) (Table 1). LTR retroTns generating the most esiRNAs in S2 cells all belong to the Gypsy family of retroTns (Chung *et al.* 2008; Czech *et al.* 2008; Kawamura *et al.* 2008). Because Gypsy retroTns Dm297, blood, and mdg1 generate abundant esiRNAs and produce ample AS transcripts (Table S3), these retroTns were chosen for further investigation. Additionally, only Tns in the Jockey family of non-LTR retroTns generate both S and AS transcripts (Table 1) and esiRNAs (Kawamura *et al.* 2008). Jockey family members jockey and juan produce the highest levels of AS RNAs (Table S3). Therefore, to explore the importance of LTR sequences in retroTn AS transcription, juan and jockey were chosen for further study. TIR DNA Tns were not further investigated.

LTR retroTn AS transcription initiates from within or near LTRs

Bedgraphs of nonunique, strand-specific RNA-seq reads mapping to Dm297 (Figure 1B), blood (Figure 1C), and mdg1 (Figure 1D, right half of mdg1{}1720) representative full-length elements are shown. These data indicate that both S and AS transcripts are distributed across the elements including the three ORFs (Figure 1, B–D). Dm297, blood, and mdg1 AS reads tend to be concentrated in the LTRs, while S RPM are higher in the ORFs (Figure 1, B–D). In all three cases, total S transcript levels were higher than AS (Figure 1, B–D, red numbers). Some sequences were removed by splicing (data not shown).

To identify S and potential AS transcription start sites (tss), we remapped publicly available short-capped RNA high-throughput sequencing datasets (Nechaev *et al.* 2010; Henriques *et al.* 2013) and filtered to isolate only nonunique reads. Potential S and AS tss were observed for all three LTR retroTns (Figure 1, B–D, blue). RPM for tss on the S strand were higher than RPM for AS tss for Dm297 and blood (Figure 1, B and C, blue numbers). These data correlate with S and AS transcription levels for each retroTn. For Dm297 and blood, S and AS transcription could begin near the 3′ end of the LTR (Figure 1, B and C, top, blue) and transcription initiating from this location could result in S and AS RNAs spanning the entire element (Figure 1, B and C, blue).

In contrast, no LTR AS tss were observed for full-length mdg1 elements. An AS tss is visible (Figure 1D, right, bottom, blue); however, transcription from this tss would only produce AS RNAs of the 5′ LTR. One mdg1 element, mdg1{}1720, was identified that could produce the observed AS transcripts. Mdg1{}1720 consists of two tandemly repeated mdg1 elements with an inverted and centralized LTR in the RT ORF of the first mdg1 retroTn (Figure 1D). Transcription initiating from the non-LTR tss in the downstream mdg1 element could result in AS transcription of the first mdg1 retroTn. While only nonuniquely mapping AS reads are shown in Figure 1D, AS reads corresponding to unique sequences in the upstream mdg1 repeat are observed (data not shown), indicating that this specific retroTn, mdg1{}1720, is transcribed.

We next performed strand-specific RT-qPCR (Purcell *et al.* 2006; Vashist *et al.* 2012) to confirm S and AS transcription of Dm297, blood, and mdg1. Each potential transcript was reverse transcribed with a strand-specific, gene-specific primer having a unique nucleic acid tag (Table S5). The unique tag provided a primer binding site for the downstream qPCR reaction to ensure detection of only the transcript of interest. Random priming was evaluated in the absence of an RT primer and no target transcripts were detected (data not shown). For Dm297, blood, and mdg1, we detected both S and AS transcription using several primer sets spanning the coding sequence (Table S5). We calculated the difference between S and AS Cts (ΔCt(S − AS)) for Dm297, blood, and mdg1 (Table S6). AS transcripts were less abundant in all cases and the differences between S and AS transcription correlated with RNA-seq data (Figure 1, B–D).

Lastly, we performed Northern blot analysis to detect S and AS transcripts (Table S7). RNA from Dmel-2 cells was transferred to a membrane and probed with radioactively labeled Dm297, blood, and mdg1 complementary S and AS probes. S (Figure 1E) and AS (Figure 1F) transcripts were observed for Dm297, blood, and mdg1. S Dm297 (6995 nt) appears slightly smaller than its predicted size while S blood (7410 nt) and mdg1 (7480 nt) are slightly larger (Figure 1E). No other prominent bands were detected. The Dm297 S transcript is most abundant while mdg1 and blood S transcripts are much less prevalent. Each lane contained equal amounts of 28S rRNA (Figure 1E, bottom). These data correlate with LTR retroTn S transcript levels observed in the RNA-seq data (Figure 1, B–D). We observe multiple Dm297 and blood AS transcripts resulting in a smear between 8 and 10 kb, while a single AS mdg1 transcript is present at ∼8 kb (Figure 1F). An ∼2 kb mdg1 AS transcript is also visible (Figure 1F). The sizes of these RNAs support AS transcription of full-length Dm297 and blood LTR retroTns from the bioinformatically identified tss (Figure 1, B–D). These data also indicate that the mdg1 AS tss (Figure 1D, first and third blue peaks) is functional, producing the predicted ∼8 kb and ∼2 kb RNAs, while the inverted LTR S tss is not active in this context. Multiple Dm297 AS transcripts may result from inefficient RNAPII termination. A lighter exposure (Figure 1F, bottom "L") of the AS transcripts reveals that the ∼8 kb AS mdg1 transcript is most abundant, while Dm297 and blood AS RNAs are less prevalent, mirroring the RNA-seq data (Figure 1, B–D). S2 culture cells have amplified Tn content (Potter *et al.* 1979; Junakovic *et al.* 1988; Wen *et al.* 2014) and RNA-seq reads cannot be mapped to the S2-specific mdg1 copies as the S2 genome has not been sequenced. Regardless, the data presented here support AS transcription of mdg1{}1720 from non-LTR tss.

To gain a more complete view of AS retroTn transcription, we considered the genomic context of each annotated full-length Dm297, blood, and mdg1 element (Table S4). Individual retroTns were found both intergenically and within introns. Intronic Dm297 and mdg1 retroTns were more often AS to their host coding gene, while blood elements were in the S orientation. These data indicate that transcription of intronic LTR retroTns, together with RNAs produced from intraelement tss, could contribute to AS transcript abundance.

We also investigated S and AS transcription of individual Dm297, blood, and mdg1 elements (Table S4). If unique intra-element RNA-seq reads and RNA-seq reads corresponding to the retroTn-intergenic/intronic sequence junction could be identified, we concluded that the individual Dm297, blood, or mdg1 element was transcribed. S transcription was confirmed for 9/18 (50%) of Dm297, 2/15 (13%) of mdg1, and 9/22 (41%) of blood retroTns (Table S4). AS transcription was verified for 4/18 (22%) of Dm297, 1/15 (7%) of mdg1, and 3/22 (14%) of blood elements (Table S4). Of all mdg1 retroTns, only AS transcription of mdg1{}1720 could be confirmed using this analysis. These numbers of transcribed individual elements are probably an underestimate of the total as not all elements have mutations allowing observation of unique internal reads. Also, RPM for unique reads were low, reflecting less RNA from one individual element as compared to nonunique RPM corresponding to total transcription of all individuals of a retroTn type. Collectively, the RNA-seq analyses, strand-specific RT-qPCR, and Northern blots indicate that individual LTR retroTns in Dmel-2 cells undergo convergent S and AS transcription and that transcription can initiate within or near Dm297, blood, and mdg1 LTRs.

Non-LTR retroTns juan and jockey produce AS transcripts

To investigate the role of LTRs in AS transcription, we examined non-LTR retroTns juan and jockey. Similar to LTR elements, strand-specific nonunique RNA-seq reads map the entire length of jockey and juan (Figure 2, B and C, red); however, non-LTR S and AS transcripts are less abundant than corresponding LTR retroTn RNAs. Additionally, jockey is the only retroTn investigated for which the AS transcript is more highly expressed than the S transcript.

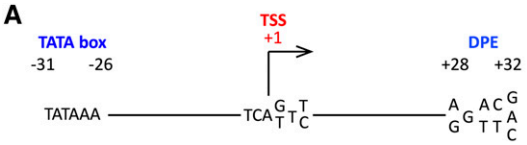
A juan S tss is observed at the 5′ end of the retroTn (Figure 2B, top, blue) and initiation of transcription from this location could result in a single complete element S transcript. Several AS tss were also observed; however, these tss cannot be responsible for reads mapping to the 3′ half of juan (Figure

2B, bottom, blue). The source of these AS RNA-seq reads is unclear, although unique AS reads were identified for both intergenic and intragenic juan elements (Table S4), indicating AS transcription of individual elements (see previous section). Two S tss are observed for jockey. Transcription beginning at these tss could produce a S transcript the entire length of the element (Figure 2C, top, blue). For jockey, a potential AS tss is observed at the 5′ end, but the number of reads mapping to this tss do not correlate with higher AS RPM (Figure 2C, bottom, blue). Collectively, these data indicate that non-LTR retroTns are transcribed in both S and AS directions, albeit at lower levels than LTR retroTns.

To verify S and AS transcription, we performed strand-specific qPCR of non-LTR retroTns juan and jockey as described in the previous section (Table S5). S and AS transcription were detected for juan and jockey and the ΔCt(S − AS) mirrored the ratio of S to AS RPM of each ORF investigated (Table S6). Additionally, Northern blot analysis revealed juan S and AS, and jockey AS transcription (Figure 2, D and E); however, the S jockey transcript was not visible presumably because of its low abundance. Previously, S jockey transcripts initiating from an internal promoter were identified in *Drosophila* cell culture (Mizrokhi *et al.* 1988). A probe to the S juan transcript (4236 nt) reveals one band ~5 kb, while AS jockey (5020 nt) and AS juan probes show smears indicating multiple transcripts (Figure 2, D and E). Juan AS RNAs range from ~2 kb to ~4 kb (Figure 2E). If juan AS transcription initiates from bioinformatically identified tss (Figure 2B) and RNAPII termination is inefficient, multiple ~2 to ~4 kb AS transcripts could be produced. Jockey AS RNAs range from ~7 kb to greater than 10 kb (Figure 2D). We hypothesize that smaller transcripts in this range could be produced from the bioinformatically identified tss (Figure 2C). Additionally, one jockey retroTn (jockey{}1630), has a LTR retroTn, roo (9092 nt), inserted in the *gag* ORF making this element 14,127 nt. Transcripts originating from an observed tss (data not shown) at the 5′ end of jockey{}1630 could result in jockey AS RNAs greater than 10 kb. These data suggest that non-LTR retroTns juan and jockey are transcribed in both the S and AS directions from intraelement tss.

S and AS tss have canonical Drosophila promoter elements

We next wanted to determine if the observed LTR and non-LTR retroTn tss were flanked by traditional *Drosophila* promoter elements. *Drosophila* transcription initiates at T-C-A_{+1}-G/T-T-T/C (where A_{+1} is the tss) within a promoter composed of a TATA box (−31 to −26) and/or a downstream promoter element (DPE) located between +28 to +32 (Butler and Kadonaga 2002) (Figure 3A). The TATA box or DPE occur in core promoters 29% and 26% of the time, respectively, while 14% contain both a TATA box and a DPE (Butler and Kadonaga 2002). Further evaluation of Dm297 revealed near-canonical tss and DPEs in appropriate locations for both S and AS promoters (+28 and +29, respectively, Figure 3B). Blood S and AS transcripts initiate from tss having two non-



B

Tn	TSS	TSS 5′-3′	S DPE	AS DPE	S DPE Pos.	AS DPE Pos.
Dm297	S-TSS	TTAGTC	AGTCG	-	28	-
Dm297	AS-TSS	GCAGTC	-	GGATG	-	29
blood	S-TSS	TTAGAC	AGAAC	-	28	-
blood	AS-TSS	TGACTT	-	GGTCA	-	28
mdg1	S/AS-TSS	TCAGTT	AGATC	-	28	-
mdg1	AS-TSS_2	TCACAT		GGTCA		38
juan	S-TSS	ACAGTC	GGATG	-	28	-
juan	AS-TSS	TCAGTT		AGTTA	-	28
jockey	S-TSS		-	-	-	-
jockey	AS-TSS		-	-	-	-

Figure 3 S and AS tss have canonical *Drosophila* RNAPII promoter elements. (A) A schematic representing canonical *Drosophila* promoter elements is shown. (B) HTS analysis at nucleotide resolution of LTR and non-LTR retroTns tss is depicted. The tss, tss sequence, S DPE, AS DPE, and position of each DPE are shown for each retroTn. Bold nucleotides represent divergence from canonical nucleotide/s shown in A.

canonical bases but also have canonical DPEs +28 from S and AS tss (Figure 3B). The mdg1 LTR has a canonical tss with a perfect DPE +28 downstream, while the AS tss not located in the LTR has two nonideal bases and an inappropriately spaced DPE (Figure 3B). These data support previous characterization of the mdg1 S promoter (Arkhipova and Ilyin 1991). The non-LTR retroTn juan has a near canonical S tss and a canonical AS tss. Both tss have canonical DPEs +28 bases downstream of the tss (Figure 3B). These data support bonafide S tss for all three LTR-retroTns and non-LTR retroTn juan. Finally, promoter analysis revealed no canonical initiation site or DPE for either S or AS jockey tss. We hypothesize that this promoter is unique compared to more canonical core *Drosophila* promoters.

LTR and non-LTR retroTn AS transcripts lack strong polyadenylation

Our data suggest that S and AS LTR and non-LTR retroTns are convergently transcribed from canonical *Drosophila* promoters. As these RNAs are likely RNAPII transcripts (Gogvadze and Buzdin 2009), we wanted to determine their polyadenylation status. S retroTn transcripts have canonical polyadenylation signals (Gogvadze and Buzdin 2009; this work) and polyadenylation of these RNAs has previously been reported (Gogvadze and Buzdin 2009). We first fractionated total RNA using an oligo d(T) column and then performed strand-specific qPCR to the RT ORF of each retroTn on total RNA, polyA+ RNA and polyA− RNA (Table S5). We used the amount of transcript in total RNA to normalize polyA+ and polyA− fractions by subtracting polyA+ and polyA− Cq values from those of total RNA (ΔCq = total-polyA +/−) (Figure 4A). We then determined a fold enrichment of polyadenylation by calculating the difference between these ΔCq values (ΔΔCq = [(total-polyA +)-(total-polyA −)]) (Figure 4B). 18s rRNA (polyA−) and Actin (polyA+) were used as controls. Total

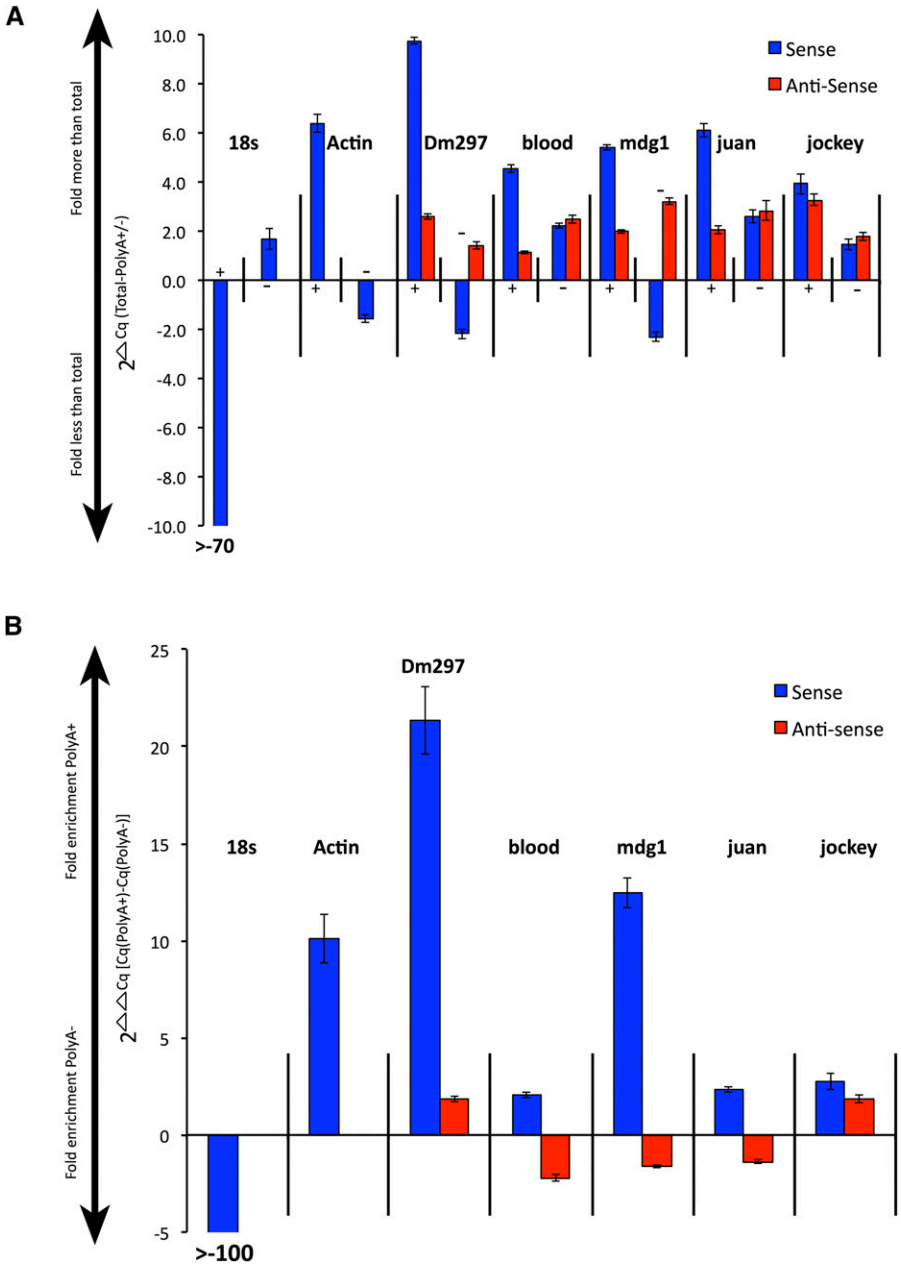


Figure 4 LTR and non-LTR retroTn AS transcripts lack strong polyadenylation. (A) Graph of S and AS transcript fold differences in polyA+ or polyA− fractions compared to total RNA. $2^{\Delta Cq}$ is the *y*-axis and represents (total RNA – polyA+ or polyA−). PolyA+ or polyA− fractions are indicated along the *x*-axis as + or – signs. S transcripts are blue bars and AS transcripts are red bars. The name of each retroTn or control is listed above the appropriate group. Error bars represent standard deviation of strand-specific qPCR technical triplicates. (B) Graph of direct comparison of polyA+ to polyA− levels for each retroTn S/AS transcript pair. Fold enrichment values for polyA+ or polyA− fractions are shown along the *y*-axis ($2^{\Delta\Delta Cq}$) where the ΔΔCq equals [(total-polyA +)-(total-polyA −)]). S and AS transcripts are shown along the *x*-axis; S bars are blue and AS bars are red. Error bars represent standard deviation of strand-specific qPCR technical triplicates.

RNA was efficiently separated into polyA+ and polyA− fractions as 18s S transcripts were >70-fold less in the polyA+ fraction than in total RNA (Figure 4A, 18s +) and actin S transcripts were ∼6-fold increased in the polyA+ fraction as compared to total RNA (Figure 4A, actin +). 18s S RNAs were more than 100-fold depleted in polyA+ transcripts, while Actin S RNAs were approximately 10-fold enriched in polyA+ transcripts (Figure 4B), indicating that our assay to assess polyadenylation was working properly.

S Dm297 and mdg1 transcripts were ∼10- and ∼5-fold more, respectively, in the polyA+ fraction than in total RNA and ∼2-fold less in the polyA− fraction than in total RNA (Figure 4A, Dm297 and mdg1, blue). These S transcripts are enriched for polyadenylation at least as much (mdg1) if not more (Dm297) than the polyadenylated Actin control (Figure 4B). Dm297 and mdg1 AS transcripts, and blood, juan, and jockey S and AS transcripts were between ∼1.5- and ∼3-fold more in both polyA+ and polyA− fractions than in total RNA, indicating a mixture of both polyA+ and polyA− transcripts (Figure 4A). ΔΔCq calculations suggest that blood, juan, and jockey S transcripts are enriched for polyadenylation although much less than Dm297 and mdg1 (Figure 4B). None of the AS transcripts are highly enriched with polyA+ transcripts. Blood, mdg1, and juan are slightly enriched in polyA− RNAs (Figure 4B). These data suggest that while all retroTn S transcripts have polyadenylation signals, only Dm297 and mdg1 S transcripts are polyadenylated. Bioinformatic assessment did not reveal strong polyadenylation sites for any of the AS transcripts examined (data not shown).

Collectively, these data support a hypothesis that retroTn AS transcripts are not strongly polyadenylated.

Dcr-2 depletion decreases retroTn-derived esiRNA levels

Previous studies show that esiRNAs, many of which map to retroTns, are cleaved from long dsRNA precursors by Dcr-2 (Tomari *et al.* 2007; Ghildiyal *et al.* 2008; Chung *et al.* 2008; Kawamura *et al.* 2008; Siomi *et al.* 2008). Knockdown of Dcr-2 results in decreased esiRNA levels and a corresponding increase in precursor RNAs (Chung *et al.* 2008; Ghildiyal *et al.* 2008). Small RNA (<200 nt) high-throughput sequencing (HTS) libraries were constructed in triplicate from Dcr-2-depleted and control (LacZ) cells (Figure 5A). Greater than 99% of reads from all six libraries mapped to the *Drosophila* genome (Table S1). Dcr-2 knockdown resulted in a statistically significant ($P = 0.00154$) decrease in nonunique reads (22.7%) compared to the LacZ control (34.4%) (Table S1), indicating that Dcr-2 is required for global production of nonuniquely mapping esiRNAs.

Nonunique siRNA-seq reads map across Dm297, blood, mdg1, juan, and jockey for both the LacZ control and the Dcr-2-depleted samples (Figure 5, B–F) and esiRNA patterns are generally similar for both the control and the Dcr-2 knockdown. RPM vary considerably among the five retroTns with the most esiRNAs mapping to Dm297 and blood, and fewer mapping to mdg1, juan, and jockey (Figure 5, B–F, red numbers). Both Dm297 and mdg1 have a higher concentration of reads mapping to LTRs, as previously described (Chung *et al.* 2008; Ghildiyal *et al.* 2008) (Figure 5, B and D). RPM for esiRNAs mapping to all five retro Tns were decreased in Dcr-2-depleted Dmel-2 cells (Figure 5, B–F). Average RPM calculated from triplicate sequencing experiments for each retroTn in both control and Dcr-2-depleted samples (Figure 5G) were used to determine the ratio of esiRNAs in the LacZ control as compared to the Dcr-2 knockdown (Figure 5H). Dcr-2 depletion led to statistically significant reduction of the number of esiRNAs mapping to Dm297 (2.6-fold), jockey (2.9-fold), mdg1 (2.0-fold), juan (1.7-fold), and blood (1.4-fold) (Figure 5H). These data indicate that depletion of Dcr-2 causes a decrease in retroTn-derived esiRNA levels without changing the specific esiRNAs produced. These data strengthen the previous hypothesis that Dcr-2 produces retroTn-derived esiRNAs in Dmel-2 cells.

Sense and antisense retroTn transcript levels increase with Dcr-2 knockdown

Previous RT-qPCR studies suggest that some retroTn transcript levels increase after knockdown of Dcr-2 in S2 cells (Chung *et al.* 2008; Ghildiyal *et al.* 2008). To determine S and AS retroTn RNA levels globally, we performed strand-specific RNA-seq on Dcr-2-depleted, large RNA (>200 nt) resulting in an average read depth of 78.3× and ≥98% of reads mapping to the *Drosophila* genome (Table S1). Dcr-2 depletion resulted in a lower read depth as compared to the control (Table S1). The percentage of nonuniquely mapping reads was significantly increased in the Dcr-2 knockdown (46.7%) compared to the LacZ knockdown (41.9%) ($P = 9.6 \times 10^{-5}$, Table S1).

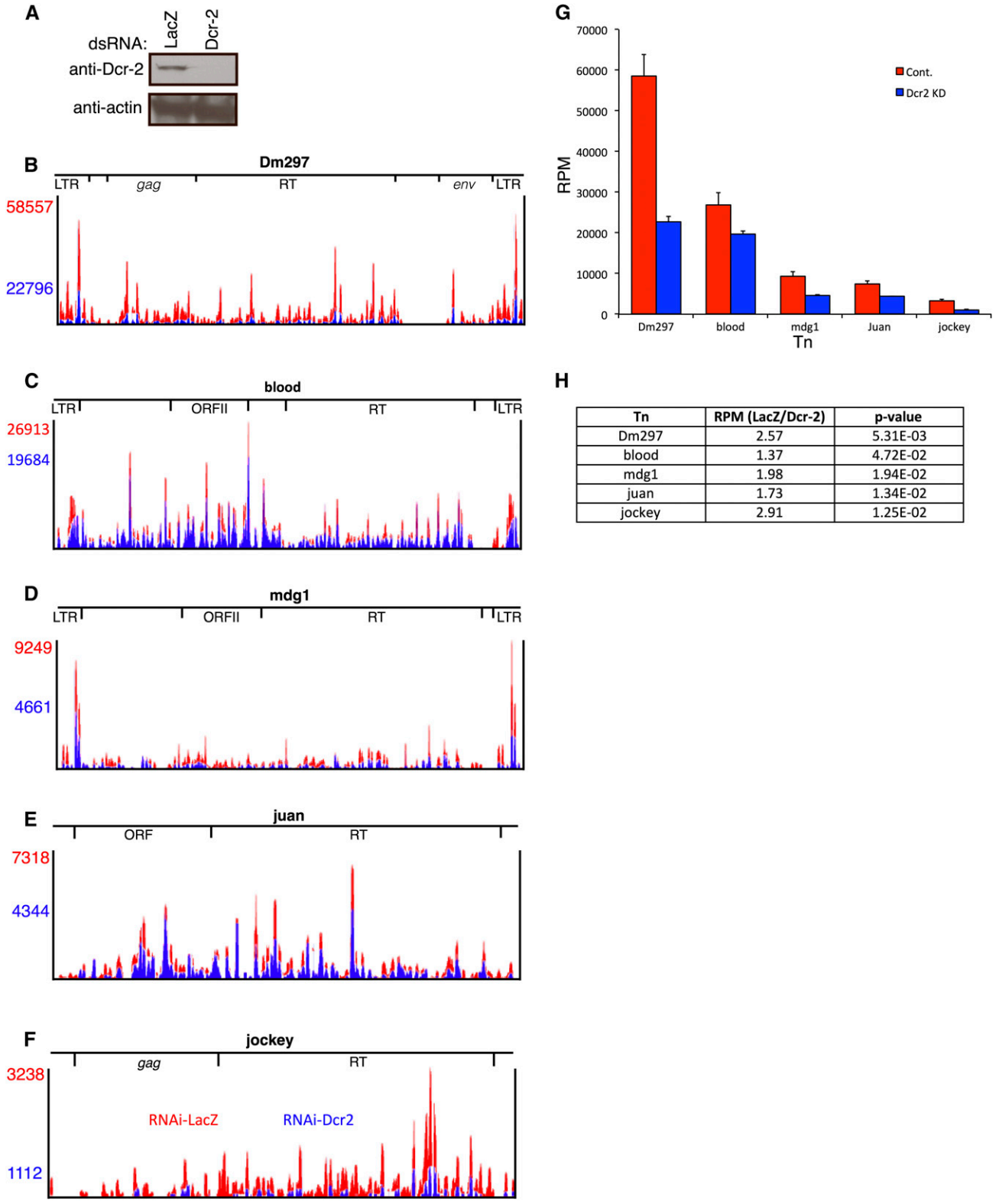
We compared Dcr-2 knockdown and LacZ control RPM for Dm297, blood, mdg1, juan, and jockey (Figure 6A). Generally, S retroTn transcript levels were increased in the Dcr-2-depleted samples, as previously reported (Figure 6, A, red, and B) (Chung *et al.* 2008; Ghildiyal *et al.* 2008). We observed a similar trend for AS retroTn transcripts except for Dm297, which showed a slight reduction in AS transcript levels following Dcr-2 knockdown (Figure 6, A, blue, and B). These results suggest that Dcr-2 generates esiRNAs from dsRNA precursors consisting of S and AS retroTn transcripts.

Discussion

Understanding the mechanisms that balance retroTn amplification and repression in eukaryotes is critical, as misregulation can lead to detrimental genomic damage. Many retroTns are active in *Drosophila* (Kofler *et al.* 2015), providing a unique opportunity to understand molecular mechanisms of retroTn repression. In *Drosophila* somatic cells, silencing of retroTns requires a dsRNA precursor that is processed into esiRNAs by Dcr-2 (Tomari and Zamore 2005; Ghildiyal *et al.* 2008; Marques *et al.* 2010). To better understand the origin of this retroTn-derived dsRNA precursor, we performed RNA-seq, strand-specific qPCR, and Northern blotting of control Dmel-2 cells. Most Tns produce S and AS transcripts, although S and AS expression are highest for LTR retroTns (Table 1). Bioinformatic analysis of representative LTR retroTns Dm297 and blood, a specific mdg1 element (mdg1{}1720), and representative non-LTR retroTns juan and jockey showed S and AS transcripts originating from intraelement transcription start sites for all elements investigated (Figure 1 and Figure 2). These initiation sites are generally canonical RNAPII transcription start sites with conserved DPEs (Figure 3). Collectively, these data suggest that AS retroTn RNAs are convergently transcribed from these start sites. Interestingly, we also observed that AS transcripts derived from retroTns are not strongly polyadenylated (Figure 4). By sequencing small RNAs, we determined that esiRNAs are globally derived from locations of retroTn S/AS convergent transcription and that Dcr-2 knockdown decreases esiRNA levels (Figure 5). Consistently, we showed that both S and AS retroTn transcript levels increase when Dcr-2 is knocked down (Figure 6). Taken together, these data support a model in which AS retroTn transcripts hybridize to their S counterparts forming dsRNAs that are substrates for esiRNAs production by Dcr-2 (Figure 7).

Drosophila retroTns are convergently transcribed from independent, canonical S and AS tss

Unlike in mammals, *Drosophila* RNAPII transcription does not initiate bidirectionally from promoters to generate AS transcripts (Lapidot and Pilpel 2006; Nechaev *et al.* 2010; Core *et al.* 2012). Also, the >100 predicted overlapping *cis*-natural pairs in *Drosophila* are most often complementary ORF 3′

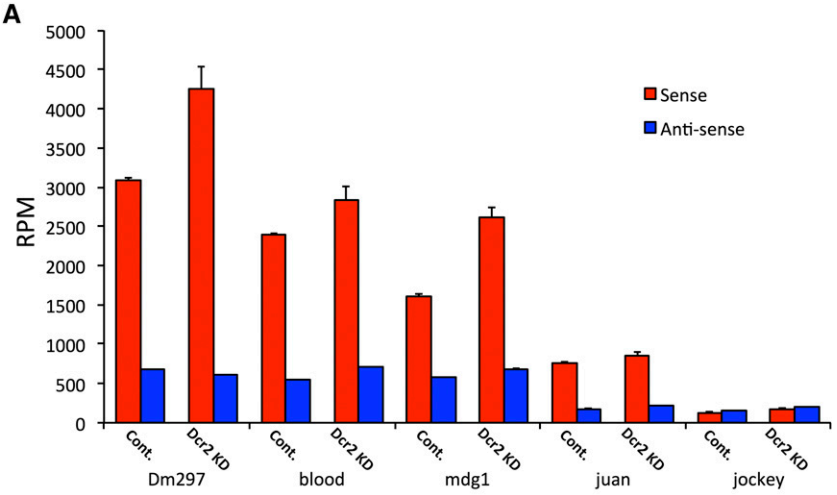


H

Tn	RPM (LacZ/Dcr-2)	p-value
Dm297	2.57	5.31E-03
blood	1.37	4.72E-02
mdg1	1.98	1.94E-02
juan	1.73	1.34E-02
jockey	2.91	1.25E-02

Figure 5 Dcr-2 depletion decreases retroTn-derived esiRNA levels. (A) Representative Western blot of Dcr-2 depletion. The antibody used is indicated to the left of blots and dsRNA for RNAi is labeled above blots. (B–F) Bedgraphs of esiRNAs mapping to retroTns in control and Dcr-2-depleted Dmel-2 cells. The control is red and the Dcr-2 knockdown is in blue. RPM values are listed to the left of bedgraphs and are color coordinated. Relative locations of specific ORFs are shown above the bedgraphs. (G) Graphs of esiRNA levels in control and Dcr-2-depleted Dmel-2 cells (control, red; Dcr-2 knockdown, blue). RPM values are on the *y*-axis and retroTns are indicated along the *x*-axis. Error bars represent standard deviation of technical triplicates. (H) A table reporting the ratio of esiRNA levels (RPM) between the control and Dcr-2 knockdown for each retroTn (middle column) is shown. RetroTn is indicated in the left column and *P*-value (unpaired *t*-test) is indicated in the right column.

A



B

Tn	Sense RPM Dcr2/RPM LacZ	Sense p-value	Anti-sense RPM Dcr2/RPM LacZ	Anti-sense p-value
Dm297	1.38	1.76E-02	0.92	3.02E-04
blood	1.18	4.29E-02	1.29	2.39E-04
mdg1	1.63	3.34E-03	1.18	7.11E-04
Juan	1.11	7.04E-02	1.23	2.16E-01
jockey	1.36	8.99E-04	1.35	1.52E-04

Figure 6 S and AS retroTn transcript levels increase with Dcr-2 knockdown. (A) A graph of retroTn transcript levels upon RNAi depletion of Dcr-2 is shown (sense, red; antisense, blue). RPM values are shown on the *y*-axis and retroTn (control *vs.* Dcr-2 knockdown) is indicated on the *x*-axis. Error bars represent standard deviation of technical triplicates. (B) A table reporting the ratio of retroTn S and AS transcript levels (RPM) between the control and Dcr-2 knockdown for each retroTn is shown. RetroTn is indicated in the first column. *P*-values (unpaired *t*-test) are reported for these comparisons.

UTRs (Okamura *et al.* 2008a), not S transcript-noncoding AS RNA pairs as in other organisms (Pelechano and Steinmetz 2013). Because protein coding genes do not produce AS transcripts in *Drosophila*, mechanisms of AS transcription and downstream regulatory functions have not been fully elucidated. AS retroTn transcription of *Drosophila* telomere LTR retroTns has been observed previously (Danilevskaya *et al.* 1999). Herein, we provide the first evidence of global AS transcription of LTR and non-LTR retroTns, and TIR DNA Tn families in *Drosophila* (Table 1 and Table S3). Additionally, we identify and quantitate AS transcription of individual Tns and specific elements (Figure 1, Figure 2, and Table S4).

Bioinformatically identified AS transcription start sites and promoter analysis provide the first clues about how retroTn AS transcripts are produced (Figure 3). Interestingly, Dm297, blood, and mdg1 AS transcription start sites and promoter elements are located within the retroTn and AS transcripts initiating from these locations could explain all observed AS RNA-seq reads, suggesting that external sequences are not required for LTR retroTn AS transcription. Evidence to support this hypothesis comes from identifying several individual, intergenic LTR retroTns that are transcribed in the AS direction (Table S4). It seems unlikely that multiple intergenic LTR retroTns simultaneously evolved external promoters in different genomic locations and is more plausible that the observed internal Dm297, blood, and mdg1 AS transcription start sites are functional. AS transcription start sites are observed for non-LTR retroTn juan, but these cannot be responsible for transcription of the 5′ end of the element (Figure 2). The AS transcription start site identified for jockey does not have adequate normalized read counts to account for the amount of AS transcription (Figure 2). We hypothesize that the additional AS jockey and 5′ juan transcripts originate from intragenic elements oppositely oriented to mRNA S transcripts (Table S4). Collectively, these data suggest that LTR retroTns are transcribed from intraelement transcription start sites, while some non-LTR retroTns RNAs are generated indirectly by transcription of protein coding genes.

AS transcripts can arise from bidirectional transcription at RNAPII promoters (Core *et al.* 2008; Guil and Esteller 2012) or convergent transcription from strand-specific promoters (Gullerova and Proudfoot 2012). Bidirectional transcription initiates in both directions from one promoter, while convergent transcription requires independent transcription start sites. Bidirectional transcription of a retroTn from a single promoter would not result in a full-length dsRNA (Figure 7). Therefore, our results suggest that double-stranded retroTn RNAs are derived from convergent transcription of S and AS RNAs from independent transcription start sites (Figure 7). Transcriptional gene silencing mediated by convergent transcription is highly efficient in both fission yeast and mammalian cells (Gullerova and Proudfoot 2012). Our data support a model in which formation of dsRNAs by convergent transcription of retroTns is the first step in *Drosophila* somatic cell retroTn silencing.

Production of dsRNAs by convergent transcription is a novel retroTn regulatory mechanism

A well-studied mammalian non-LTR retroTn, L1, initiates AS transcription in the S RNA 5′ UTR in humans (Speek 2001; Nigumann *et al.* 2002) and the S RNA ORF1 in mice (Li *et al.* 2014). Most full-length intragenic L1 elements are oriented AS to protein coding genes (Szak *et al.* 2002). Therefore, AS

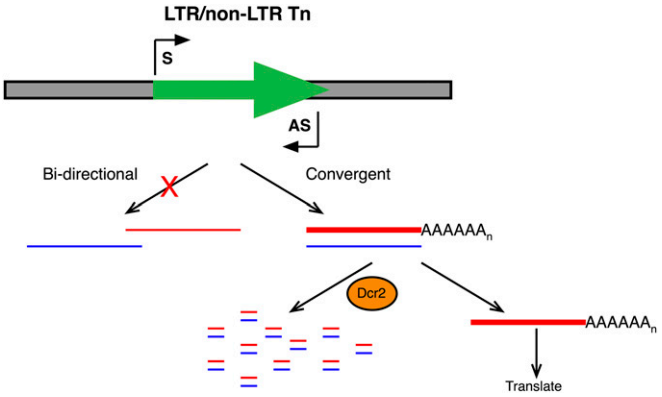


Figure 7 Dcr-2 generates esiRNAs from dsRNAs derived from convergent S and AS transcription of retroTns. Shown is a model depicting convergent S and AS transcription (arrows, black, "S" and "AS," respectively) of retroTns (arrow, green) in *Drosophila*. S transcripts (red) are polyadenylated and more abundant (thick line) compared to AS transcripts (blue, thin line). AS transcripts act as a molecular sponge isolating a portion of S transcripts resulting in the formation of dsRNA Dcr-2 substrates. Some S transcripts are translated promoting mobility of retroTns.

transcription from the identified transcription start sites proceeds into neighboring mRNAs forming fusion transcripts that regulate expression of numerous genes (Speek 2001; Nigumann *et al.* 2002; Mätlik *et al.* 2006; Cruickshanks and Tufarelli 2009) and affect mobility of L1 elements (Li *et al.* 2014). The L1 retroTn is closely related to *Drosophila* non-LTR retroTns jockey and juan (Mizrokhi *et al.* 1988; Speek 2001). No AS transcription start sites were observed in jockey analogous to L1 AS transcription start sites (Figure 2C). Juan AS transcription start sites were located in the S RNA ORFI (Figure 2C), but only two full-length juan elements (of seven total) are intragenic (Table S4) limiting the impact of a potential L1-like AS fusion transcript regulatory mechanism. Additionally, juan AS transcripts were identified upstream of the observed transcription start sites. Together, these data indicate that the mechanism of *Drosophila* non-LTR retroTn AS transcript initiation and the functions of these AS RNAs may differ from their mammalian L1 counterparts.

In fission yeast and the *Drosophila* germline, movement of repetitive sequences and retroTns, respectively, are repressed by a transcriptional gene silencing mechanism wherein siRNAs induce heterochromatin formation (Huisinga and Elgin 2009; Wang and Elgin 2011; Sienski *et al.* 2012; Huang *et al.* 2013; Le Thomas *et al.* 2013; Rozhkov *et al.* 2013). In *Schizosaccharomyces pombe*, siRNAs are produced from RNA-dependent RNA polymerase generated dsRNA precursors by Dicer 1 (Volpe *et al.* 2002; Yu *et al.* 2014; Holoch and Moazed 2015). In *Drosophila*, retroTns are silenced in the germline by piRNAs cleaved from single-stranded substrates and amplified via a mechanism that does not include long dsRNAs (Saito *et al.* 2006; Aravin *et al.* 2007; Gunawardane *et al.* 2007). Therefore, our proposed model that esiRNAs are generated from hybridized convergently transcribed S and AS retroTn transcripts, is novel as other mechanisms do not require dsRNA substrates, utilize an RNA-dependent RNA polymerase to produce dsRNA substrates, or use AS transcripts to regulate gene expression in a way that does not require siRNAs.

Lack of AS retroTn polyadenylation may lead to nuclear retention of dsRNAs

Efficient polyadenylation of transcripts can promote export to the cytoplasm and removal of polyA signals may cause nuclear retention of RNAs (Dower *et al.* 2004). We propose that convergently transcribed S and AS retroTn transcripts hybridize in the nucleus forming dsRNAs. Because only Dm297 and mdg1 S transcripts are polyadenylated (Figure 4), all double-stranded retroTn RNAs investigated would contain at least one polyA− component, encouraging nuclear retention of these dsRNAs. As the number of retroTn AS transcripts is often significantly less than the number of S transcripts (Figure 1 and Figure 2), unhybridized S RNAs are exported to the cytoplasm for translation, leading to a balance of repression and expansion of retroTns. A nuclear pool of Dcr-2 (Cernilogar *et al.* 2011, and data not shown) may use nuclear-retained retroTn dsRNA as substrates for esiRNAs biogenesis.

Dcr-2 generates esiRNAs from dsRNAs derived from convergent S and AS transcription of retroTns

LTR and non-LTR retroTns produce esiRNAs from dsRNA precursors through Dcr-2-dependent mechanisms in *Drosophila* somatic cells (Ghildiyal *et al.* 2008; Kawamura *et al.* 2008; Siomi *et al.* 2008). Here, we show that expression of both S and AS retroTn transcripts is regulated by Dcr-2. Specifically, depletion of Dcr-2 leads to reduction in esiRNA levels (Figure 5) and a corresponding increase in both S and AS retroTn transcript levels (Figure 6). The mechanism of Dcr-2 mediated AS silencing is likely similar to S silencing as S and AS esiRNAs are often equally abundant (Ghildiyal *et al.* 2008).

Others have hypothesized that esiRNAs are processed from double-stranded LTR hairpins because of higher concentrations of small RNA-seq reads from LTRs (Chung *et al.* 2008; Ghildiyal *et al.* 2008). Generally, our data do not support this model. EsiRNA reads from retroTns span the entire element, suggesting that LTR hairpins cannot be the only dsRNA substrates (Figure 5). Additionally, S and AS RNA-seq reads map to all regions of retroTns, indicating that the entire element has the potential to form a dsRNA precursor. Also, we observe convergent transcription of non-LTR retroTns and Dcr-2-regulated esiRNAs mapping to these retroTns (Figure 5). Thus, an LTR is not required for dsRNA formation and subsequent siRNA biogenesis.

Mechanisms of AS transcription and esiRNA biogenesis are conserved in tissue culture and Drosophila

The data presented here were collected in Dmel-2 cells, a derivative of Schneider 2 (S2) cells, a somatic cell line derived from *Drosophila* embryos (Schneider 1972). Previous parallel investigation of esiRNA biogenesis in S2 cells and *Drosophila* heads indicated no mechanistic differences between these two tissues (Ghildiyal *et al.* 2008). Most importantly, esiRNAs were equally derived from S and AS retroTn strands and

mapped evenly across retroTn precursors (Ghildiyal *et al.* 2008), indicating that a full-element dsRNA precursor is required to generate the observed esiRNAs in both fly tissues and cell lines. These data are consistent with our results that retroTns in S2 cells produce AS transcripts (Figure 1 and Figure 2).

Previous studies show that *Drosophila* tissue culture cells have amplified Tn content (Potter *et al.* 1979; Tchurikov *et al.* 1981; Maisonhaute *et al.* 2007; Wen *et al.* 2014) and hypothesize that this amplification is necessary for creating immortal cell lines (Junakovic *et al.* 1988). Once established, Tn location and number appear stable in *Drosophila* Kc and S2 cell lines (Junakovic *et al.* 1988). This amplification is reflected as a greater portion of retroTn-derived esiRNAs mapping to Tns in S2 cells than in *Drosophila* heads (Ghildiyal *et al.* 2008). While having more Tn copies in tissue culture potentially increases the absolute levels of S and AS retroTn transcripts (and esiRNAs generated from dsRNA precursors) the molecular mechanisms required to produce AS transcripts and generate esiRNAs appear conserved between flies and culture cells (Ghildiyal *et al.* 2008). Additionally, a higher concentration of esiRNAs and dsRNA precursors is a tremendous advantage of the S2 cell system.

In conclusion we show, for the first time, that *Drosophila* retroTns are transcribed in the AS direction from intraelement transcription start sites. We observed convergent transcription of S and AS transcripts in all retroTns investigated, suggesting that this is a global dsRNA formation mechanism in *Drosophila*. The experiments described here will provide the basis for future mechanistic studies of retroTn AS transcription and allow determination of the role of convergent transcription in retroTn gene silencing.

Acknowledgments

We thank Joshua Daugherty, Michael McKain, and Michael Hughes for technical assistance; Daniel Michalski for helpful discussions; and Ambrose R. Kidd, III and Lon Chubiz for critical review of the manuscript.

Literature Cited

Aravin, A. A., G. J. Hannon, and J. Brennecke, 2007 The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761–764.

Arkhipova, I. R., and Y. V. Ilyin, 1991 Properties of promoter regions of mdg1 Drosophila retrotransposon indicate that it belongs to a specific class of promoters. EMBO J. 10: 1169–1177.

Bingham, P. M., M. G. Kidwell, and G. M. Rubin, 1982 The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell 29: 995–1004.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128: 1089–1103.

Brouha, B., J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley *et al.*, 2003 Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. USA 100: 5280–5285.

Butler, J. E. F., and J. T. Kadonaga, 2002 The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev. 16: 2583–2592.

Cernilogar, F. M., M. C. Onorati, G. O. Kothe, A. M. Burroughs, K. M. Parsi *et al.*, 2011 Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. Nature 480: 391–395.

Chen, L., J. E. Dahlstrom, S.-H. Lee, and D. Rangasamy, 2012 Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. Epigenetics 7: 758–771.

Chung, W.-J., K. Okamura, R. Martin, and E. C. Lai, 2008 Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr. Biol. 18: 795–802.

Cordaux, R., and M. A. Batzer, 2009 The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. 10: 691–703.

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

Core, L. J., J. J. Waterfall, D. A. Gilchrist, D. C. Fargo, H. Kwak *et al.*, 2012 Defining the status of RNA polymerase at promoters. Cell Reports 2: 1025–1035.

Cruickshanks, H. A., and C. Tufarelli, 2009 Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics 94: 397–406.

Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde *et al.*, 2008 An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798–802.

Danilevskaya, O. N., K. L. Traverse, N. C. Hogan, P. G. DeBaryshe, and M. L. Pardue, 1999 The two Drosophila telomeric transposable elements have very different patterns of transcription. Mol. Cell. Biol. 19: 873–881.

Deininger, P., 2011 Alu elements: know the SINEs. Genome Biol. 12: 236–247.

dos Santos G., A. J. Schroeder, J. L. Goodman, V. B. Strelets, M. A. Crosby *et al.*, 2015 FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. **43:** D690–D697.

Dower, K., N. Kuperwasser, H. Merrikh, and M. Rosbash, 2004 A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. RNA 10: 1888–1899.

Ghildiyal, M., H. Seitz, M. D. Horwich, C. Li, T. Du *et al.*, 2008 Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science 320: 1077–1081.

Gogvadze, E., and A. Buzdin, 2009 Retroelements and their impact on genome evolution and functioning. Cell. Mol. Life Sci. 66: 3727–3742.

Grant, G. R., M. H. Farkas, A. D. Pizarro, N. F. Lahens, J. Schug *et al.*, 2011 Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics 27: 2518–2528.

Guil, S., and M. Esteller, 2012 Cis-acting noncoding RNAs: friends and foes. Nat. Struct. Mol. Biol. 19: 1068–1075.

Gullerova, M., and N. J. Proudfoot, 2012 Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. Nat. Struct. Mol. Biol. 19: 1193–1201.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura *et al.*, 2007 A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science 315: 1587–1590.

Henriques, T., D. A. Gilchrist, S. Nechaev, M. Bern, G. W. Muse *et al.*, 2013 Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell 52: 517–528.

Holoch, D., and D. Moazed, 2015 RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet. 16: 71–84.
Huang, X. A., H. Yin, S. Sweeney, D. Raha, M. Snyder *et al.*, 2013 A major epigenetic programming mechanism guided by piRNAs. Dev. Cell 24: 502–516.
Huisinga K. L., Elgin S. C. R., 2009 Small RNA directed heterochromatin formation in the context of development: what flies might learn from fission yeast. Biochimica et Biophysica Acta. 1789: 3–16.
Junakovic, N., C. Di Franco, M. Best-Belpomme, and G. Echalier, 1988 On the transposition of copia-like nomadic elements in cultured Drosophila cells. Chromosoma 97: 212–218.
Kaminker, J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas *et al.*, 2002 The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. 3: 1–20.
Kawamura, Y., K. Saito, T. Kin, Y. Ono, K. Asai *et al.*, 2008 Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature 453: 793–797.
Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle *et al.*, 2002 The human genome browser at UCSC. Genome Res. 12: 996–1006.
Kidwell, M. G., J. F. Kidwell, and J. A. Sved, 1977 Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86: 813–833.
Kofler, R., V. Nolte, and C. Schlötterer, 2015 Tempo and mode of transposable element activity in Drosophila. PLoS Genet. 11: e1005406.
Lapidot, M., and Y. Pilpel, 2006 Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep. 7: 1216–1222.
Le Thomas, A., A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao *et al.*, 2013 Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 27: 390–399.
Lee, M.-C., and C. J. Marx, 2013 Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. Genetics 193: 943–952.
Lerat, E., C. Rizzon, and C. Biémont, 2003 Sequence divergence within transposable element families in the Drosophila melanogaster genome. Genome Res. 13: 1889–1896.
Li, J., M. Kannan, A. L. Trivett, H. Liao, X. Wu *et al.*, 2014 An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. Nucleic Acids Res. 42: 4546–4562.
Maisonhaute, C., D. Ogereau, A. Hua-Van, and P. Capy, 2007 Amplification of the 1731 LTR retrotransposon in Drosophila melanogaster cultured cells: origin of neocopies and impact on the genome. Gene 393: 116–126.
Marques, J. T., K. Kim, P.-H. Wu, T. M. Alleyne, N. Jafari *et al.*, 2010 Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. Nat. Struct. Mol. Biol. 17: 24–30.
Martin M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17: 10–12.
Mätlik, K., K. Redik, and M. Speek, 2006 L1 antisense promoter drives tissue-specific transcription of human genes. J. Biomed. Biotechnol. 2006: 71753.
McClintock, B., 1950 The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36: 344–355.
Mills, R. E., E. A. Bennett, R. C. Iskow, and S. E. Devine, 2007 Which transposable elements are active in the human genome? Trends Genet. 23: 183–191.
Mizrokhi, L. J., S. G. Georgieva, and Y. V. Ilyin, 1988 jockey, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell 54: 685–691.
Nechaev, S., and D. C. Fargo, G. dos Santos, L. Liu, Y. Gao *et al.* 2010 Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science 327: 335–338.
Nigumann, P., K. Redik, K. Mätlik, and M. Speek, 2002 Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79: 628–634.
Okamura, K., S. Balla, R. Martin, N. Liu, and E. C. Lai, 2008a Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat. Struct. Mol. Biol. 15: 581–590.
Okamura, K., W.-J. Chung, J. G. Ruby, H. Guo, D. P. Bartel *et al.*, 2008b The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803–806.
Pelechano, V., and L. M. Steinmetz, 2013 Gene regulation by antisense transcription. Nat. Rev. Genet. 14: 880–893.
Picard, G., J. C. Bregliano, A. Bucheton, J. M. Lavige, A. Pélisson *et al.*, 1978 Non-mendelian female sterility and hybrid dysgenesis in Drosophila melanogaster. Genet. Res. 32: 275–287.
Potter, S. S., W. J. Brorein, P. Dunsmuir, and G. M. Rubin, 1979 Transposition of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 17: 415–427.
Purcell, M. K., S. A. Hart, G. Kurath, and J. R. Winton, 2006 Strand-specific, real-time RT-PCR assays for quantification of genomic and positive-sense RNAs of the fish rhabdovirus, Infectious hematopoietic necrosis virus. J. Virol. Methods 132: 18–24.
Rozhkov, N. V., M. Hammell, and G. J. Hannon, 2013 Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev. 27: 400–412.
Rubin, G. M., M. G. Kidwell, and P. M. Bingham, 1982 The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29: 987–994.
Saito, K., and M. C. Siomi, 2010 Small RNA-mediated quiescence of transposable elements in animals. Dev. Cell 19: 687–697.
Saito, K., K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi *et al.*, 2006 Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 20: 2214–2222.
Schneider, I., 1972 Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. 27: 353–365.
Seitz, H., M. Ghildiyal, and P. D. Zamore, 2008 Argonaute loading improves the 5′ precision of both MicroRNAs and their miRNA* strands in flies. Curr. Biol. 18: 147–151.
Sentmanat, M. F., and S. C. R. Elgin, 2012 Ectopic assembly of heterochromatin in Drosophila melanogaster triggered by transposable elements. Proc. Natl. Acad. Sci. USA 109: 14104–14109.
Sienski, G., D. Dönertas, and J. Brennecke, 2012 Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151: 964–980.
Siomi, M. C., K. Saito, and H. Siomi, 2008 How selfish retrotransposons are silenced in Drosophila germline and somatic cells. FEBS Lett. 582: 2473–2478.
Speek, M., 2001 Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol. Cell. Biol. 21: 1973–1985.
Sullivan, K. D., M. Steiniger, and W. F. Marzluff, 2009 A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Mol. Cell 34: 322–332.
Szak, S. T., O. K. Pickeral, W. Makalowski, M. S. Boguski, D. Landsman *et al.*, 2002 Molecular archeology of L1 insertions in the human genome. Genome Biol. 3: 52.

Tchurikov, N. A., Y. V. Ilyin, K. G. Skryabin, E. V. Ananiev, A. A. Bayev *et al.*, 1981 General properties of mobile dispersed genetic elements in Drosophila melanogaster. Cold Spring Harb. Symp. Quant. Biol. 45(Pt 2): 655–665.

Tomari, Y., and P. D. Zamore, 2005 Perspective: machines for RNAi. Genes Dev. 19: 517–529.

Tomari, Y., T. Du, and P. D. Zamore, 2007 Sorting of Drosophila small silencing RNAs. Cell 130: 299–308.

Vagin, V. V., A. Sigova, C. Li, H. Seitz, V. Gvozdev *et al.*, 2006 A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313: 320–324.

Vashist, S., L. Urena, and I. Goodfellow, 2012 Development of a strand specific real-time RT-qPCR assay for the detection and quantitation of murine norovirus RNA. J. Virol. Methods 184: 69–76.

Volpe, T. A., C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal *et al.*, 2002 Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297: 1833–1837.

Wang, S. H., and S. C. R. Elgin, 2011 Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc. Natl. Acad. Sci. USA 108: 21164–21169.

Wen, J., J. Mohammed, D. Bortolamiol-Becet, H. Tsai, N. Robine *et al.*, 2014 Diversity of miRNAs, siRNAs, and piRNAs across 25 Drosophila cell lines. Genome Res. 24: 1236–1250.

Xie, W., R. C. Donohue, and J. A. Birchler, 2013 Quantitatively increased somatic transposition of transposable elements in Drosophila strains compromised for RNAi. PLoS One 8: e72163.

Yang, N., and H. H. Kazazian, 2006 L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat. Struct. Mol. Biol. 13: 763–771.

Yu, R., G. Jih, N. Iglesias, and D. Moazed, 2014 Determinants of heterochromatic siRNA biogenesis and function. Mol. Cell 53: 262–276.

Communicating editor: J. Birchler

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177196/-/DC1

Antisense Transcription of Retrotransposons in *Drosophila*: An Origin of Endogenous Small Interfering RNA Precursors

Joseph Russo, Andrew W. Harrington, and Mindy Steiniger

DOI: 10.1534/genetics.115.177196

Table S1: HTS mapping statistics

Sample	total # reads	% mapping	read depth	% unique	*p*-value	% non-unique	*p*-value
RNA-seq							
LacZ1	30020327	98.5	99.7	57.0	6.949E-05	41.500	9.582E-05
LacZ2	35561677	98.4	118.1	56.3		42.100	
LacZ3	26072570	98.4	86.6	56.4		42.000	
Dcr2-1	22668363	98.2	75.3	51.8		46.500	
Dcr2-2	23826043	98.0	79.2	51.3		46.700	
Dcr2-3	24236601	98.0	80.5	51.1		46.900	
smRNA-seq							
LacZ1	341316	99.1		64.5	1.209E-03	34.600	1.544E-03
LacZ2	288286	99.1		63.4		35.700	
LacZ3	285543	99.2		66.1		33.000	
Dcr2-1	207361	99.7		77.6		22.100	
Dcr2-2	272897	99.7		76.5		23.1	
Dcr2-3	196054	99.7		76.7		23.0	

The total number of reads, percent mapping, read depth, percent unique and percent non-unique are shown for each high-throughput sequencing sample. Read depth = total # reads*100/30.1 Mb. A Students T-test was performed to determine if differences observed between % unique or % non-unique for Dcr2 and LacZ samples were statistically significant. The *p*-values for these tests are indicated.

Table S2: Highly transcribed genes show little AS transcription

Gene	S RPM	AS RPM
GAPDH1	573	3
GAPDH2	615	1
Groucho	48	3
Armadillo	372	0
Pumillio	208	0
Succinate Dehydrogenase A (SdhA)	108	0
RpL 32	3850	11
RpL 0	2329	0
Stem Loop Binding Protein (SLBP)	69	2
Cleavage and Polyadenylation Specificy Protein (CPSF) 100	35	3

S and AS RPM for highly expressed genes.

Table S3: Non-LTR retroTns and TIR Tns

non-LTR

RetroTn	Family	Size (bp)	Genomic location	S (RPM)	AS (RPM)
BS{}707	Jockey	5128	2R:1984289-1989416		
BS{}1260	Jockey	5142	3R:3869566-3874707		
BS{}1292	Jockey	5122	3R:7666843-7671964		
Doc{}772	Jockey	4719	2R:3756749-3761467	+(203)	+(33)
Doc{}819	Jockey	4726	2R:9024402-9029127	+	+
Doc{}827	Jockey	4721	2R:10281792-10286512	+	+
F{}731	Jockey	4699	2R:2199985-2204683	+(78)	+(37)
F{}755	Jockey	4707	2R:2382090-2386796	+	+
F{}763	Jockey	4710	2R:3084132-3088841	+	+
I{}769	I	5133	2R:3495977-3501109		
I{}129	I	5371	X:14530558-14535928		
I{}18	I	1727	X:1461750-1463476		
jockey{}817	Jockey	5010	2R:8707006-8712015	+(265)	++(108)
jockey{}838	Jockey	4959	2R:13034810-13039768	+	++
jockey{}277	Jockey	5006	2L:47514-52519	+	++
Juan{}768	Jockey	4236	2R:3322453-3326688	++(662)	+++++(241)
Juan{}138	Jockey	4226	X:15307142-15311367	++	+++++
Juan{}1190	Jockey	4232	3R:234195-238426	++	+++++
Rt1a{}1276	R1	5177	3R:5104562-5109738		
Rt1a{}905	R1	5193	3L:1424822-1430014		
Rt1a{}1390	R1	5175	3R:15570024-15575198		
Rt1b{}334	R1	5171	2L:10138214-10143384		+(32)
Rt1b{}1218	R1	5027	3R:1164177-1169203		+
Rt1b{}1288	R1	5170	3R:6783807-6788976		+

TIR

Tn	Family	Size (bp)	Genomic location	S (RPM)	AS (RPM)
1360{}1226	protop	1107	3R:1648157-1649263	+(80)	+(31)
1360{}136	protop	1107	X:15167794-15168900	+	+
1360{}1498	protop	1084	4:315271-316354	+	+
Bari1{}1534	Tc1	1728	4:860624-862351		
Bari1{}282	Tc1	1728	2L:770516-772243		
Bari1{}1409	Tc1	1728	3R:19384173-19385900		
HB{}276	Tc1	1573	X:22243764-22245336		
HB{}512	Tc1	1636	2L:21673477-21675112		
HB{}761	Tc1	1633	2R:2876633-2878265		
hopper{}82	Transib	1433	X:7739814-7741246	+(48)	+(21)
hopper{}105	Transib	1435	X:11162651-11164085	+	+
hopper{}1432	Transib	1432	X:21841727-21843158	+	+
mariner2{}1130	Tc1	1110	3L:23131582-23132691		
mariner2{}767	Tc1	983	2R:3281653-3282635		
mariner2{}522	Tc1	879	2L:22004569-22005447		
pogo{}297	Pogo	2122	2L:2933354-2935475	++(785)	+(28)
pogo{}400	Pogo	2134	2L:17912012-17914134	++	+
pogo{}1294	Pogo	2122	3R:7848660-7850781	++	+
S{}173	Tc1	1731	X:19604159-19605889		+(23)
S{}758	Tc1	1735	2R:2551739-2553473		+
S{}1207	Tc1	1704	3R:508954-510657		+
FB{}1449	Tc1	1310	3R:24905161-24906470		+(26)
FB{}2296	Tc1	1592	2L:22250953-22252544		+
FB{}5386	Tc1	1325	4:601549-602873		+

Family, size, genomic location, S and AS transcription levels are shown for Individual Tns. The '+' represent relative transciption among Tns and the number shown in ( ) is the normalized non-unique read count (RPM) corresponding to that individual Tn.

Table S3: LTR

LTR

retroTn	Family	Size (bp)	Genomic location	S	AS
17.6{}790	Gypsy	7494	2R:5614174-5621667	+++(1120)	+++(469)
17.6{}804	Gypsy	7494	2R:6835588-6843081	+++	+++
17.6{}1287	Gypsy	7475	3R:6629950-6637424	+++	+++
297{}832	Gypsy	6992	2R:10972539-10979530	+++++(3848)	++++(571)
297{}388	Gypsy	6997	2L:16153791-16160787	+++++	++++
297{}407	Gypsy	6978	2L:19147425-19154402	+++++	++++
3S18{}853	Pao	6130	2R:14463358-14469487	++++(2171)	+(95)
3S18{}4	Pao	6127	X:322507-328633	++++	+
3S18{}35	Pao	6127	X:3309106-3315232	++++	+
412{}880	Gypsy	7567	2R:19801877-19809443	+(54)	+(40)
412{}881	Gypsy	7521	2R:20034646-20042166	+	+
412{}882	Gypsy	7428	2R:20064814-20072241	+	+
blood{}852	Gypsy	7413	2R:14375381-14382793	++++(1531)	+++++(694)
blood{}856	Gypsy	7412	2R:15603415-15610826	+++	+++++
blood{}280	Gypsy	7443	2L:347941-355383	+++	+++++
Burdock{}770	Gypsy	6412	2R:3703232-3709643	+(188)	++(152)
Burdock{}783	Gypsy	6413	2R:5038862-5045274	+	++
Burdock{}514	Gypsy	6413	2L:21691882-21698294	+	++
copia{}631	Copia	5145	2R:1105258-1110402	>+++++(64258)	+++(486)
copia{}837	Copia	5151	2R:12427198-12432348	>+++++	+++
copia{}840	Copia	5146	2R:13124069-13129214	>+++++	+++
diver{}782	Pao	6133	2R:4665472-4671604	++(785)	++(163)
diver{}839	Pao	6132	2R:13063915-13070046	++	++
diver{}873	Pao	6112	2R:18467972-18474083	++	++
HMS-Beagle{}318	Gypsy	7062	2L:6991487-6998548		
HMS-Beagle{}333	Gypsy	7062	2L:9973781-9980842		
HMS-Beagle{}333	Gypsy	7072	2L:12558375-12565446		
Invader2{}633	Gypsy	5075	2R:1115288-1120362		
Invader2{}563	Gypsy	5045	2L:22329289-22334333		
Invader2{}1169	Gypsy	5056	3L:23264861-23269916		
invader3{}695	Gypsy	5474	2R:1509626-1515099		
invader3{}240	Gypsy	5382	X:21818924-21824305		
invader3{}751	Gypsy	5477	2R:2358324-2363800		
mdg1{}831	Gypsy	7367	2R:10903017-10910383	+++(1192)	+++++(788)
mdg1{}859	Gypsy	7355	2R:15802405-15809759	+++	+++++
mdg1{}885	Gypsy	7451	2R:20615462-20622912	+++	+++++
mdg3{}119	Gypsy	5520	X:13357733-13363252	+++(1153)	++(140)
mdg3{}144	Gypsy	5520	X:16386734-16392253	+++	++
mdg3{}291	Gypsy	5520	2L:1801273-1806792	+++	++
opus{}760	Gypsy	7525	2R:2839724-2847248	+(106)	++(124)
opus{}821	Gypsy	7602	2R:9615692-9623293	+	++
opus{}127	Gypsy	7604	X:14445732-14453335	+	++
Quasimodo{}352	Gypsy	7387	2L:12781858-12789244		+(56)
Quasimodo{}360	Gypsy	7379	2L:13449517-13456895		+
Quasimodo{}1186	Gypsy	7355	3R:84350-91704		+
roo{}796	Pao	9109	2R:6064440-6073548	+(184)	++(203)
roo{}806	Pao	9116	2R:6897375-6906490	+	++
roo{}828	Pao	9094	2R:10354854-10363947	+	++
rover{}1212	Gypsy	7320	3R:713256-720575		
rover{}1275	Gypsy	7412	3R:4732111-4739522		
rover{}133	Gypsy	7470	X:14928180-14935649		
springer{}300	Gypsy	7510	2L:3251549-3259058		
springer{}1464	Gypsy	7543	3R:26900006-26907548		
springer{}59	Gypsy	7510	X:4990361-4997870		
Stalker{}174	Gypsy	7256	X:19691436-19698691	+(66)	+(35)
Stalker{}1277	Gypsy	7230	3R:5130306-5137535	+	+
Stalker{}1427	Gypsy	7255	3R:22384105-22391359	+	+
Stalker2{}1505	Gypsy	8119	4:340670-348788	++(414)	+(80)
Stalker2{}22	Gypsy	7883	X:1848974-1856856	++	+
Stalker2{}1042	Gypsy	7895	3L:18703914-18711808	++	+
tirant{}797	Gypsy	8527	2R:6473118-6481644	++(602)	++(203)
tirant{}833	Gypsy	8425	2R:11006877-11015301	++	++
tirant{}834	Gypsy	8527	2R:11228251-11236777	++	++
Transpac{}32	Gypsy	5249	X:2969722-2974970	+(182)	+(42)
Transpac{}1439	Gypsy	5248	3R:23688731-23693978	+	+
Transpac{}362	Gypsy	5249	2L:13522313-13527561	+	+

Table S4: Individual non-LTR retroTns

retroTn	Length (bp)	Genomic location	UI-S	UI-AS	S-S	S-AS	intergenic	intron	S/AS to Tn?
juan{}768	4236	2R:3322453-3326688			x		x		
juan{}138	4226	X:15307142-15311367	x	x	x			CG18210	AS
juan{}1190	4232	3R:234195-238426	x	x	x	x		CG32944	AS
juan{}257	4235	X:21953676-21957910	x	x			x		
juan{}83	4235	X:7925528-7929762				x	x		
juan{}266	4230	X:22099389-22103618			x	x	x		
juan{}2251	4236	2R:2298595-2302830				x	x		
jockey{}765	5014	2R:3123380-3128393	x	x			x		
jockey{}807	4985	2R:6907459-6912443	x	x				luna	S
jockey{}817	5010	2R:8707006-8712015		x			x		
jockey{}838	4959	2R:13034810-13039768	x	x			x		
jockey{}277	5006	2L:47514-52519			x			CG31973	AS
jockey{}307	5002	2L:4918233-4923234	x		x			hoe1	AS
jockey{}477	5017	2L:20891618-20896634			x			CG9339	AS
jockey{}1238	5017	3R:2335803-2340819	x				x		
jockey{}1447	5015	3R:24618648-24623662					x		
jockey{}51	5018	X:4625918-4630935	x	x			x		
jockey{}261	5023	X:21966895-21971917	x	x	x		x		
jockey{}973	5011	3L:9569071-9574081						CG32048	AS
jockey{}1086	5087	3L:21742484-21747570						Syn1	S
Jockey{}1630	14127	2R:14245969-14260095					x		

Size and genomic location of all full-length juan and jockey elements are shown. The presence of unique internal S or AS reads (UI-S or UI-AS) or, S or AS reads corresponding to retroTn-external sequence junctions produced by splicing (S-S or S-AS) are shown with an 'x.' Whether the retroTn is intergenic or intragenic is shown. If the retroTn is contained within an intron, the gene name is shown. The last column describes whether an intronic retroTn is S or AS to the S strand of the host gene. 'S' indicates that S retroTn RNA is also the S mRNA while 'AS' indicates that the AS retroTn transcript is also the S mRNA.

Table S4: Individual LTR retroTns

retroTn	Length (bp)	Genomic location	UI-S	UI-AS	S-S	S-AS	intergenic	intron	S/AS to Tn?
297{}832	6992	2R:10972539-10979530	x		x		x		
297{}388	6997	2L:16153791-16160787	x	x	x	x	x		
297{}407	6978	2L:19147425-19154402	x	x	x			brat	S
297{}1440	6998	3R:23701477-23708474	x		x		x		
297{}48	6978	X:4051980-4058957	x		x	x		Fas2	AS
297{}92	6995	X:9692611-9699605			x			CG32698	AS
297{}98	6996	X:10419159-10426154			x			spri	AS
197{}109	7013	X:11527505-11534517		x				ptp10D	AS
297{}327	6997	2L:8594520-8601516	x					Sema-1a	AS
297{}338	6999	2L:10876598-10883596		x	x	x	x		
297{}346	6995	2L:12067385-12074379	x	x			x		
297{}376	6995	2L:15586465-15593459	x		x		x		
297{}766	6996	2R:3149747-3156742	x		x	x	x		
297{}897	6996	3L:402740-409735	x	x	x	x	x		
297{}950	6995	3L:7786867-7793861			x	x		CG32369	S
297{}1107	6917	3L:22983017-22989933		x	x	x	x		
297{}323	6917	2L:7977135-7984051	x		x			snoo	AS
297{}1286	6917	3R:6167388-6174304	x			x	x		
mdg1{}831	7367	2R:10903017-10910383	x		x			hbs	AS
mdg1{}859	7355	2R:15802405-15809759	x	x			x		
mdg1{}885	7451	2R:20615462-20622912						CG33988	AS
mdg1{}299	7372	2L:3202794-3210165	x				x		
mdg1{}305	7384	2L:4601759-4609142	x			x	x		
mdg1{}1280	7365	3R:5663885-5671249	x	x	x			Teh1	AS
mdg1{}1403	7335	3R:17586899-17594233			x			CG42335	AS
mdg1{}1442	7373	3R:23999447-24006819				x		CG34354	S
mdg1{}900	7369	3L:925231-932599	x					Glut1	S
mdg1{}1047	7390	3L:19110198-19117587	x			x	x		
mdg1{}1678	7403	2R:6546757-6554159	x				x		
mdg1{}29	7353	X:2738988-2746340	x	x			x		
mdg1{}914	7369	3L:2966778-2974144	x				x		
mdg1{}1610	7273	3L:17724006-17731278		x				C cn	AS
mdg1{}1720	18802	2R:6509904-6528705	x	x		x		CG11883	S
blood{}852	7413	2R:14375381-14382793	x		x			CG30116	AS
blood{}856	7412	2R:15603415-15610826	x			x	x		
blood{}280	7443	2L:347941-355383	x		x	x	x		
blood{}285	7409	2L:1220184-1227592	x	x	x	x		CG42329	S
blood{}289	7408	2L:1679032-1686439		x	x			chinmo	S
blood{}335	7413	2L:10156377-10163789		x			x		
blood{}344	7411	2L:11713562-11720972			x	x	x		
blood{}356	7410	2L:12933905-12941314				x	x		
blood{}375	7412	2L:15446106-15453517	x		x		x		
blood{}468	7411	2L:20303216-20310626	x				x		
blood{}472	7411	2L:20576274-20583684	x		x		x		
blood{}488	7415	2L:21347605-21355019	x					Tsp39D	S
blood{}1369	7417	3R:11444508-11451924	x		x		x		
blood{}1376	7409	3R:13401108-13408516	x		x		x		
blood{}1389	7411	3R:15531296-15538706			x		x		
blood{}1462	7412	3R:26505296-26512707	x	x		x	x		
blood{}1470	7413	3R:27823769-27831181	x		x	x	x		
blood{}218	7415	X:21407620-21415034	x	x			x		
blood{}409	7417	2L:19347541-19354957	x		x			dnt	AS
blood{}959	7410	3L:8491452-8498861					x		
blood{}1092	7413	3L:22548925-22556337	x	x	x		x		
blood{}1094	7395	3L:22610345-22617739	x	x	x	x	x		

See **Table S4:** Individual non-LTR retroTns for legend.

Table S5: Strand specific qPCR primers

Description	5' to 3' Sequence	Position (bp)
Dm297-RT RT sense primer*	caagactcagctggttcctgcgacttcttcttcttcaagc	4674-4693
Dm297-RT sense qPCR-F*	ggcagacagagacggag	4629-4645
Dm297-RT sense qPCR-R (tag)*	caagactcagctggttcctg	unique tag
Dm297-RT RT antisense primer*	gagaagctcatagtacctcgggcagacagagacggag	4629-4693
Dm297-RT antisense qPCR-F (tag)*	gagaagctcatagtacctcg	unique tag
Dm297-RT antisense qPCR-R*	cgacttcttcttcttcaagc	4673-4693
Dm297-env RT sense primer	gcctgtcccgatatatgaacctcaataatgtcgttgtg	6317-6335
Dm297-env sense qPCR-F	gacaccactatacacaccac	6269-6289
Dm297-env sense qPCR-R (tag)	gcctgtcccgatatatgaac	unique tag
Dm297-env RT antisense primer	gttattaatcgtaataacgggacaccactatacacaccac	6269-6288
Dm297-env antisense qPCR-F (tag)	gttattaatcgtaataacgg	unique tag
Dm297-env antisense qPCR-R	ctcaataatgtcgttgtg	6317-6335
blood-ORFII RT sense primer	ccagaaaaccgcgtgtctacgctgcttacgcatactgtc	2624-2643
blood-ORFII sense qPCR-F	cgtaaaaggcgaatcgcctg	2534-2554
blood-ORFII sense qPCR-R (tag)	ccagaaaaccgcgtgtctac	unique tag
blood-ORFII RT antisense primer	cccatacgcgagatacactgcgtaaaaggcgaatcgcctg	2534-2554
blood-ORFII antisense qPCR-F (tag)	cccatacgcgagatacactg	unique tag
blood-ORFII antisense qPCR-R	gctgcttacgcatactgtc	2624-2643
blood-RT RT sense primer*	ctcgtcgctttcggatttgccaaagcctcgttaagtggcg	4726-4746
blood-RT sense qPCR-F*	cctataccaacagatgccgac	4647-4668
blood-RT sense qPCR-R (tag*)	ctcgtcgctttcggatttgc	unique tag
blood-RT RT antisense primer*	gactgcagacatcagatcggcctataccaacagatgccgac	4647-4668
blood-RT antisense qPCR-F (tag)*	gactgcagacatcagatcgg	unique tag
blood-RT antisense qPCR-R*	caaagcctcgttaagtggcg	4726-4746
mdg1-ORFII RT sense primer	cgtttaaaccagaccgacaccgggtaatgttattaccgctg	2133-2154
mdg1-ORFII sense qPCR-F	ctgagatcggtgaggatatcg	2053-2074
mdg1-ORFII sense qPCR-R (tag)	cgtttaaaccagaccgacac	unique tag
mdg1-ORFII RT antisense primer	ggcacacttatgctcagcacctgagatcggtgaggatatcg	2053-2074
mdg1-ORFII antisense qPCR-F (tag)	ggcacacttatgctcagcac	unique tag
mdg1-ORFII antisense qPCR-R	cgggtaatgttattaccgctg	2133-2154
mdg1-RT RT sense primer*	ctacgatgccgctaagaacctcctgctctgtagtggac	4923-4942
mdg1-RT sense qPCR-F*	gtaaacaagcatgtggagcg	4824-4844
mdg1-RT sense qPCR-R (tag)*	ctacgatgccgctaagaacc	unique tag
mdg1-RT RT antisense primer*	gatcggcgaccatttgtgaggtaaacaagcatgtggagcg	4824-4844
mdg1-RT antisense qPCR-F (tag)*	gatcggcgaccatttgtgag	unique tag
mdg1-RT antisense qPCR-R*	ctcctgctctgtagtggac	4923-4942
jockey-gag RT sense primer	gcctagaattacctaccgcgtgctccatattctccgtttcag	919-897
jockey-gag sense qPCR-F	acctatcctcacccttctc	776-795
jockey-gag sense qPCR-R (tag)	gcctagaattacctaccgcg	unique tag
jockey-gag RT antisense primer	ctacgttacagcgtgcatagacctatcctcacccttctc	776-795
jockey-gag antisense qPCR-F (tag)	ctacgttacagcgtgcatag	unique tag
jockey-gag antisense qPCR-R	tgctccatattctccgtttcag	919-897
jockey-RT RT sense primer*	gcctagaattacctaccgcgggaagttgaagtggtctgaag	2922-2943
jockey-RT sense qPCR-F*	gtggacattgataatgccacaag	2841-2864
jockey-RT sense qPCR-R (tag)*	gcctagaattacctaccgcg	unique tag
jockey-RT RT antisense primer*	ctacgttacagcgtgcataggtggacattgataatgccacaag	2841-2864
jockey-RT antisense qPCR-F (tag)*	ctacgttacagcgtgcatag	unique tag
jockey-RT antisense qPCR-R*	ggaagttgaagtggtctgaag	2922-2943
juan-ORFI RT sense primer	gctgctcttatcacatttgccctaggttttgtagcatggatttg	586-609
juan-ORFI sense qPCR-F	ctgtgagttctacacgtacgatac	499-522
juan-ORFI sense qPCR-R (tag)	gctgctcttatcacatttgc	unique tag
juan-ORFI RT antisense primer	gccagtcgtattccttctcgctgtgagttctacacgtacgatac	499-522
juan-ORFI antisense qPCR-F (tag)	gccagtcgtattccttctcg	unique tag
juan-ORFI antisense qPCR-R	cctaggttttgtagcatggatttg	586-609
juan-RT RT sense primer*	gctgctcttatcacatttgcctgtgagcagttgacaaccac	2168-2189
juan-RT sense qPCR-F*	gcgcaatgtaaaaacatatccg	2082-2104
juan-RT sense qPCR-R (tag)*	gctgctcttatcacatttgc	unique tag
juan-RT RT antisense primer*	gccagtcgtattccttctcggcgcaatgtaaaaacatatccg	2082-2104
juan-RT antisense qPCR-F (tag)*	gccagtcgtattccttctcg	unique tag
juan-RT antisense qPCR-R*	ctgtgagcagttgacaaccac	2168-2189
Actin5C RT sense primer*	gtgcgtacaccttaatcaccgggtgccacacgcagctcat	280-298
Actin5C sense qPCR-F*	ggcgcagagcaagcgtggta	175-195
Actin5C sense qPCR-R (tag)*	gtgcgtacaccttaatcacc	unique tag
18s RT sense primer*	ctctcctcctcagcatgctgaccagacttgccctccaat	553-571
18s sense qPCR-F*	ctgagaaacggctaccacatc	400-422
18s sense qPCR-R (tag)*	ctctcctcctcagcatgctg	unique tag

'*' indicates that primers were used for detection of strand-specific transcripts in poly A+/- fractions.

Table S6: Strand specific RT-qPCR of retroTn S and AS transcripts

LTR:

Tn-amplicon	Ave. ΔCt(S-AS)	SD
Dm297-RT	-5.23	0.18
Dm297-env	-4.38	0.12
blood-ORFII	-4.31	0.10
blood-RT	-0.86	0.08
mdg1-ORFII	-3.02	0.13
mdg1-RT	-2.00	0.07

Non-LTR:

Tn-amplicon	Ave. ΔCt(S-AS)	SD
juan-ORF1	-2.44	0.12
juan-RT	-2.01	0.22
jockey-gag	-0.46	0.47
jockey-RT	1.86	0.23

Strand specific qPCR with primers to multiple retroTn ORFs (first column) was performed in triplicate. The average difference and between S and AS Ct values and the standard deviation (SD) were calculated.

Table S7: Northern probes for detection of transcripts

Description	Oligonucleotide sequence (5’ to 3’)
Dm297 sense probe	gatgagtcttgctttaagggtaggccaatcttcgatgttcggaagtcccaaa
Dm297 antisense probe	ttgatttttagtcttaagctgagatccaaagaataaagtcgtgaaactatt
blood sense probe	aattcccaaatcaaatcggcaatattagcagcatttcctcagtagtcctcaga
blood antisense probe	gacactctgtagaggttaagcgggcagaaccgtttctgctactcgaagagat
mdg1 sense probe	tcccatcacactgacactctactcactcagatcgctttttttcttcataattgcc
mdg1 antisense probe	acaccctaatactaaatatgcgaattcagcatgtacgcctttaggggtcgcac
jockey sense probe	gcaaccttggtcctgaacgcttgctgaatatttgtgatgtgcctgctgaag
jockey antisense probe	cttcagcaggcacatcacaaatattcagcaagcgttcaggaccaaggttgc
juan sense probe	gtaggcaatgagatctggggttgatttccaaagagagcagatggagcgatg
juan antisense probe	catcgctccatctgctctctttggaaatcaaccccagatctcattgcctac