

Lariat intronic RNAs in the cytoplasm of vertebrate cells

Gaëlle J. S. Talhouarne^{a,b,1} and Joseph G. Gall^{a,2}

^aDepartment of Embryology, Carnegie Institution for Science, Baltimore, MD 21218; and ^bDepartment of Biology, Johns Hopkins University, Baltimore, MD 21218

Contributed by Joseph G. Gall, July 5, 2018 (sent for review May 22, 2018; reviewed by Lynne E. Maquat and Sandra L. Wolin)

Most intronic RNAs are degraded within seconds or minutes after their excision from newly formed transcripts. However, stable intronic sequence RNAs (sisRNAs) have been described from oocytes of the frog *Xenopus*, from *Drosophila* embryos, and from human cell lines. In *Xenopus* oocytes, sisRNAs are abundant in both the nucleus and cytoplasm, they occur in the form of lariats, and they are stable for days. In this study we demonstrate that cytoplasmic sisRNAs are also found in human, mouse, chicken, and zebrafish cells. They exist as circular (lariat) molecules, mostly 100–500 nucleotides in length, and are derived from many housekeeping genes. They tend to have an unusual cytosine branchpoint (with the exception of those from the frog). Stable lariats are exported from the nucleus to the cytoplasm by the NXF1/NXT1 system, demonstrating that their presence in the cytoplasm is not due to passive diffusion. Lariats in the cytoplasm are not associated with transcripts of the genes from which they are derived. The biological significance of cytoplasmic sisRNAs remains obscure.

RNA | intron | cytoplasm | vertebrate

Intronic RNAs represent a large fraction of the unstable transcriptome of higher eukaryotes. These RNAs are spliced from pre-mRNA as branched circular RNAs (lariats). Within minutes of splicing, a dedicated debranching enzyme (DBR1) recognizes the branchpoint and linearizes the lariat, promoting its rapid degradation (1). This pathway is also necessary for processing noncoding RNAs such as small nucleolar RNAs (2). An earlier study from our laboratory revealed the existence of unusually abundant intronic RNAs in the nuclei of *Xenopus* oocytes (3). Many of these intronic RNAs persist for as long as 2 d and were given the name stable intronic sequence RNA, or sisRNA. In a later study we found thousands of different sisRNAs in the cytoplasm of the *Xenopus* oocyte (4). We were certain of the cytoplasmic localization, because the gigantic nucleus can be removed manually from the oocyte before the cytoplasmic RNAs are isolated and sequenced. The cytoplasmic sisRNAs are stored as circular molecules (lariats without tails).

Until now, however, it has remained unknown whether cytoplasmic circular sisRNAs are unique to frog oocytes or are more widely distributed. A major technical problem in searching for cytoplasmic sisRNAs is the difficulty of obtaining cytoplasmic fractions completely free of nuclear contamination. For this reason, we began our search for cytoplasmic sisRNAs in other species by examining mammalian red blood cells (RBCs). Mammalian RBCs lose their nuclei during their maturation from precursor erythroblasts and so provide a convenient source of pure cytoplasm. We found numerous circular intronic sequences in RBCs, thereby demonstrating the existence of sisRNA in mammalian somatic cells. We also found circular sisRNAs in cultured cells and in tissues from the mouse, human, chicken, and zebrafish. We discuss functions that cytoplasmic circular sisRNAs may perform independently of their host genes.

Results

Circular SisRNAs in RBCs. Mammalian RBCs extrude their nuclei before entering the circulation and hence are a source of pure cytoplasm. Although extensive RNA degradation occurs during

RBC maturation (5), some RNAs persist (6), including circular RNA or circRNA (7, 8). Intronic RNA has not yet been analyzed in these cells. We purified RBCs from mouse blood by fractionation on a Percoll gradient (9) and subjected the small amount of recovered RNA to high-throughput sequencing. The 77 million single-end 100-bp reads that we obtained were analyzed using a method developed by Taggart et al. (10). Of particular interest, this method allows one to recover inverted reads that would otherwise be thrown out by standard pipelines. We searched for and found 524 mouse introns that were mapped by conventional reads and also by inverted reads that contained sequences from both ends of the intron (Fig. 1*A* and Dataset S1). Because we were looking for circular sisRNA, we disregarded any intron with conventional reads but no inverted reads. We confirmed the circularity of the intronic RNA by RT-PCR and Northern blotting of selected RNAs (Fig. 1*B* and *C*). Two classes of circular sisRNA were detected. The most abundant circular sisRNAs presumably arise from failure of the excised intronic RNA to be linearized by the debranching enzyme DBR1. An unusual feature of these stable lariats was the presence of a cytosine at the branchpoint rather than the more common adenine (Fig. 1*D*). Approximately 80% of the stable lariats contained cytosine at the branchpoint compared with the 10% expected rate (10). On the other hand, a fraction of the circular sisRNAs did not resemble conventional spliced lariats. Specifically, the junctions of these

Significance

Introns are noncoding DNA sequences interspersed among the coding sequences of genes. Shortly after transcription, the intronic sequences are spliced out of the primary RNA transcript as lariat RNAs (circular molecules with a short tail). Most of these lariats are destroyed within minutes in the cell nucleus. We report here that many such intronic RNAs are, in fact, exported to the cytoplasm, where they remain as stable circular molecules. These cytoplasmic introns are derived from hundreds of different genes of widely different functions. We find them in cells of human, mouse, chicken, frog, and zebrafish. The widespread occurrence of so many stable lariat RNAs in the cytoplasm suggests that they play some as-yet unexpected role in cell metabolism.

Author contributions: G.J.S.T. and J.G.G. designed research; G.J.S.T. and J.G.G. performed research; G.J.S.T. contributed new reagents/analytic tools; G.J.S.T. and J.G.G. analyzed data; and G.J.S.T. and J.G.G. wrote the paper.

Reviewers: L.E.M., University of Rochester; and S.L.W., National Cancer Institute.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: RNA sequencing data have been deposited in the NCBI Sequence Read Archive (SRA BioProject ID [PRJNA479418](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA479418)).

¹Present address: Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520.

²To whom correspondence should be addressed. Email: gall@carnegiescience.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808816115/-DCSupplemental.

Published online August 6, 2018.

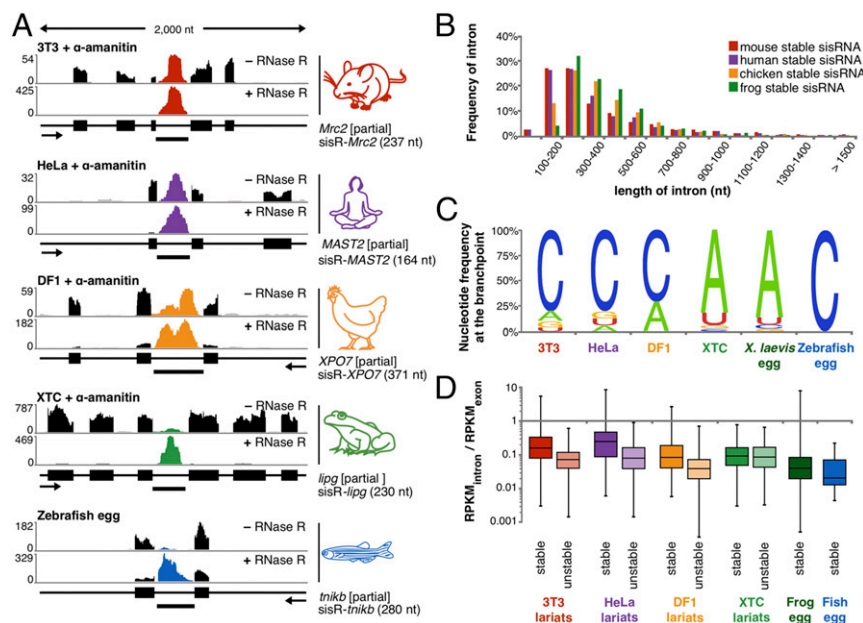


Fig. 2. Identification of lariat RNAs in five vertebrate species. (A) A typical prominent sisRNA detected in cultured cells of mouse (red), human (purple), chicken (orange), and frog (green) and zebrafish eggs (blue). RNase R was used to eliminate mRNA transcripts (black). Transcript annotation and orientation are shown below the coverage for each gene. (B) Percent of stable circular RNAs as a function of intron length (50-nt bins). Mouse (red), human (purple), chicken (orange), frog (green), and zebrafish (blue) sisRNAs derive from short introns. (C) Nucleotide frequency at the branchpoint of stable lariats. (D) Abundance of “stable” and “unstable” circular intronic RNA relative to their cognate mRNA in untreated cells.

sensitive to α -amanitin (*SI Appendix, Fig. S3A*). In untreated cells, unstable lariats are less abundant relative to their cognate mRNA than are stable lariats (Fig. 2D). Also, we find that these two groups have distinct branchpoint signatures. Most stable lariats have a C branchpoint at a frequency similar to the C branchpoint detected in α -amanitin-treated cells, as expected. However, most unstable lariats have an A branchpoint (*SI Appendix, Fig. S3B*). Furthermore, we found that for equally abundant lariats, a C branchpoint predicts stability in the 3T3 cell line. On the other hand, the distance between the branchpoint and the 3' splice site is comparable between the two groups (*SI Appendix, Fig. S3C*). Similarly, whereas stable lariats have a somewhat higher GC content, the average predicted minimum free energy is indistinguishable between the two groups (*SI Appendix, Fig. S3D and E*).

In summary, we found that 3T3 cells produce stable circular intronic RNAs. These sisRNAs are characterized by (i) their short length, (ii) their C branchpoint, and (iii) their molecular ratio compared with their cognate mRNA.

SisRNAs in Different Vertebrate Species. To determine whether sisRNAs are found in other vertebrate species, we extracted RNA from control and α -amanitin-treated HeLa cells, DF1 cells (a chicken cell line), and XTC cells (a *Xenopus laevis* cell line). For all samples, RNase R was used to enrich for circular RNA. We detected sisRNAs in all three species (Fig. 2A).

In HeLa cells, 318 sisRNAs were derived from 270 genes; 46 of these genes also hosted an sisRNA in mouse 3T3 cells (*Dataset S1*), and 14 sisRNA-encoding introns were orthologous. Similarities between human and mouse sisRNAs included the length of the sisRNAs (Fig. 2B) and the molecular ratio between the sisRNA and its cognate mRNA (Fig. 2D). Among these, most were stabilized as lariats with a bias for a C branchpoint (Fig. 2C), and a fifth (14) were common to human RBCs. In a previous study, Zhang et al. (15) analyzed 20 abundant lariats and determined that the branchpoint in untreated HeLa cells could be either an A or a C nucleotide with about equal frequency. We confirmed 16 out of the 20 and mapped the branchpoint of 13. Among

these 13 lariats, 8 lariats persisted after transcription inhibition, all of which had a C branchpoint. Conversely, we found an A branchpoint in all five lariats that were detected in untreated cells only (*SI Appendix, Fig. S2*). These data suggest a common pathway for stability of human and mouse lariats.

Similar observations were made on chicken DF1 cells. We detected 277 circular sisRNAs derived from short introns of 214 genes (*Dataset S1*). Twenty-one of these sisRNAs had corresponding inverted reads, and most were stabilized as lariats with a cytosine branchpoint (71%), unlike unstable DF1 lariats with predominantly A branchpoints (65%). Among the chicken sisRNA host genes, 10 were the same as in HeLa cells and 15 were the same as in mouse 3T3 cells (including two orthologous introns). Overall six genes were host to at least one sisRNA in all three species. The six common genes have very different functions, including transcription regulation, motors, ribosome biogenesis, fatty acid synthesis, and protein localization.

In *X. laevis* XTC cells, we found a somewhat different pattern—about 4,000 circular sisRNAs derived from 2,000 genes (*Dataset S1*). In the most extreme case, one gene hosted 17 sisRNAs. We are sure of intronic RNA stability, because many unstable transcripts are no longer detectable after transcription is inhibited for 12 h (*SI Appendix, Fig. S4*). A total of 570 stable intronic RNAs had inverted reads, 98% of which were lariats. These lariats were barely detectable in RNase R minus samples, and most molecules had an A branchpoint and an AU-rich body, unlike those in mouse, human, and chicken. The sisRNAs from XTC cells resembled those from frog oocytes: thousands of circular sisRNAs, among which nearly 90% of mapped branchpoints were adenine (Fig. 2C and D). Among the 4,000 circular sisRNAs, 20% were found in the cytoplasm of the *X. laevis* oocyte.

To study stable lariat RNAs from a fish, we extracted RNA from freshly expressed eggs of the zebrafish (*Danio rerio*). These eggs contain a large pool of stable RNAs used during early embryonic development. We treated egg RNA with RNase R to enrich for circular molecules but found only 13 examples among more than 10,000 spliced transcripts. Fig. 2A shows reads for the

tnikb gene and the stable lariat derived from its intron. This sisRNA, which has a C branchpoint and an AU-rich body, is the most abundant zebrafish sisRNA in our sample. Unfortunately, we could not map the junction of other sisRNAs because of their low abundance. An important limitation of our zebrafish data is the dependence on genome annotation for sisRNA discovery. We anticipate that more sisRNAs will be detected when a more complete genome assembly becomes available.

In summary, we found hundreds of stable lariat intronic RNAs in six vertebrate species—two mammals, two amphibians, one bird, and one fish. These stable lariats derive from short introns and are characterized by a C branchpoint (except in *Xenopus*). They come from coding and noncoding genes related to nearly all known cellular functions.

In Situ Hybridization Detects Intronic RNAs in the Cytoplasm. Definitive evidence that lariats are stored in the cytoplasm of vertebrate cells (other than RBCs) comes from single molecule in situ hybridization. We chose probes for three highly abundant sisRNAs, one each from human, mouse, and chicken (*SI Appendix, Fig. S5*). The human and mouse sisRNAs were known to be abundant in RBCs as well as in cultured HeLa and 3T3 cells. The chicken probe was selected as the most abundant sisRNA in cultured DF1 cells. Because all of these introns are less than 250 nucleotides in length (*SI Appendix, Fig. S5*), we used probes consisting of three to four short oligonucleotides, a multistep signal amplification series, and detection by Fast Red (BaseScope probes from Advanced Cell Diagnostics). Positive (exonic *PpiB* gene) and negative (bacterial *dapB* gene) controls were included. In each case, we detected weak but clear positive signals in the cytoplasm corresponding to the intronic probe (Fig. 3). Because the tissues had not been treated with α -amanitin to inhibit ongoing transcription, we also saw positive signals in the nucleus that probably correspond to nascent transcripts at the sites of transcription and to unstable spliced introns.

Although our molecular analysis of lariat RNAs began with examination of RBCs, we did not carry out confirming in situ hybridization experiments on RBCs. In a recently published study on 7SL RNA in amphibian and mammalian blood cells (16), we encountered unexpected and still unexplained resistance of these cells to standard in situ hybridization protocols. The simplest explanation is that conventional in situ hybridization probes do not penetrate the intact RBC membrane. But until these technical issues are overcome, we are unable to provide confirming in situ hybridization data for lariat RNAs in RBCs.

SisRNAs in Mouse Liver and Brain. It was convenient to begin our study of sisRNAs with cell lines, which provide pure populations of cells in which transcription is easily inhibited. To examine whether stable lariats also occur in somatic tissues, we collected total RNA from mouse liver and brain. As before, we used RNase R to enrich for circular RNAs before sequencing the samples. We found ~3,500 different circular intronic RNAs in the liver and ~1,000 in the brain. Fig. 4A shows the most abundant circular RNA in these tissues, derived from an intron of *Pex6*.

Unlike RNA from α -amanitin-treated cultured cells, which is enriched for metabolically stable molecules, total RNA derived from whole tissues is a mixture of short lived and more stable molecules. We analyzed intronic RNAs derived from short introns (<1,500 nt) and determined the branchpoint usage. In both the liver and brain, roughly half of the mapped branchpoints were cytosine, suggesting the existence of stable lariats in these tissues (Fig. 4B). Next, we examined intronic RNAs that were expressed at 10% or more of their mRNA level (as seen for most sisRNAs detected in cell culture). We found 946 and 242 such intronic RNAs in the liver and brain, respectively. These pools were enriched for C-branched lariats, consistent with their being stable lariats (Fig. 4B). Taken together, the data suggest that we

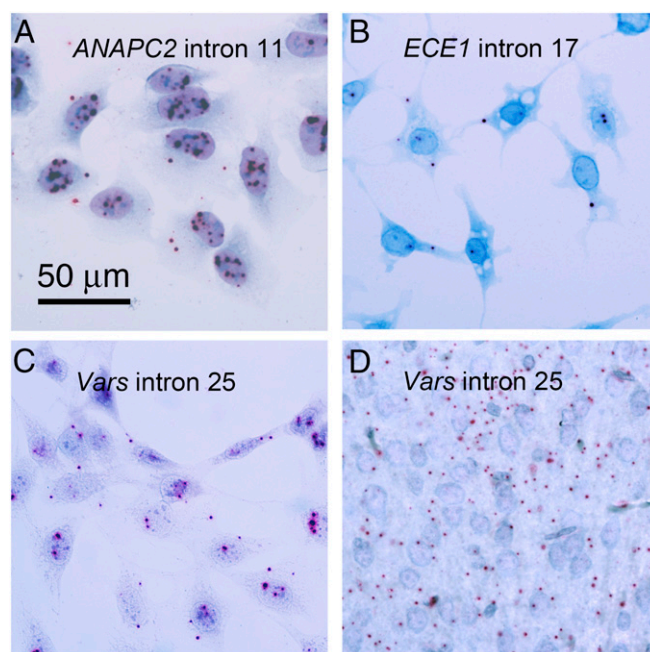


Fig. 3. Single-molecule in situ hybridization of introns. (A) Human *ANAPC2* intron 11 in cultured human cells (HeLa). (B) Chicken *ECE1* intron 17 in cultured chicken cells (DF1). (C) Mouse *Vars* intron 25 in cultured mouse cells (3T3). (D) Mouse *Vars* intron 25 in section of mouse hind brain. Tissues were fixed in 4% paraformaldehyde and hybridized according to the protocol supplied by the probe manufacturer (Advanced Cell Diagnostics).

detect both stable and unstable lariats, but importantly some of the lariats have the characteristics of those detected in α -amanitin-treated cells: They are short, they have a C branchpoint, and they are expressed at 10% or more of their cognate mRNA level. These data provide a suggestion that sisRNAs exist in liver and brain.

To explore these presumed sisRNAs further, we divided the lariats from liver and brain into two categories: those with a C branchpoint and those with A, G, or T. Most of the C-branched lariats (72% from the liver and 75% from the brain) were also detected in α -amanitin-treated 3T3 cells, strongly suggesting the presence of sisRNA in liver and brain. Conversely, many of the presumably unstable lariats (those with an A, G, or T branchpoint) were detected in untreated cells only (Fig. 4C). To confirm the presence of sisRNA in the brain, we carried out in situ hybridization on conventional formaldehyde-fixed tissue sections using a probe against intron 25 of the mouse *Vars* gene. Positive signals were evident in both the cerebrum and cerebellum (Fig. 3D).

Cytoplasmic SisRNAs Are Not Associated with Their Cognate mRNAs.

A possible function for cytoplasmic sisRNAs would be to regulate translation or localization of their cognate mRNAs. To explore this possibility, we tested whether sisRNAs might be physically associated with their cognate mRNAs in the cytoplasm. We carried out in situ hybridization for the products of two genes in the cytoplasm of *Xenopus* oocytes: *uggt1* (UDP-glucose glycoprotein glucosyltransferase 1) and *pphln1* (periphilin 1). These two genes each produce a very long sisRNA, making it possible to design probes consisting of multiple fluorescent oligonucleotides for maximal signal intensity. Probes were labeled with Quasar 570 (mRNA) or Quasar 670 (intron) and hybridized simultaneously. Very small transparent oocytes were chosen so that the probes could be visualized in whole mounts. In situ hybridization of probes against *uggt1* mRNA and its longest intron (Fig. 5) show essentially no colocalization. Similar results were obtained for *pphln1* mRNA and its longest intron. Although the

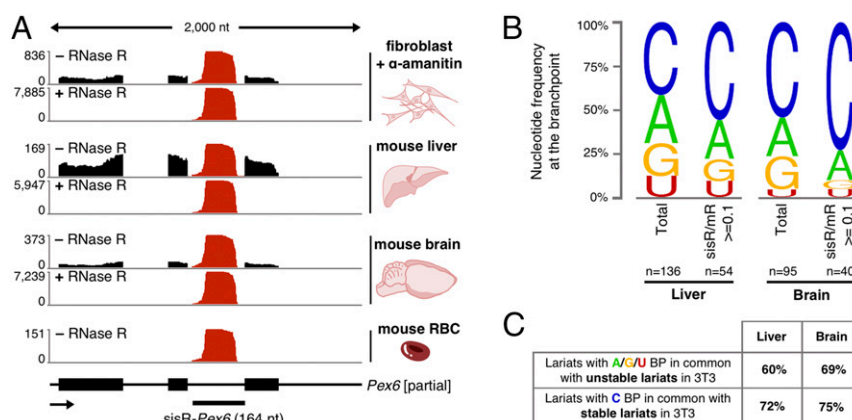


Fig. 4. SisRNAs in various mouse tissues. (A) A very abundant sisRNA, sisR-Pex6 (red), is detected in fibroblasts, liver, brain, and RBCs. (B) Branchpoint nucleotide frequency of all lariats detected in liver and brain. "n" represents the number of mapped branchpoints. (C, Upper) Lariats with A, G, and U branchpoints in common with lariats detected in control 3T3 cells and (Lower) lariats with C branchpoint in common with lariats detected in α -amanitin-treated 3T3 cells.

mRNA and cognate sisRNA from each of these two genes might colocalize transiently in the cytoplasm, they do not form a permanent association detectable by in situ hybridization.

SisRNAs Are Exported from the Nucleus by NXF1/NXT1. Because the oocyte nuclear envelope remains intact until the very end of oogenesis, it seemed probable that sisRNAs reach the cytoplasm by an active transport mechanism. To test this possibility, we injected purified sisRNA or pre-mRNA into the oocyte nucleus. However, these RNAs were rapidly degraded, suggesting that sisRNAs are normally stabilized at or near the time of their transcription. Hence, to study the export mechanism, we expressed sisRNAs from DNA constructs that we injected into the nucleus. After 48 h, we measured the amount of expressed sisRNAs that remained in the nucleus and the amount exported to the cytoplasm. Such experiments are possible because the oocyte stores large amounts of the proteins required for RNA export (14). Also, because export is a saturable process, one can perform export competition assays by injecting an RNA of interest along with large amounts of competitor RNA. After a few minutes, the oocytes can be dissected and cytoplasmic accumulation can be determined (17, 18).

In our system, we expressed ectopic sisRNAs independently of their endogenous cognate mRNA. *X. tropicalis* introns were inserted into a partial *X. laevis* *ncl* gene between exons 14 and 15, a combination that shows high splicing efficiency. The partial *ncl* gene was placed downstream of an mCherry gene (SI Appendix, Fig. S6A). Once injected into the nucleus, the DNA construct was transcribed, giving rise to both *X. tropicalis* sisRNA and mCherry mRNA (SI Appendix, Fig. S6C). By selecting for red oocytes, we could monitor the efficiency of injection, transcription, and translation (SI Appendix, Fig. S6B). Because intronic sequences are highly divergent between *X. tropicalis* and *X. laevis*, we could easily distinguish ectopic from endogenous sisRNAs.

Two days after injection, we examined the amount of sisRNA that had accumulated in the cytoplasm (Fig. 6A). For two abundant *X. tropicalis* cytoplasmic sisRNAs, sisR-*faf2* and sisR-*trappc9*, 50% of the newly transcribed lariats were detected in the cytoplasm. On the other hand, for sisR-*act1* and sisR-*ift43*, sisRNAs found primarily in the nucleus, only 5% of the ectopic transcripts accumulated in the cytoplasm. To determine which export system is used by sisRNA, we coinjected three constructs whose transcripts should compete for different export pathways: an mRNA (GFP), a *Xenopus* tRNA, and a short hairpin RNA (against a mouse transcript not conserved in *Xenopus*) (Fig. 6B). The export of sisRNA was inhibited only by overexpression of GFP mRNA, suggesting that mRNA and sisRNA share a common export pathway. Importantly, the fact that sisRNA export

could be inhibited in this way strongly implies that sisRNA export is not due to simple leakage from the nucleus.

Most mRNAs are exported via NXF1/NXT1 (19). To test whether NXF1 is sufficient for sisRNA export, we coinjected a sisR-*faf2* DNA construct along with increasing concentrations of GFP DNA constructs with or without *nxf1* DNA. As expected, increasing amounts of GFP mRNA led to a decrease in sisRNA

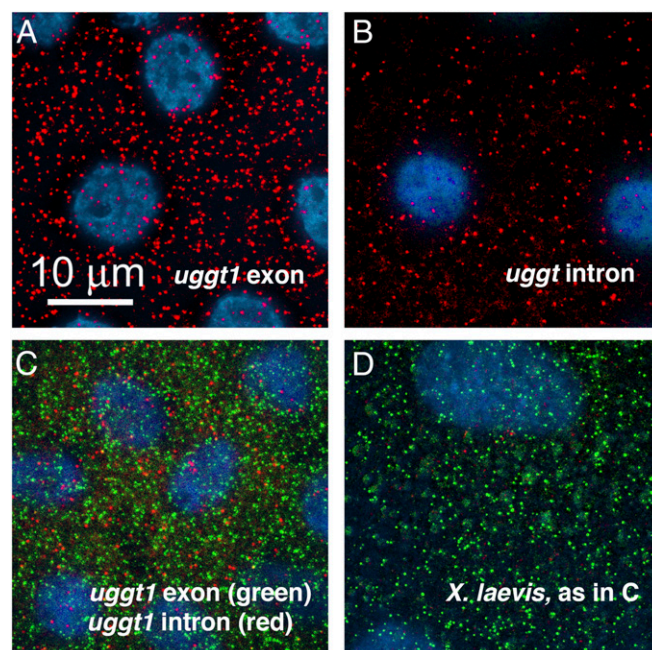


Fig. 5. Single-molecule in situ hybridization of exon and intron probes from the *X. tropicalis* *ugg1*-1 gene. Oocytes (200–400 μ m diameter) were hybridized and examined as whole mounts. Images are deconvolved stacks extending from the follicle cells (blue nuclei, DAPI stain) surrounding the oocyte into the peripheral cytoplasm of the oocyte. All of the signal is in the oocyte itself. (A) *X. tropicalis* oocyte hybridized against the *ugg1*-1 exon probe (red). (B) *X. tropicalis* oocyte hybridized against the *ugg1*-1 intron probe (red). (C) *X. tropicalis* oocyte hybridized against both the exon probe (green) and the intron probe (red). The two probes give nonoverlapping signals. (D) *X. laevis* oocyte hybridized with both probes, as in C. Only the green exonic probe gives a positive signal in this species. Oocytes were fixed in 4% paraformaldehyde and hybridized according to the protocol supplied by the probe manufacturer (Stellaris probes from LGC Biosearch Technologies).

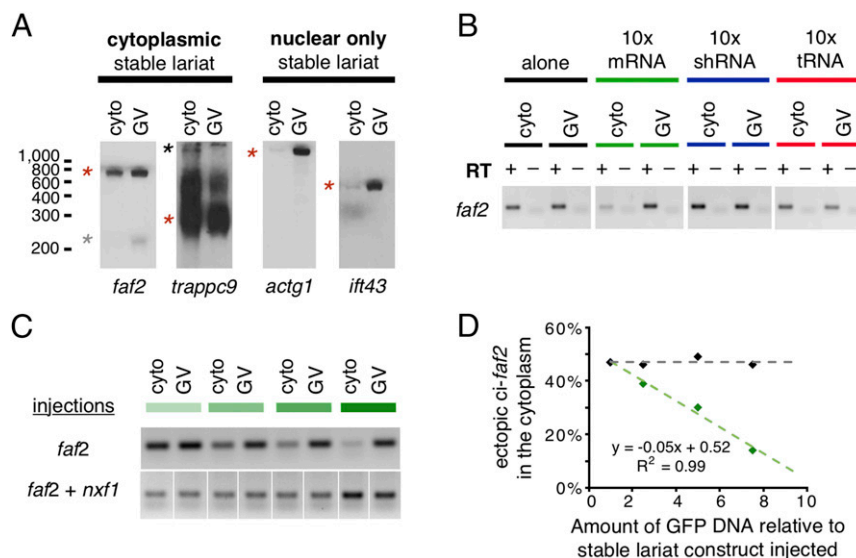


Fig. 6. (A) Northern blots against two cytoplasmic and two nuclear *X. tropicalis* sisRNAs expressed ectopically in *X. laevis*. Red stars mark circular transcripts. The gray star marks the linearized *sisR-faf2*, and the black star marks the unprocessed intronic RNA *trappc9*. (B) RT-PCR against *sisR-faf2* on cytoplasmic and nuclear fractions. Oocytes were injected with a *sisR-faf2* intron construct alone (black) or with GFP DNA (mRNA, green), an shRNA construct (blue), and a tRNA construct (red). All 16 lanes were run on the same gel. After imaging, the lanes were cut (digitally) and rearranged to give the final panel. (C) Upper lanes: increasing amounts of GFP DNA were injected into the nucleus along with the *sisR-faf2* intron construct. The amount of cytoplasmic *sisR-faf2* decreases as more GFP DNA is injected. Lower lanes: coinjection of *nx1* DNA along with the *sisR-faf2* intron construct rescues the cytoplasmic accumulation of *sisR-faf2*. The samples were run on nonadjacent lanes on one gel. (D) Semi-quantitative analysis of the RT-PCR shown in C.

export (Fig. 6 C and D, green data points). On the other hand, when NXF1 was overexpressed, the export of sisRNA was restored, presumably because there was now an excess of the correct export system (Fig. 6 C and D, black data points). From these experiments, we conclude that sisRNAs are selectively exported to the cytoplasm using the NXF1/NXT1 machinery.

Discussion

During or shortly after transcription, introns are generally spliced out of new transcripts in the form of lariats (circles with a tail). They are then attacked by the nuclear enzyme DBR1 (1), which hydrolyzes the 2'-5' phosphodiester bond (branchpoint) to give rise to a linear molecule that is degraded within minutes in the nucleus. A variety of exceptions to this canonical pathway have been described. If the excised intron is not debranched, it can remain as a circular molecule in the nucleus. Zhang et al. (15) described numerous circular intronic long noncoding RNAs (ciRNAs) in cultured human cells and suggested that these RNAs have a general role as regulators of their parent coding genes.

Previously, we described stable circular introns from oocytes and early cleavage stages of the frog *Xenopus*. Stable introns from thousands of genes were found in the cytoplasm of both immature and mature oocytes. They persist during the maturation stages of the oocyte and are transmitted to the early embryo. Hence, they were given the name sisRNA. Molecular analysis showed that these sisRNAs in the cytoplasm are circular in form.

The studies reported here were undertaken to determine whether cytoplasmic sisRNAs are limited to oocytes and early embryos or exist in other organism and other cell types. We began by examining stable transcripts in mouse and human RBCs. Mature RBCs lack nuclei and hence provide an ideal source of pure cytoplasm. Circular sisRNAs were readily detected, ~500 in the mouse and 150 in human RBCs. To extend our search, we examined cultured cells of four species—human, mouse, chicken, and frog. The cells were treated for 12 h with α -amanitin, an inhibitor of transcription, to allow time for unstable introns to be degraded. In each case, we found numerous transcripts that were resistant to the inhibitor. The cytoplasmic

localization of a few transcripts was demonstrated by single-molecule in situ hybridization. From these studies, we feel confident that cytoplasmic sisRNAs are widespread in normal cells of a variety of vertebrate species.

Cytoplasmic sisRNA from human, mouse, and chicken cells share several general characteristics. They are usually derived from a single short intron per gene with a strong bias for a C branchpoint instead of the canonical A. It is known that the debranching enzyme DBR1 is relatively inefficient in linearizing C-branched lariats. We speculate that a delay in linearization of short, C-branched introns may contribute to the formation of a stable complex that is more resistant to debranching. However, about 10% of circular sisRNAs in mammals are A-branched, and not all C-branched lariats are stabilized. Therefore, other features, such as folding of the lariat, must also be involved in the failure to linearize.

The final step in stabilizing the sisRNA is its export to the cytoplasm, where there is no DBR1 enzyme. Export does not occur by simple diffusion; instead, sisRNAs are actively transported to the cytoplasm by the NXF1/NXT1 export pathway. In summary, C-branched sisRNAs may avoid debranching by a combination of resistance to the DBR1 enzyme and dedicated export to the cytoplasm, where there is no enzyme.

Xenopus sisRNAs, which have a preponderance of A branchpoints in both cultured cells and oocytes, are 10 times less abundant on average than mammalian sisRNAs, perhaps reflecting their sensitivity to DBR1. Nevertheless, there are 4,000 different sisRNAs in *Xenopus* XTC cells, compared with only 500 in 3T3 cells and 400 in HeLa cells. *Xenopus* sisRNAs, like those of other species, escape debranching in part by export to the cytoplasm by way of the NXF1/NXT1 pathway.

Two other chemically distinct sisRNA species have been observed in the cytoplasm. The first occurs in *Drosophila*, where Pek et al. (20) detected a linear sisRNA, sisR-1, in the cytoplasm of the early embryo. SisR-1 is processed from the fourth intron of the *rga* gene, presumably after splicing and linearization of the full-length intronic RNA. The sisR-1 export pathway remains unknown. The second example involves an unexpected source of stable introns derived from tRNA transcripts. Intronic RNAs are

spliced out of the pretransfer RNA by an endonuclease and released as linear molecules. Instead of being targeted for degradation by Xrn1, some molecules are circularized and stabilized in the cytoplasm (21).

Postsplicing circularization has also been observed for pol II intronic transcripts, including the stable ones reported here. Intronic RNAs are presumably linearized and can then be either recircularized without further trimming (5'ss to 3'ss circle) (10) or trimmed before recircularization (22). The stability or localization of these circular molecules has not yet been examined.

We considered the hypothesis that sisRNAs are derived from and somehow regulate a specific class of mRNAs in the cytoplasm. Using the admittedly imprecise gene ontology categories, we did not find that sisRNAs are derived from any specific class of genes. It seems unlikely, therefore, that sisRNAs regulate mRNAs in a specific metabolic pathway or pathways.

One simple hypothesis is that each cytoplasmic sisRNA interacts directly with and regulates the mRNA derived from the same gene. To look for such physical association between a sisRNA and its cognate mRNA, we carried out single-molecule FISH on two mRNAs and the sisRNAs from the same genes (*uggt1* and *pphln1*). We found no tendency for the sisRNA to be associated physically with its cognate mRNA. These observations rule out a permanent association between the mRNA and its cognate sisRNA but leave open the possibility of more transient interactions.

Clues about the function of cytoplasmic sisRNA may be extrapolated from studies of intronic lariat RNAs that accumulate in cells with a minimal level of the DBR1 enzyme. These induced or artificial sisRNAs are associated with RNA-binding proteins (RBPs)—dicer, TDP-43, and small nuclear ribonucleoproteins (snRNPs)—and they contribute to an abnormal depletion of RBPs (23–25). Therefore, it is possible that cytoplasmic sisRNAs play a role in the regulation of RBPs.

Whatever may be the function(s) of cytoplasmic sisRNA in normal somatic tissues, they cannot be involved in transcription or translation in RBCs, which lack these functions entirely. Cytoplasmic sisRNAs in RBCs must have some as-yet undetermined function, or they may simply represent stable molecules that will be discarded when the RBCs themselves are later degraded.

In summary, our studies document the existence of stable intronic RNAs in the cytoplasm of somatic tissues from five vertebrate species (human, mouse, chicken, fish, and frog). These intronic RNAs are circular (lariat) in form and are characterized by a predominance of an unusual C branchpoint (except in the frog). A specific function for these lariat intronic RNAs remains to be determined.

Materials and Methods

Cell Culture. Human HeLa and mouse 3T3 cells were cultured at 37 °C in DMEM (Dulbecco's modified Eagle's medium; Gibco) with 10% FBS, L-alanyl-L-glutamine (Glutamax; Gibco), and antibiotics. Chicken DF1 cells were cultured at 39 °C in the same DMEM. *X. laevis* XTC cells were cultured in 66% Leibovitz's L-15 medium (Gibco) with 10% FBS, Glutamax, and antibiotics at room temperature.

Tissue Collection. Mouse tissues were provided by Safia Malki from the Bortvin laboratory, Department of Embryology, Carnegie Institution for Science, Baltimore, MD. Blood was collected from recently euthanized CD-1 or C57BL/6 animals by cardiac puncture with an EDTA-coated 1 mL syringe and a 22-gauge needle. Brain and liver were dissected from the same animals and immediately processed for RNA extraction. Human RBCs were purchased from the Interstate Blood Bank (through Zen-Bio, Inc.). Before shipment, the sample tested negative for hepatitis B, hepatitis C, HIV, and syphilis. The samples were stored between 2 °C and 8 °C in anticoagulant citrate dextrose solution. Eggs from wild-type zebrafish were provided by the Halpern laboratory, Department of Embryology, Carnegie Institution for Science, Baltimore, MD.

RBC Purification. Blood samples were centrifuged at 300 g for 10 min at room temperature. The supernatant was removed, and the pelleted cells were

resuspended in 1× PBS. The suspension was layered on a 50–80% Percoll gradient and centrifuged at 1,000 g for 30 min at room temperature. RBCs settled near the bottom of the tube and were collected free of white blood cells. Before usage, RBCs were washed three times in 1× PBS.

RNA Purification. RNA was extracted with TRIzol reagent (Ambion) and purified with the Direct-zol kit (Zymo Research). DNase treatment was performed on the Direct-zol column. RNA was quantitated with a Nanodrop instrument (Thermo Fisher Scientific) and its quality assessed on a Bioanalyzer 2100 electrophoresis instrument (Agilent).

RNAse R. RNA was denatured 5 min at 72 °C in 8 μL of RNase R 1× buffer. Then, 1 μL of RNase R (Epicentre) was added, and the reaction was incubated overnight at 37 °C. RNA degradation was prevented with RNasin (Promega).

Library Preparation, Sequencing, and Sequence Analysis. Libraries were prepared using TrueSeq stranded total RNA sample preparation (Illumina). Sequencing was performed on an Illumina HiSeq. 2000 sequencer with either 100 bp or 150 bp single-end reads. Reads were aligned with TopHat (v2.0.7) to the mouse genome (v10), the human genome (v19), the chicken genome (v5), or the *X. laevis* genome (v9.0). Intronic reads were quantified using Bedtools (v2.15.0), and lariat reads were obtained using lariat_find.pl (10).

RT-PCR Analysis. To confirm the bioinformatic prediction, we performed RT-PCR analysis according to a previously published protocol (4). In short, cDNA was made using Episcript (Epicentre) according to the manufacturer's protocol and amplified with Taq polymerase (Qiagen) and various primers. In vitro transcribed RNA was used as a control to demonstrate the efficiency of the primers used.

Northern Blots. Up to 5 μg of RNA was separated on an 8% polyacrylamide–8 M urea gel and transferred onto a nylon membrane (Zeta Probe GT; Bio-Rad). RNA was probed with dsDNA labeled with digoxigenin (DIG)-dUTP in hybridization buffer and detected using an anti-DIG antibody conjugated with alkaline phosphatase and CDP-Star chemiluminescent substrate (Roche).

Single-Molecule in Situ Hybridization. The probes consisted of “ZZ” oligonucleotide pairs from Advanced Cell Diagnostics, which are complementary to the following gene regions:

Mus musculus Vars intron 25, bases TGAGGATCTTGTCGGTGGCCATCGGTGACAGTGCCT GTGCCACACAGTCTGTGGCTCCCTCCACTCAGTGTCAAGACTAAAGCTGCTGCCCTGAGATCCTCTTTAGTCAGGGACAGACAGGTGAAGTTGATGATTGATGGGAAGTCTCTGCAGCTTGGGGCC

Gallus gallus ECE1-intron17, bases GCCATCCATGCCTTCTCCATCCACGATGGTTTTGGGAGAGCATGACAGCAAGCCGAGGTGTCCAGAGATGCCCCAGAGAATGAGATCAGGAGGTCTGGTTAAAGAGCGAGCCAGGCTTGGCGAGAGTGGGTCTGCAGCTGCGGGCATCA

Homo sapiens ANAPC2-intron11, bases AGCTGGCCTGCCCGGCCACAGCGCCAGCTGTGGGTGAGGGGTGTCCCCAGTGAGGGTGTGGTTGGGGCAGCTGCGCCAGAGCGCGCACCTTCCCT and TGACGGGAGGAGGGGAGGTGCAGCCCGTGCCCA

The ZZ oligonucleotide pairs are attached to a proprietary amplification target, which in turn is detected by a proprietary set of oligonucleotides. The latter are finally detected by Fast Red dye. Each red “dot” in the final preparation presumably represents a single RNA molecule.

Xenopus Oocytes and Injections. *X. laevis* females were anesthetized with 1% MS222 (tricaine methane sulfonate, pH 7), and pieces of ovary were surgically removed and cultured in OR2 medium (26) at 16°. Manually separated oocytes were microinjected at the animal pole with 100 pg of plasmid in a 2.4 nL volume using the Nanoject II microinjection apparatus (Drummond Scientific). Needles were pulled from capillary tubes (0.5-mm inner diameter, 1.2-mm outer diameter) with a horizontal micropipette puller (model P-97; Sutter Instrument). Different amounts of competitor plasmid were coinjected: 100, 250, 500, 750, and up to 1,000 pg. Injection of more than 1,000 pg of plasmids was often lethal for oocytes. At 48 h after injection, oocytes were dissected in OR2 isolation solution (83 mM KCl, 17 mM NaCl, 6.0 mM Na₂HPO₄, 4.0 mM KH₂PO₄, 1 mM MgCl₂, 1.0 mM DTT, pH 7.8). Nuclei and cytoplasm were collected on dry ice before adding TRIzol for RNA extraction.

Plasmids. A construct that expresses sisRNAs was generated as follows: The mCherry-CAAX gene, kindly provided by the Halpern laboratory, Department of Embryology, Carnegie Institution for Science, Baltimore, MD, was subcloned into the pGEM-T plasmid. Multicloning sites were used to add a minimal CMV promoter and a partial *X. laevis* gene (exon 14–15 of *ncf*) in which the intron was replaced by intron 2 of the *X. tropicalis* *faf2* gene.

A plasmid that encoded an HA-tagged (N terminus) human NXF1 gene was purchased from ABclonal. The pCS2 GFP plasmid was generated in the Brown laboratory, Department of Embryology, Carnegie Institution for Science, Baltimore, MD. An *X. tropicalis* tRNA gene was amplified by PCR and cloned in the pGEM-T plasmid. An shRNA construct was kindly provided by the Zhao

Zhang laboratory, Department of Embryology, Carnegie Institution for Science, Baltimore, MD.

Data Deposition. RNA sequencing data have been deposited in the NCBI Sequence Read Archive (SRA BioProject ID: PRJNA479418).

ACKNOWLEDGMENTS. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award R01 GM33397 (to J.G.G.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. J.G.G. is an American Cancer Society Professor of Developmental Genetics.

- Ooi SL, et al. (2001) RNA lariat debranching enzyme. *Methods Enzymol* 342:233–248.
- Ooi SL, Samarsky DA, Fournier MJ, Boeke JD (1998) Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: Intron length effects and activity of a precursor snoRNA. *RNA* 4:1096–1110.
- Gardner EJ, Nizami ZF, Talbot CC, Jr, Gall JG (2012) Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev* 26:2550–2559.
- Talhouarne GJ, Gall JG (2014) Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *RNA* 20:1476–1487.
- Bertles JF, Beck WS (1962) Biochemical aspects of reticulocyte maturation. I. Fate of the ribonucleic acid. *J Biol Chem* 237:3770–3777.
- Doss JF, et al. (2015) A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes. *BMC Genomics* 16:952.
- Alhasan AA, et al. (2016) Circular RNA enrichment in platelets is a signature of transcriptome degradation. *Blood* 127:e1–e11.
- Qian Z, et al. (2018) Potential diagnostic power of blood circular RNA expression in active pulmonary tuberculosis. *EBioMedicine* 27:18–26.
- Sedgwick AD, Morris T, Russell BA, Lees P (1986) Single step purification procedure for the rapid separation of equine leucocytes. *Vet Res Commun* 10:445–452.
- Taggart AJ, et al. (2017) Large-scale analysis of branchpoint usage across species and cell lines. *Genome Res* 27:639–649.
- An X, et al. (2014) Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* 123:3466–3477.
- Gall JG, Wu Z (2010) Examining the contents of isolated *Xenopus* germinal vesicles. *Methods* 51:45–51.
- Nam K, et al. (1994) Yeast lariat debranching enzyme. Substrate and sequence specificity. *J Biol Chem* 269:20613–20621.
- Panté N (2006) Use of intact *Xenopus* oocytes in nucleocytoplasmic transport studies. *Methods Mol Biol* 322:301–314.
- Zhang Y, et al. (2013) Circular intronic long noncoding RNAs. *Mol Cell* 51:792–806.
- Talhouarne GJS, Gall JG (2018) 7SL RNA in vertebrate red blood cells. *RNA* 24:908–914.
- Dargemont C, Kühn LC (1992) Export of mRNA from microinjected nuclei of *Xenopus laevis* oocytes. *J Cell Biol* 118:1–9.
- Jarmolowski A, Boelens WC, Izaurralde E, Mattaj JW (1994) Nuclear export of different classes of RNA is mediated by specific factors. *J Cell Biol* 124:627–635.
- Delaleau M, Borden KL (2015) Multiple export mechanisms for mRNAs. *Cells* 4:452–473.
- Pek JW, Osman I, Tay ML, Zheng RT (2015) Stable intronic sequence RNAs have possible regulatory roles in *Drosophila melanogaster*. *J Cell Biol* 211:243–251.
- Lu Z, et al. (2015) Metazoan tRNA introns generate stable circular RNAs in vivo. *RNA* 21:1554–1565.
- Panda AC, et al. (2017) High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic Acids Res* 45:e116.
- Armakola M, et al. (2012) Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat Genet* 44:1302–1309.
- Li Z, et al. (2016) Intron lariat RNA inhibits MicroRNA biogenesis by sequestering the dicing complex in *Arabidopsis*. *PLoS Genet* 12:e1006422.
- Han B, et al. (2017) Human DBR1 modulates the recycling of snRNPs to affect alternative RNA splicing and contributes to the suppression of cancer development. *Oncogene* 36:5382–5391.
- Wallace RA, Jared DW, Dumont JN, Segal MW (1973) Protein incorporation by isolated amphibian oocytes. III. Optimum incubation conditions. *J Exp Zool* 184:321–333.