Trends in Biochemical Sciences

CellPress REVIEWS

Review

Stable Intronic Sequence RNAs (sisRNAs): An Expanding Universe

Seow Neng Chan[1] and Jun Wei Pek[1,2,*]

Intronic sequences are often regarded as 'nonsense' transcripts that are rapidly degraded. We highlight here recent studies on intronic sequences that play regulatory roles as long noncoding RNAs (lncRNAs) which are classified as sisRNAs. Interestingly, sisRNAs come in different forms and are produced via a variety of ways. They regulate genes at the DNA, RNA, and protein levels, and frequently engage in autoregulatory feedback loops to ensure cellular homeostasis under normal and stress conditions. Future directions, evolutionary insights, and potential implications of dysregulated sisRNAs are also discussed, especially in relation to human pathogenesis.

Highlights

A revised definition of sisRNAs is provided that unifies recently discovered intronic sequence-containing lncRNAs.

sisRNAs can be produced through splicing-dependent/independent pathways, and interact with satellite bodies for their decay.

A conserved sisRNA–target base-pairing regulatory mechanism is proposed.

sisRNAs are involved in stem cell homeostasis via regulatory feedback loops.

A New Member: sisRNAs

Recent years have witnessed the discovery and characterization of abundant **noncoding RNAs** (ncRNAs; see Glossary) in different biological systems. Introns are noncoding nucleotide sequences that are interspersed between the coding sequences (also known as exons) [1]. During the production of mature mRNA in eukaryotic cells, the precursor mRNA (consisting of both introns and exons) undergoes splicing to remove introns, and exons are assembled together for translation. Thus, intronic RNAs are commonly regarded as 'junk' sequences by researchers. Furthermore, they are rapidly degraded, except for some exceptional cases such as small Cajal body associated RNAs (scaRNAs), small nucleolar RNAs (snoRNAs), and microRNAs (miRNAs), which are involved in various processes in the cell [2,3]. Thus, many regulatory intronic lncRNAs were often neglected or left undetected in the past.

In 2012, a high-throughput RNA sequencing study in the *Xenopus tropicalis* oocyte nucleus revealed intronic sequences that are abundant and unusually stable – for at least 1–2 days [4]. Further experiments concluded that these intronic sequences are maternally deposited and are derived from the coding strand, where a few intronic sequences may be formed from a single long intron. They constitute a new class of ncRNAs known as **stable intronic sequence RNAs** (sisRNAs) [4,5]. Following that observation, sisRNAs have also been found in other organisms such as viruses [6,7], yeast [8], *Arabidopsis* [9,10], *Drosophila* [11–14], and mouse [15]. These sisRNAs were reported to be involved in regulation of gene expression via various mechanisms.

Recent studies also report different forms of sisRNAs that are generated through a variety of mechanisms (Table 1). In this review we provide a revised definition of sisRNAs to include the expanding pool of intronic sequence RNAs that have been identified over the past few years. We also discuss the biogenesis, molecular mechanisms, and biological functions of sisRNAs, and conclude with future research directions and the potential implication of sisRNAs in human health and disease.

[1]Temasek Life Sciences Laboratory, National University of Singapore, 1 Research Link, 117604, Singapore
[2]Department of Biological Sciences, National University of Singapore, 14 Science Drive, 117543, Singapore

*Correspondence:
junwei@tll.org.sg (J.W. Pek).

© 2018 Elsevier Ltd. All rights reserved.

Trends in Biochemical Sciences

CellPress REVIEWS

A Revised Definition of sisRNAs

Generally, sisRNAs are defined as intronic sequences that are excised during splicing, and which are not degraded immediately in the cell. Lately, emerging studies have reported the presence of intronic sequences that are **polyadenylated** (may also contain a 5′ cap) [16] or the retention of introns in a truncated transcript (containing both exonic and intronic sequences) [17–21]. These transcripts are also unexpectedly stable and do not enter the canonical degradation pathways such as **nonsense-mediated decay** (NMD). Although the mechanisms leading to the production of these intronic sequence-containing transcripts are still unclear, some have been shown to be functional. We therefore suggest a broader definition of sisRNAs that accommodates these newly discovered ncRNAs. The revised definition of sisRNAs is as follows: the intronic sequences (i) are derived from the coding strands of their cognate host genes, (ii) are not rapidly degraded, and (iii) may contain exonic sequences, 5′ caps, and/or polyA tails.

Biogenesis and Decay of sisRNAs

The biogenesis of sisRNAs can be generally divided into two pathways, namely splicing-dependent and splicing-independent mechanisms. Nevertheless, the pathways used for sisRNAs biogenesis are not mutually exclusive because a sisRNA can be produced using either pathway [16]. This idea is similar to the production of other ncRNAs such as snoRNAs and miRNAs where different biogenesis pathways can be utilized. Generally, sisRNAs can exist in different structures (linear or circular), localize at different compartments (nucleus or cytoplasm), or undergo different downstream processing depending on their function, the type of tissue, and developmental stage. Although the decay pathway of sisRNAs has not been well studied, it was shown that a double-stranded RNA (dsRNA)-binding protein is involved in the degradation of sisRNAs in *Drosophila* [13]. Interestingly, because this dsRNA-binding protein also localizes to nuclear bodies, it suggests an involvement of biomolecular condensation in the metabolism of sisRNAs.

Splicing-Dependent Mechanism

The canonical splicing-dependent mechanism can be considered as the earliest proposed mechanism for the biogenesis of sisRNAs (Figure 1A). During pre-mRNA splicing, the intron is spliced out in the form of a lariat (circular intronic RNA with 3′ tail) [1,2]. The lariat is further processed by the **lariat debranching enzyme** or undergoes 3′-end trimming to form a linear or circular sisRNA, respectively [6,11,12,22–24]. Recircularization of linear sisRNAs can also result in the formation of circular sisRNAs after debranching [25–27].

Alternatively, some sisRNAs are formed through an exon **back-splicing** mechanism whereby the 5′ end of the exon preceding the intron is ligated to the 3′ end of the succeeding exon. This process of circularization is highly promoted by the presence of complementary elements (such as *Alu* repeats) flanking the exons [28] (Figure 1A). It produces a class of circular sisRNAs (*circEIF3J* and *circPAIP2*) that contain both exons and retained intronic sequences (termed **exon–intron circular RNAs**, ElciRNAs) in mammalian cells. It was suggested that the biogenesis of sisRNAs by back-splicing occurs at the site of transcription, and that this process competes with the production of linear mRNA during splicing of the single pre-mRNA.

Splicing-Independent Mechanism

For splicing-independent mechanisms, a sisRNA can be transcribed directly from an intronic promoter or from a cognate host gene promoter, and does not require canonical or back-splicing of the pre-mRNA. Intronic sequences can be fully retained in **intron-retained (IR) transcripts** or partially retained via intronic cleavage and polyadenylation.

Glossary

Back-splicing: a process in which the 5′ end of one or more exons is joined to the 3′ of the succeeding exon to form a circular structure.
Biomarker: a quantifiable biological parameter, including biomolecules such as nucleic acids and proteins, which can be used as indicators of disease risk or for the diagnosis of health conditions.
Circular intronic RNA (ciRNA): a type of circularized RNA consisting of an intronic sequence which escapes debranching.
***Cis* natural antisense transcript (*cis*-NAT):** an RNA that is transcribed from the antisense strand of a coding gene.
Enhancer: a short DNA region that can be bound by an activator to promote the transcription of a gene.
Exon–intron circular RNA (ElciRNA): a type of circularized RNA that consists of a retained intron between exons.
Feedback loop: a type of mechanism whereby the output response of a reaction subsequently returns to influence the input signal to modulate the continued response.
Germline stem cell (GSC): an adult stem cell that can self-renew and differentiate to produce a daughter cell that can differentiate to produce gametes.
Homeostasis: maintenance of a stable internal state at equilibrium by regulatory processes despite changing environmental or external conditions.
Intron-retained (IR) transcripts: stably expressed and fully spliced transcripts that contain intronic sequences.
Lariat debranching enzyme: a protein involved in the debranching of intronic lariats by catalyzing the hydrolysis of 2′,5′-phosphodiester bond at the branchpoint.
Non-coding RNA (ncRNA): a transcript that does not code for protein but may possess enzymatic, structural, and regulatory functions.
Nonsense-mediated decay (NMD): an RNA degradation pathway in the cytoplasm that is activated by the presence of premature translation-termination codon in the mRNA.
Nuclear body: a membraneless dynamic organelle structure found in

An example of sisRNAs produced via direct transcription from intronic promoters is the biogenesis of polyadenylated maternally deposited sisRNAs in *Drosophila* (such as *sisR-1*, *sisR-2*, and *sisR-3*) [16]. These sisRNAs are transcribed from the 5′ ends of intronic promoters and are cleaved at the polyadenylation signals (PAS) at the 3′ ends of the cognate introns (Figure 1B). These sisRNAs are then polyadenylated by Wispy, a cytoplasmic polyA polymerase (PAP), which stabilizes the sisRNAs.

In human and mouse cell lines, sisRNAs are manifested as IR transcripts that are transcribed from cognate host gene promoters, and the full-length unspliced intronic sequences are retained in the mature polyadenylated transcripts (Figure 1B). These transcripts with 'detained' introns (DIs) are stored in the nucleus and are unusually stable (do not undergo NMD) [19].

Studies of sisRNAs from the Epstein–Barr virus (EBV) [6,29], such as *ebv-sisRNA-1* and *ebv-sisRNA-2*, suggest that sisRNAs can also be independently transcribed to form unspliced transcripts of different sizes with or without polyA tails. Proteins that bind to *ebv-sisRNA-1* and *ebv-sisRNA-2* also bind to unspliced RNAs containing intact splice-junctions, such as *W1-ebv-sisRNA-1-W2-ebv-sisRNA-2*, suggesting that some sisRNAs may be transcribed from the promoters of cognate host genes and are terminated at the PAS sequences, resulting in the production of stable unspliced sisRNAs [18].

Finally, not all sisRNAs transcribed from cognate host gene promoters contain fully retained introns because premature termination at the 3′ ends of intronic sequences can also generate sisRNAs (Figure 1B). Upon UV irradiation or during an immune response, widespread early termination and polyadenylation at introns occur, producing sisRNAs that localize to both the nucleus and cytoplasm. Both the inclusion of alternative last exons and slower transcription rates are suggested to activate the polyA sites on these shorter RNA isoforms [17,20,21].

sisRNA Decay and Satellite Bodies

Regulation and degradation of a plethora of cellular RNAs expressed is crucial, and involves various RNA degradation/surveillance pathways [30]. However, the mechanism for sisRNA regulation is still not well understood. Because sisRNAs escape the canonical degradation pathway during splicing and exhibit unusual stability, it is perhaps not surprising that other yet-to-be-discovered mechanisms are utilized for their regulation. In *Drosophila*, Disco-interacting protein 1 (DIP1), a dsRNA-binding protein, was found to be involved in regulating the abundance of *sisR-1*. DIP1 binds to and negatively regulates *sisR-1* in a post-transcriptional manner (Figure 2A). Interestingly, DIP1 forms a **nuclear body** (also known as a **satellite body**) that localizes around the fourth chromosomes (Figure 2B) [13]. The *Drosophila* fourth chromosome is highly enriched in *INE-1* transposable elements inserted in the introns. During transcription, sisRNAs that contain *INE-1* sequences (called *INE-1* sisRNAs) are transcribed from various genes, making the fourth chromosome a prominent locus for *INE-1* sisRNAs. These *INE-1* sisRNAs localize in close proximity to the fourth chromosome and bind to DIP1, forming microscopically visible satellite bodies. Satellite bodies are distinctly different from other known nuclear bodies such as Cajal bodies and histone locus bodies in *Drosophila*. Thus, a high local concentration of sisRNAs can impact on the organization of nuclear architecture.

Molecular Functions of sisRNAs

Recent studies suggest that sisRNAs play important roles in regulating biochemical processes in the cell. They can either repress or enhance gene expression through **feedback loops** or affect splicing events by acting as **protein decoys**.

the nucleus and which is involved in the regulation of numerous nuclear functions.

Satellite body: an uncharacterized membraneless subnuclear organelle defined by the presence of DIP1 that is localized around the fourth chromosome in the nucleus, and which may be involved in the metabolism of *INE-1* sisRNAs.

Polyadenylation: addition of a polyadenylic acid (polyA) tail to the 3′ end of a transcript (such as mRNA) following the polyadenylation signal (PAS).

Polycomb repressive complex 2 (PRC2): a multisubunit protein complex that catalyzes the methylation of chromosomal histone H3, leading to epigenetic repression.

Protein decoy: competitive RNAs that have structures similar to the actual target of a protein, and are able to sequester the protein from its targets.

Stable intronic sequence RNA (sisRNA): a stable noncoding intronic RNA (may contain exons) that is transcribed from the coding strand and may play regulatory roles in the cell.

Synthetic biology: a field in biology that combines engineering in building novel biological compounds, structures, functions, and systems.

Table 1. List of sisRNAs and Their Known Functions

Organism/source	Name	Type of sisRNA	Location	Functions	Refs
Arabidopsis thaliana	*AG-incRNA4*	Linear intronic sequence	Second intron of the *AGAMOUS* gene	Represses the expression of *AGAMOUS* through the PRC2 complex by associating with *CURLY LEAF*	[9]
	lariat41	Intronic lariat	N/a[a]	Regulates the level of miRNA through sequestration of the dicing complex, DCL1/HYL1	[10]
	lariat42	Intronic lariat	N/a		
Drosophila melanogaster	*sisR-1*	Linear intronic sequence	Second intron of the *rga* gene	Regulates GSC occupancy by modulating *rga* expression in a negative feedback loop through repression of the antisense transcript of *rga* (*ASTR*)	[11]
	sisR-2	Linear intronic sequence	Third intron of the *mbt* gene	Regulates GSC maintenance in a negative feedback loop with *bantam* by repressing the expression of dFAR1	[31]
	sisR-3	Linear intronic sequence	Fourth intron of the *csp* gene	Represses lncRNA *CR44148* which is essential for larval development	[16]
	sisR-4	Intronic lariat	Second intron of the *dpn* gene	Promotes the expression of *deadpan* in a positive feedback loop by activating an enhancer during embryonic development	[12]
Epstein–Barr virus	*ebv-sisRNA-1*	Linear intronic sequence	Internal repeat 1 region of EBV	Interacts with human regulatory proteins. May be involved in the pathogenesis of EBV	[7,18]
	ebv-sisRNA-2	Long hairpin intronic sequence	Internal repeat 1 region of EBV		
Human (MRC5-VA cell line)	*ASCC3* sisRNA	Short ncRNA with alternative last exon	N/a	Restores the expression of *ASCC3* long mRNA after UV irradiation by interacting with ASCC3 protein complex	[21]
Human (H9 cell line)	*ci-ankrd52*	Circular intronic RNA	Second intron of the *ANKRD52* gene	Promotes the expression of the *ANKRD52* gene in a positive feedback manner by associating with RNA Pol II.	[23]
Human (HeLa and Hek293 cell lines)	circEIF3J	Exon–intron circRNAs	Third intron of the *EIF3J* gene	Enhances transcription of parental gene in *cis* in a positive feedback loop by interacting with U1 snRNP complexes and RNA Pol II. May have other functions in *trans*	[28]
	circPAIP2	Exon–intron circRNAs	Second intron of the *PAIP2* gene		
Human (MCF7 cell line)	*Eleanors*	Linear intronic sequence	Intergenic region around the *ESR1* gene locus (*upstream-Eleanor*)	*Upstream-Eleanor* enhances the expression of *ESR1* and *Eleanor* RNAs by associating with RNA Pol II, and are involved in resistance adaptations of breast cancer cells	[39]
Mouse (RAW264.7 cell line)	*lnc-Lsm3b*	Intron-retained transcript	Exon 1, exon 2, and a fragment of intron 2 of the *Lsm3* gene	Maintains homeostasis of the immune response by sequestering the RIG-I viral sensor, leading to downregulation of type 1 IFN at the end of the immune response	[17]
Human (HCT-116 cell line)	SMYD3-CLIP	Linear intronic sequence	First intron of the *SMYD3* gene	Represses host gene transcription by associating with EZH2 of the PRC2 complex	[32]

[a]N/a, not available.

Gene Repression

Several linear sisRNAs are known to downregulate gene expression, as demonstrated for *sisR-1*, *sisR-2*, and *sisR-3* which repress *ASTR*, *dFAR1* (fatty-acyl-CoA reductase), and the lncRNA *CR44148* respectively [11,13,16,31]. Interestingly, these linear sisRNAs, including *ebv-sisRNA-1*, share similar predicted secondary structures where they each contain an exposed 3′ tail (Figure 3). It is hypothesized that the exposed 3′ tail plays a crucial role as a guide sequence to locate its target through complementary base-pairing, subsequently leading to target degradation. In the case of *sisR-2* and *sisR-3*, it is hypothesized that they facilitate the recruitment of RNA decay machineries to their targets [16,31].

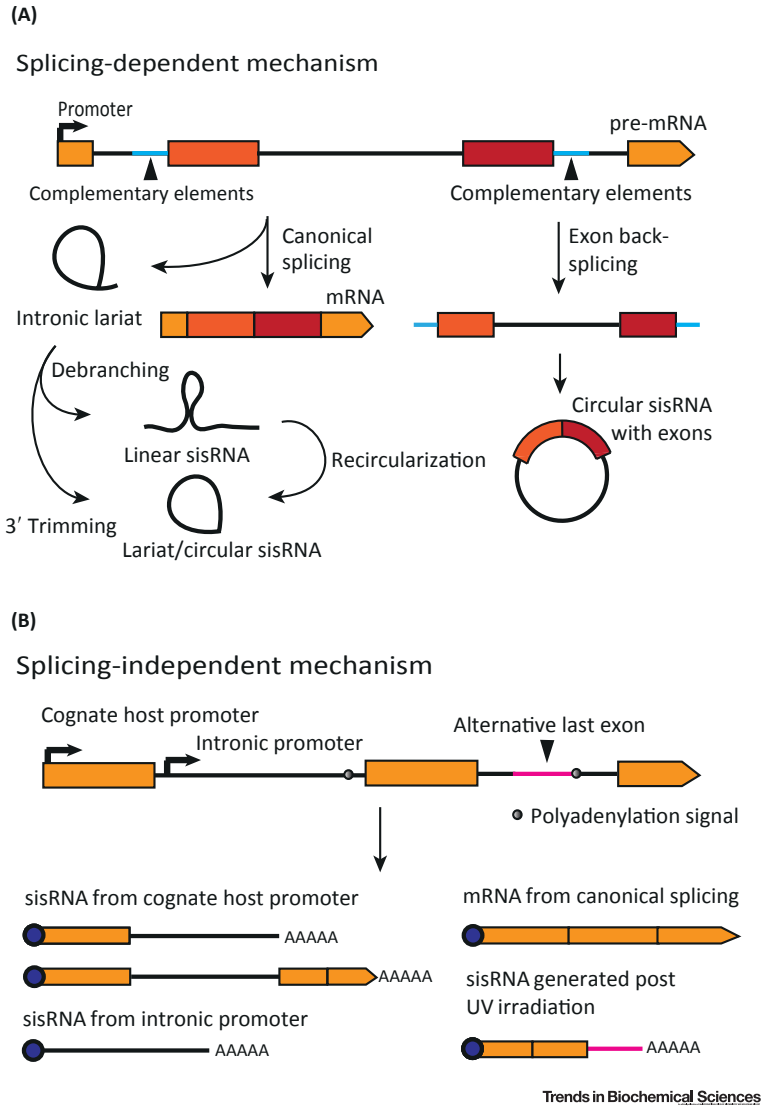

Figure 1. Biogenesis of Stable Intronic Sequence RNAs (sisRNAs). (A) Splicing-dependent mechanism. Intronic lariats derived from canonical pre-mRNA splicing are debranched to form linear sisRNAs or undergo 3′ end trimming to form circular sisRNAs. Some debranched linear sisRNAs can also recircularize to form circular sisRNAs. For pre-mRNAs with complementary repeats flanking the exons, exon back-splicing may occur to generate circular sisRNAs containing both exonic and intronic sequences. (B) Splicing-independent mechanism. sisRNAs are transcribed from promoters found in cognate host genes or introns, and terminate at the polyadenylation signal site(s). This generates 5′-capped mature polyadenylated sisRNA transcripts and sisRNAs with exonic sequences (intron-retention transcripts). Post UV irradiation, sisRNAs are generated by widespread polyadenylation and utilization of alternative last exons, which are otherwise spliced out during canonical splicing.

In a human colorectal cancer cell line (HCT-116), sisRNAs were shown to guide the **Polycomb repressive complex 2** (PRC2) by associating with EZH2 (a core component of PRC2) [32]. These sisRNAs possess EZH2 binding sites, and their binding to EZH2 leads to epigenetic silencing of endogenous cognate genes. Overexpression of intronic fragments of *SMYD3* was able to downregulate its host gene in *trans*. These gene repression effects were partially

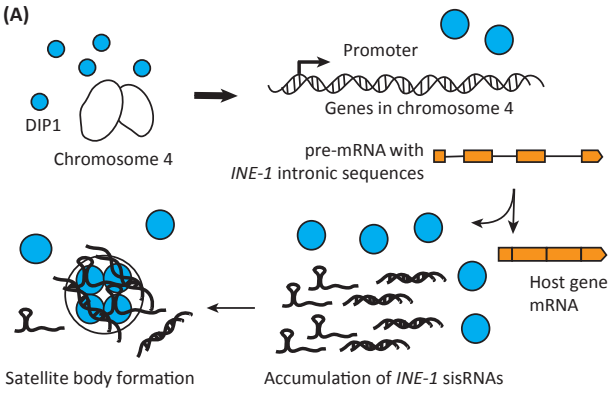

Figure 2. Formation of Satellite Bodies by DIP1. (A) A working model for the regulation of the abundance of *INE-1* stable intronic sequence RNAs (sisRNAs) by DIP1 through the formation of satellite bodies. (B) Confocal image of satellite bodies (green) formed by DIP1 at DAPI-dense (magenta) regions of the fourth chromosome, as indicated by the white arrows. Scale bar, 10 μM. Abbreviation: DAPI, 4′,6-diamidino-2-phenylindole.

diminished when EZH2 expression was knocked down, or when the EZH2 binding sites were mutated. A similar mechanism occurs in *Arabidopsis* where *AG-incRNA4* sisRNA (derived from the second intron of the *AGAMOUS* gene, *AG*) is suggested to repress the expression of its host gene via PRC2 by associating with *CURLY LEAF* (*CLF*) [9].

Based on the well-studied *sisR-1* from *Drosophila*, some insights involving its molecular mechanism in gene repression have been provided [11,13]. In the *sisR-1* model, it was shown that *sisR-1* downregulates the expression of its host gene *rga* in a negative feedback loop by repressing the expression of its target, termed *ASTR* (Figure 4A). *ASTR* is a lncRNA or ***cis* natural antisense transcript** (*cis*-NAT) that is known to positively regulate *rga* pre-mRNA. A high level of *ASTR* in 2–14 h embryos promotes the expression of *rga.* This subsequently leads

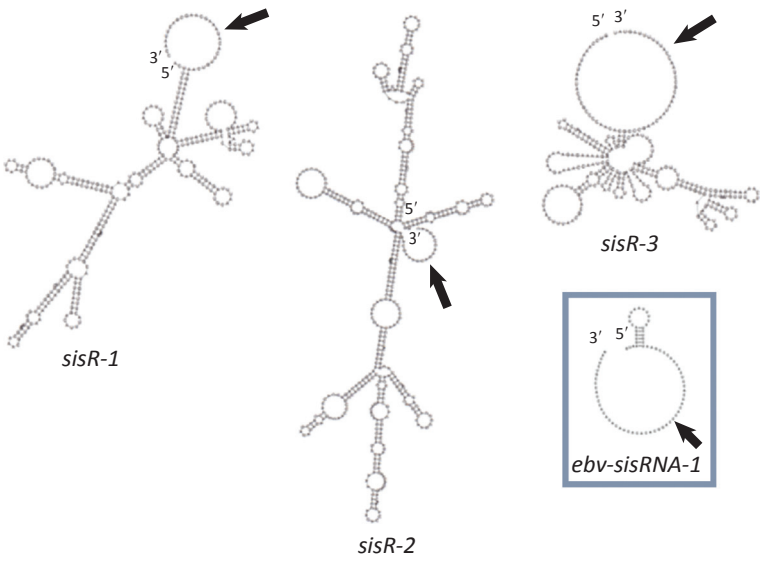

Figure 3. Predicted Structures of Linear Stable Intronic Sequence RNAs (sisRNAs). Linear sisRNAs are predicted to share a conserved secondary structure with the 3′ tails being exposed (as indicated by black arrows).

to the accumulation of the splicing byproduct, *sisR-1*, which eventually represses the expression of *ASTR* in 14–24 h embryos via a negative feedback loop. Given that the 3′ end of *ASTR* shows 76 nt complementary base-pairing with the 3′ end of *sisR-1*, a molecular mechanism was proposed. *sisR-1* assembles into a stable ribonucleoprotein complex and the 29 nt exposed 3′ tail searches for the binding site in *ASTR*. Following binding to the exposed 3′ tail, it promotes local unwinding of *sisR-1* and subsequently forms a more stable 76 nt duplex with *ASTR*, which is targeted by dsRNA nucleases such as Dicer-2 for decay [13].

Gene Activation

Several sisRNAs such as **circular intronic RNAs** (ciRNAs) and ElciRNAs in human, and *sisR-4* in *Drosophila*, were reported to enhance the transcription of their host genes [12,23,28] (Figure 4B). Studies in human cell lines and breast cancer cells have shown that these sisRNAs localize near their parental gene loci; however, the exact molecular mechanism leading to enhanced transcription remains to be determined.

In human embryonic stem cells (H9), *ci-ankrd52* was found to enhance the transcriptional activity of its parental gene, *ANKRD52* [23]. Depletion of *ci-ankrd52* led to significantly reduced transcription of *ankrd52*. Importantly, alternative mechanisms for *ci-ankrd52*-mediated regulation of gene expression have been ruled out, such as functioning as a miRNA sponge or leading to alternative splicing that can lead to NMD. Consistent with this hypothesis, *ci-ankrd52* accumulates at the transcription start-site of the *ANKRD52* gene locus and colocalizes with the RNA polymerase II (Pol II) complex, presumably to enhance the transcriptional rate of the *ANKRD52* gene [23]. A similar gene activation role was proposed for ElciRNAs, such as in *cirEIF3J* and *circPAIP2*, that were shown to link with U1 snRNP through RNA–RNA interactions before associating with RNA Pol II complexes at the promoter regions of host genes [28]. In *Drosophila*, a maternally deposited sisRNA *sisR-4* upregulates transcription of its host gene *deadpan* during early embryonic development by activating an **enhancer** in its intronic region.

This leads to the production of *sisR-4*, which encourages a higher transcriptional activity of the *deadpan* locus, forming a positive feedback loop [12].

Short polyadenylated RNA isoform sisRNAs (such as *ASCC3)*, which are generated after UV-induced DNA damage, were also found to be involved in regulation of its long mRNA isoform [21]. Under normal conditions, pre-mRNA transcriptional elongation processes occur at a fast rate. After UV irradiation there is an immediate reduction in transcriptional activity, and elongation progresses in a much slower manner, with majority of the RNA Pol II activity locating only at promoter-proximal sites (~25 kb). This results in a shift of the *ASCC3* protein-coding long isoform to a much shorter *ASCC3* sisRNA. The *ASCC3* long mRNA isoform is involved in the suppression of transcription, and *ASCC3* sisRNA is hypothesized to repress the activity of the ASCC protein complex. *ASCC3* sisRNA is primarily chromatin-associated in the nucleus, and it functions through the binding of its 2.8 kb 3′ untranslated region (UTR). This negative feedback loop mediated by *ASCC3* sisRNA ensures that transcription of the long *ASCC3* mRNA isoform resumes 24 h after UV irradiation.

Protein Decoy

In addition to binding to RNA and DNA, sisRNAs can also function by binding to proteins. sisRNAs regulate the activity of RNA-binding proteins by binding to them and sequestering them from their normal functions (Figure 4C). Such regulation can have profound effects on gene expression because RNA-binding proteins often engage in widespread functions in RNA processing.

In *Arabidopsis*, sisRNAs (specifically lariat RNAs) are able to act as protein decoys to regulate the **homeostasis** of miRNA concentration [10]. It was proposed that lariat sisRNAs function as a layer of regulation in the biogenesis and maintenance of miRNA homeostasis by sequestering the dicing complex. At steady-state, small amounts of lariat RNAs, which have 'escaped' debranching by dbr1, serve as molecular sponges and compete with pre-miRNAs to bind to the dicing complex (DCL1/HYL1). It is believed that this mechanism regulates processing of pre-miRNAs and maintains the level of miRNAs. Mechanistically, the dicing complex 'mistakenly' binds to the GU/UG motif found at the 5′ splice sites of lariat sisRNAs, leading to sequestration of their function.

In the mouse macrophage cell line RAW264.7, sisRNA *lnc-Lsm3b* is involved in homeostasis of the immune response [17]. During infection, foreign viral RNAs in the cytoplasm are detected by the innate viral RNA sensor, retinoic acid-inducible gene-I (RIG-I). Activated RIG-I subsequently elicits an immune response involving upregulation of type 1 interferons (IFNs), which lead to further upregulation of the expression of RIG-I monomers. During this process *lnc-Lsm3b* sisRNA is also produced. Towards the late stage of the immune response, *lnc-Lsm3b* sisRNA functions in a negative feedback manner by sequestering excess RIG-I monomers, and as a consequence downregulates the expression of type 1 IFNs. *lnc-Lsm3b* (self-RNA) sisRNA binds to RIG-I via its GA-rich sequences and structural motifs at the long stem structure, and competes with viral RNA (non-self). During early infection, binding of viral dsRNA to the C-terminal domain of RIG-I results in an open structure, which activates the innate response. Later on, upon binding to *lnc-Lsm3b*, it restricts the conformation change of RIG-I, and hence prevents continued production of type 1 IFNs, reverting the immune response to normal.

The *lnc-Lsm3b* sisRNA is an example of an extensive class of sisRNAs that are produced during the immune response [20]. These sisRNAs, which are formed by intronic cleavage and poly-adenylation, have been predicted to bind to RNA-binding proteins such as FUS, ELAVL1, and PUM2, supporting a role as protein decoys. In yeast, cytoplasmic circular sisRNAs have been shown to sequester TDP-43, suggesting a conserved role for sisRNAs as protein decoys [33].

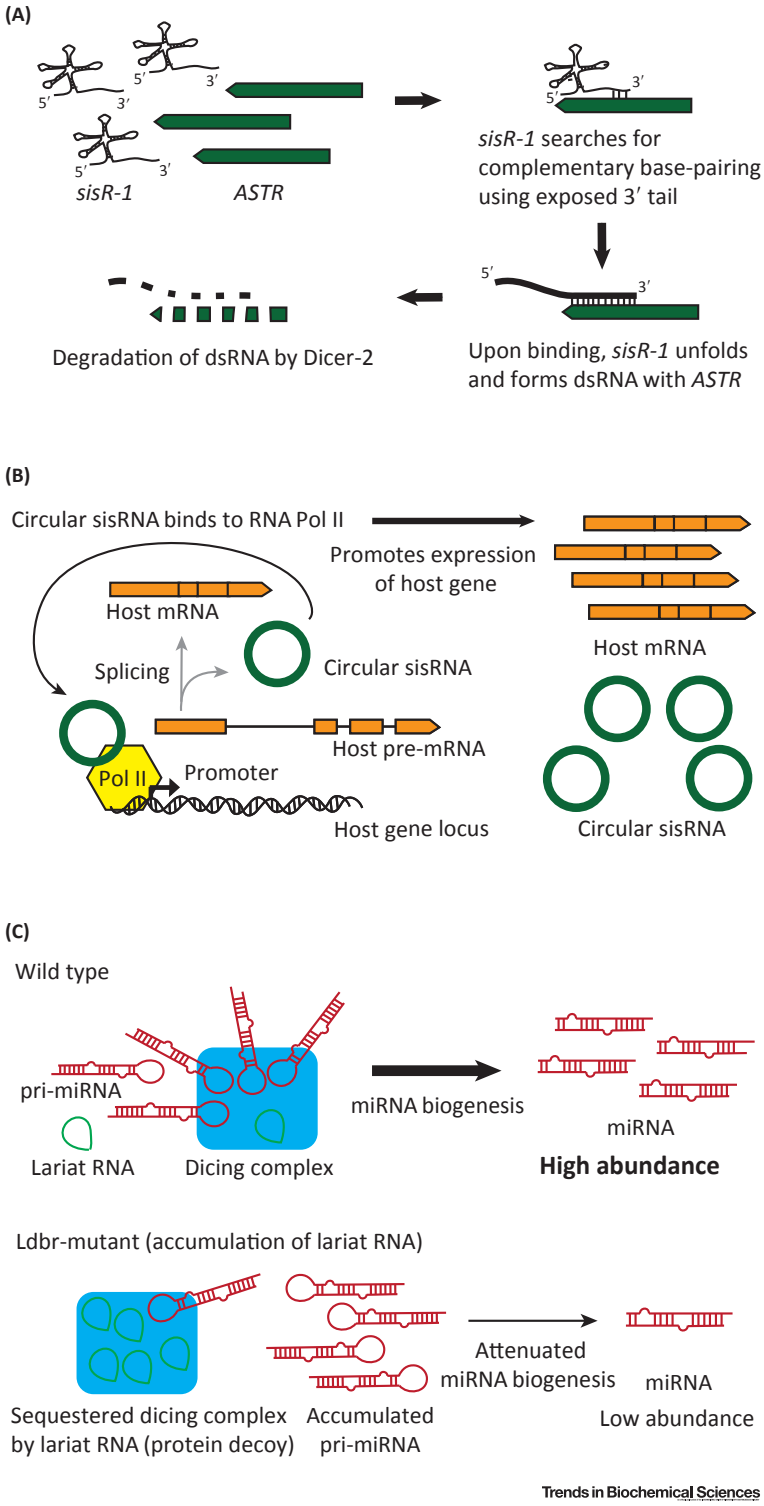

Figure 4. Gene Regulation by Stable Intronic Sequence RNAs (sisRNAs). (A) Gene repression: proposed mechanism for how *sisR-1* leads to the degradation of its target *ASTR*. (B) Gene activation: a working mechanism for how a circular sisRNA activates the expression of its own parental gene by associating with the RNA Pol II complex at the

(See figure legend on the bottom of the next page.)

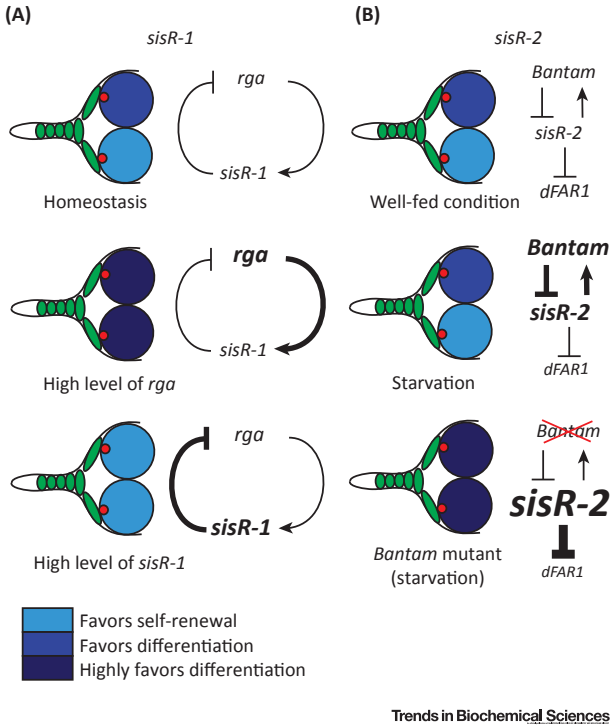

Figure 5. Germline Stem Cell (GSC) Maintenance by *sisR-1* and *sisR-2*. Illustrations of how (A) *sisR-1* and (B) *sisR-2* feedback loops regulate the fate of (GSCs; self-renewal vs. differentiation) under normal and stress conditions (starvation).

Biological Functions of sisRNAs

Because sisRNAs mainly function to fine-tune gene expression, it is perhaps not surprising to discover that they are involved in the maintenance of cellular homeostasis and developmental robustness in normal and stress conditions [34].

In *Drosophila*, *sisR-1* promotes **germline stem cell** (GSC) niche occupancy by repressing the expression of *rga* gene via *ASTR* in a negative feedback loop (Figure 5A). In ovarian GSCs, *rga* functions as a differentiation-promoting factor [13,35]. In the niche, GSCs exhibit heterogeneity in *rga* expression (high in one GSC and low in an adjacent GSC), presumably owing to oscillatory gene expression [35]. Although the mechanism on the coordination of heterogeneous *rga* expression in different GSCs is still unknown, it is proposed that a *rga-sisR-1* feedback loop may be involved in balancing GSC differentiation and self-renewal [13,35]. In general, cellular heterogeneity has been linked to oscillatory gene expression [35], which can be a consequence of negative feedback [36]. Consistent with previous studies, introns or sisRNAs may serve as ideal mediators in relaying such local feedback [34].

Another *Drosophila* sisRNA *sisR-2* functions in the maintenance of GSCs during starvation (Figure 5B). *sisR-2* negatively regulates GSC maintenance by repressing the expression of *dFAR1* [31]. *dFAR1* is responsible for the maintenance of GSCs, and knockdown of *sisR-2*

promoter region. (C) Protein decoy: working model for how sisRNAs act as protein decoys in *Arabidopsis*. The accumulated lariat sisRNAs sequester the dicing complex from pri-miRNAs and attenuate miRNA biogenesis. Abbreviations: ds, double-stranded; ldbr, lariat debranching enzyme; Pol II, polymerase II.

leads to increased dFAR1 expression and a higher number of GSCs. Importantly, *sisR-2* engages with *bantam* miRNA via a negative feedback loop: *sisR-2* promotes transcription of *bantam*, while *bantam* negatively regulates activity of *sisR-2*. This *sisR-2*/*bantam* feedback loop is crucial because expression of *sisR-2* is upregulated during starvation. However, the activity of *sisR-2* is counteracted by concomitant upregulation of *bantam*, providing a buffer to prevent GSC loss. This biological mechanism is probably important in the wild because flies experience changing environments with varying amounts of food. They need to safeguard a stable pool of GSCs for the generation of eggs when a food source becomes available.

Future Directions

Introns Are a Direct Readout of Transcription

The above description suggests that sisRNAs play a conserved role in regulatory feedback loops in plants and animals. Although they fall into common themes in regulating their parental genes, there are also variations regarding the underlying molecular mechanisms. Feedback loops can be mediated by miRNAs or nascent RNAs to ensure robustness of gene expression [37,38]. What selective advantages do sisRNAs have that make them suitable for relaying feedback loops? An important consideration is that introns are cotranscribed with exons. While exons are being spliced together to form a single mRNA, intronic RNAs are released as discrete molecules. The cell can therefore make use of sisRNAs as a direct readout of transcriptional activity of a particular locus. How this readout is interpreted will depend on their interacting partners and on the type of feedback loops that the sisRNAs engage in. In cases where pre-mRNAs are alternatively spliced in specific cell types, an increase in protein diversity may be also accompanied by an increase in sisRNA diversity. Therefore, methods have been developed to enrich for low abundant sisRNAs. Furthermore, it will be useful to study the effects of sisRNAs on gene expression at the cellular level by using more sensitive single-molecule fluorescent *in situ* hybridization and single-cell RNA sequencing technologies (Figures 6 and 7 and Boxes 1–3).

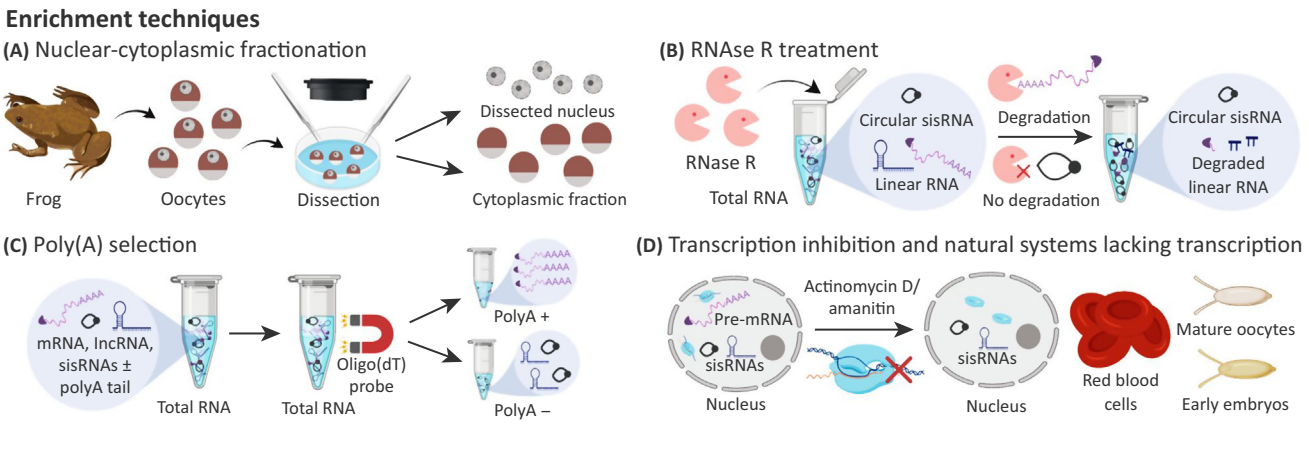

Figure 6. Methods to Identify sisRNAs. (A) Nuclear-cytoplasmic fractionation. Nuclear sisRNAs, which are often in lower abundance, can be enriched by separation from the cytoplasmic RNAs (noise) as demonstrated using frog oocytes. (B) RNase R treatment. RNase R degrades all linear RNA and enriches for circular sisRNAs and intronic lariats. (C) Poly(A) selection. sisRNAs are separated and enriched for improved detection based on the presence of poly(A) tail. (D) Transcription inhibition and natural systems lacking transcription. Transcription inhibition treatment separates sisRNAs from pre-mRNA and enriches the stable sisRNAs. Similarly, red blood cells that lack nuclear fraction, transcriptionally quiescent mature oocytes, and early embryos serve as natural systems for sisRNA detection.

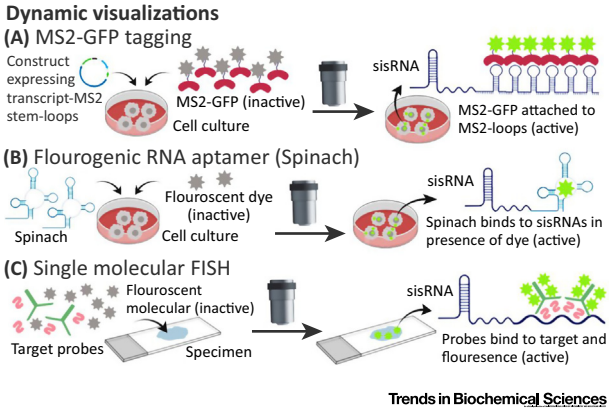

Figure 7. Methods to Characterize sisRNAs. (A) MS2-GFP tagging and (B) Flourogenic RNA aptamer (Spinach). Dynamic interactions of sisRNAs in live cell culture can be characterized by using exogenously expressed sisRNAs tagged with MS2 stem loop or tagging of endogenous sisRNAs with Spinach and visualized by the addition of MS2-GFP proteins or fluorescent dyes. (C) Single molecular FISH. A high accuracy detection and quantification of sisRNAs in a single cell by using specific target probes and visualization using fluorescent molecules.

Close Proximity to the Site of Transcription

The use of sisRNAs for host gene regulation is also ideal because they are generated near to their sites of action. This scenario provides a high concentration of sisRNAs at the parental locus for efficient activity, and at the same time minimizes potential non-specific actions at other genomic regions. This theme can be enhanced by the possibility that sisRNAs provide sequence-specific regulation by forming complementary base-pairing with DNA or *cis*-NATs transcribed locally. Because *cis*-NATs have been shown to regulate the parental gene in a myriad of ways, involving transcription interference, alternative splicing, chromatin changes, RNA editing, and RNA interference, potential sisRNA/*cis*-NAT axes will likely have major regulatory consequences for gene expression. New tools such as CRISPR/Cas13 that allow RNA-specific interference (without changing the DNA) will make it easier to tease apart the roles of sisRNA in *cis* (Figure 8 and Box 4).

Box 1. Challenges and Technologies in sisRNA Research

Stable intronic sequence RNAs (sisRNAs), which comprise of different exceptionally stable intronic transcripts, are the latest addition to the noncoding RNAs (ncRNAs) family. SisRNAs can be divided into linear intronic sequences, intron or exon-intron circular RNAs and intron-retained transcripts. Emerging evidences suggest that sisRNAs are present in different organisms and play essential roles in regulating cellular functions. Interestingly, dysregulation of sisRNAs are also believed to lead to numerous human health disorders including tumorigenesis. The advent of next-generation sequencing has enabled genome-wide and high-throughput identification of novel sisRNAs. Commonly, the functions of sisRNAs can be studied *in vivo* and *in vitro* by RNA knockdown (RNA interference or short-hairpin RNA), over-expression and exploration of potential associated targets via complementary base-pairing and immunoprecipitation techniques. Nevertheless, studies on sisRNAs are still at the infant stage and are often deterred by challenges such as: (i) low or no detection due to low abundance; (ii) inability to distinguish between pre-mRNA and sisRNA; and (iii) lack of methods to achieve potent RNA-specific knockdown with no off-target effects. To refine and address these challenges in sisRNAs studies, several techniques and new technologies can be incorporated. Firstly, different enrichment steps can be performed, such as nuclear-cytoplasmic fractionation, RNase R treatment (circular sisRNAs selection) and polyA tail selection to improve detection and minimize noise. Alternatively, natural systems that lack transcription (e.g., red blood cells) or transcription inhibition (by using Actinomycin D) can be utilized to effectively enrich sisRNAs and distinguish them from pre-mRNAs. Secondly, characterization and visualization of sisRNAs can be enhanced via single-cell live imaging techniques such as MS2-GFP tagging, fluorogenic RNA aptamers (Spinach) and single molecular FISH. Live single-cell technologies provide accurate measurements on the abundance and subcellular localization in real-time as opposed to biochemical assays that utilize population of cells. Finally, to achieve a higher specificity in RNA knockdown, RNA gene editing technologies such as Cas13 CRISPR protein can be used to silent sisRNAs-of-interest. Importantly, Cas13 CRISPR protein has been shown to have minimum to none off-targets effects in eukaryotes.

Box 2. Methods to Identify sisRNAs

Enrichment step is one of the essential procedures in the study of sisRNAs. The first high-throughput discovery of sisRNAs was reported in 2012 from Joseph Gall's laboratory where this was made possible by using clean and well-separated nuclear and cytoplasmic fractions from the *Xenopus tropicalis* oocytes [4]. It is crucial to have a clean separation as even a small amount of cytoplasmic cross-contamination will completely mask the detection of sisRNAs in the nuclear fraction. The mature *X. tropicalis* oocyte is a good model for studying sisRNAs and nuclear-cytoplasmic fractionation as it contains a giant nucleus that can be manually dissected in an optimal isolation medium. For cell culture, nuclear-cytoplasmic fractionation can be conducted with centrifugation following homogenization in cold cell disruption buffer [6]. Other enrichment strategies can be employed along with nuclear-cytoplasmic fractionation to further enrich for different types of sisRNAs. RNase R, a 3′ to 5′ exoribonuclease that degrades all linear and Y-structure RNAs except for circular RNAs, is used for enrichment of circular sisRNAs and intronic lariats [12,23]. PolyA tail ± selections coupled with ribosomal RNA depletion can increase read depth of certain sisRNAs with or without polyA tails [4,16]. Additional strategies like transcription inhibition with actinomycin D or α-amanitin help to differentiate sisRNAs from pre-mRNAs and determine the stability of sisRNAs [27]. The quality of these enriched samples is subjected to different technical challenges associated with the procedures. Ideally, researchers may opt to look for natural systems that lack nuclear compartments such as red blood cells or are transcriptionally quiescent such as mature oocytes and early embryos [11,16,27].

Box 3. Methods to Characterize sisRNAs

To characterize sisRNAs, one can utilize single-cell live imaging techniques such as MS2-GFP tagging [43], fluorogenic RNA aptamers (Spinach) [44] and single molecule fluorescence in situ hybridization (smFISH) [45]. Expression of engineered sisRNAs with MS2 loops or tagging of endogenous sisRNAs with Spinach can be visualized through their interactions with fluorescence proteins or dyes to reveal their spatial and temporal dynamics. One can study the expression and localization of sisRNAs, their interaction with other biomolecules, formation of cellular bodies and decay at the single-cell level. For Spinach, its advantages include the ability to exhibit a palette of visible spectrum, multiple targets for different fluorophores, compact and highly responsive (visible shortly after transcription). For studies that do not require the dynamic information of sisRNAs, smFISH is an alternative approach for high accuracy detection, localization, and quantification in a single cell [27]. smFISH does not affect the nature of the endogenous sisRNAs under study, which may be altered in some live cell-imaging techniques.

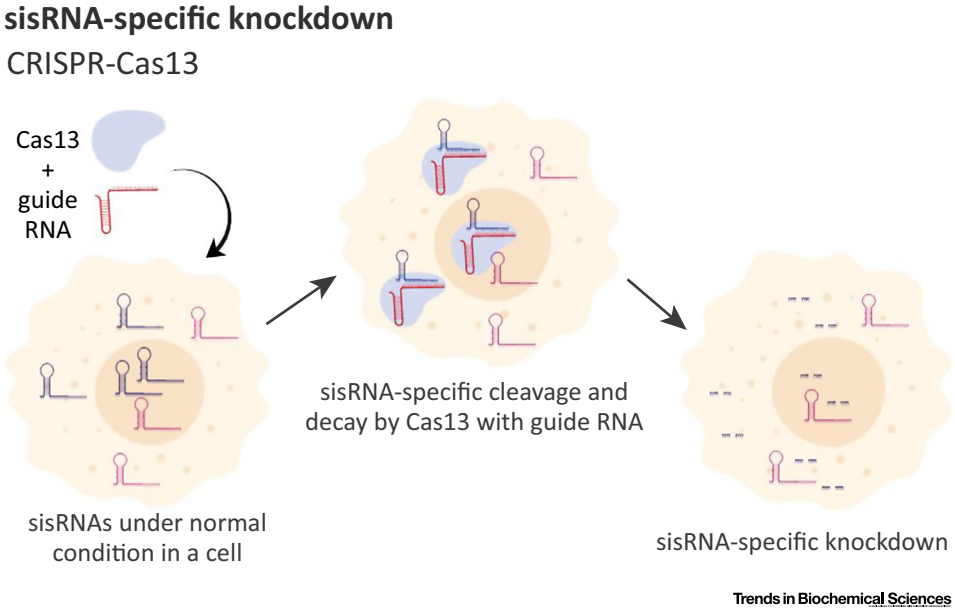

Figure 8. Phenotypic Analyses of sisRNAs. sisRNAs can be targeted specifically for knockdown via CRISPR-Cas13 technique by heterogeneous expression of Cas13 protein and sisRNA-specific guide RNA.

Box 4. Phenotypic Analyses of sisRNAs

Finally, for phenotypic analyses *in vivo*, the expression of sisRNAs are conventionally knocked down via RNAi or shRNA. Although RNAi and shRNA are common and reliable RNA knockdown tools, there are also concerns on their specificity and off-target effects. Genetic mutations that disrupt sisRNA loci can also simultaneously mutate cis-NAT loci or regulatory DNA elements (such as enhancers), thus resulting in nonspecific effects. With the discovery of CRISPR-Cas13 from *Leptotrichia wadei* (LwaCas13), it was reported that it had comparable knockdown efficiency to RNAi and shRNA, and exhibited higher target specificity, minimal to none off-target effects and no collateral activity in mammalian cells [46]. Similarly to the CRISPR-Cas9 system, the Cas13 protein and its RNA guide can be expressed heterogeneously for multiplex and nuclear RNA knockdown. Development and incorporation of CRISPR-Cas13 as tools to study the functions of sisRNAs would be advantageous and more precise functional characterization of sisRNAs can be achieved.

Outstanding Questions

What are the other possible biogenesis and decay pathways of sisRNAs?

Do sisRNAs play any significant roles in nuclear body formation and maintenance? Are there any unifying/conserved molecular mechanisms for sisRNAs?

What proteins do sisRNAs bind to?

Do the biogenesis and behavior of sisRNAs change from normal to stress conditions?

What selective pressures guide the evolution of sisRNAs?

Does dysregulation of sisRNAs in cells contribute to disorders including cancers?

Evolution of sisRNAs

In the context of evolution, the regulatory relationships between sisRNAs and their targets (complementary DNA or *cis*-NATs from the same parental loci) are less sensitive to point mutations. Nucleotide changes to the intronic DNA sequences will produce sisRNAs that contain the corresponding mutations. Provided that their secondary structures are not severely affected, these mutated sisRNAs will continue to base-pair with the DNA or *cis*-NATs bearing the complementary mutations. In other words, sisRNAs that act in *cis* have the tendency to coevolve with their targets. This hypothesis may explain why sisRNAs are generally less conserved in sequence than are structured intronic ncRNAs such as snoRNAs and scaRNAs that act in *trans*.

The need for greater layers of gene regulation for fine-tuning expression is evident for the development of multicellular organisms, where gene expression is controlled in a coordinated manner during cellular differentiation. More complex organisms such as humans have more and longer introns, potentially allowing more 'raw materials' for natural selection to work with, eventually crafting fine-tuned spatial and temporal gene expression.

Preclinical and Clinical Perspectives

sisRNAs are potential targets for the development of diagnostic and therapeutic interventions. Similarly to protein-coding genes, dysregulated expression and alteration in sisRNAs can contribute to complications in human, including the development of cancer. Although the involvement of sisRNAs in cancer has yet to be well established, sisRNAs have been shown to play essential roles in regulating numerous cellular functions. In fact, studies have shown that sisRNAs are differentially expressed in estrogen-deprived breast cancer cell lines, and lead to upregulation of genes [39], induce epigenetic silencing in colorectal cell lines [32], and associate with an array of RNA-binding proteins [18]. A recent study reported that mutations in *ebv-sisRNA-1* can inhibit B cell transformation by EBV [40]. As discussed earlier, by being relatively stable, sisRNAs serve as ideal diagnostic targets. Hence, sisRNAs that exhibit aberrant expression in diseased states can be used as cancer **biomarkers** for non-invasive screening using biofluids or biopsied tumors. Furthermore, because some sisRNAs function upstream of gene expression, targeting sisRNAs may provide a better outcome or prognosis than targeting mRNA or proteins, which are downstream of gene expression. Upregulation of specific sisRNAs, which can lead to pathogenesis, may be attenuated or controlled by developing drugs based on RNA interference therapy or gene-editing tools.

Concluding Remarks

It will be interesting in the future to examine the possibility of searching for conserved sisRNAs by synteny instead of sequence [41] (see Outstanding Questions). Furthermore, the discovery of *INE-1* sisRNAs suggests that transposable elements play a major role in the evolution of sisRNAs. Such dynamic changes in sisRNAs may provide insights to species-specific gene regulatory networks. Understand sisRNA-mediated regulatory networks may have implications for diseases such as cancer [34] and for the design of circuits in **synthetic biology** [42].

Acknowledgments

We thank Rayhan Al-Awdady Binte Ismail for comments on the manuscript. The authors are supported by the Temasek Life Sciences Laboratory.

References

1. Hesselberth, J.R. (2013) Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* 4, 677–691
2. Osman, I. *et al.* (2016) Stable intronic sequence RNAs (sisRNAs): a new layer of gene regulation. *Cell. Mol. Life Sci.* 73, 3507–3519
3. Cech, T.R. and Steitz, J.A. (2014) The noncoding RNA revolution – trashing old rules to forge new ones. *Cell* 157, 77–94
4. Gardner, E.J. *et al.* (2012) Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev.* 26, 2550–2559
5. Pek, J.W. and Okamura, K. (2015) Regulatory RNAs discovered in unexpected places. *Wiley Interdiscip. Rev. RNA* 6, 671–686
6. Moss, W.N. and Steitz, J.A. (2013) Genome-wide analyses of Epstein–Barr virus reveal conserved RNA structures and a novel stable intronic sequence RNA. *BMC Genomics* 14, 1–16
7. Moss, W.N. *et al.* (2014) RNA families in Epstein–Barr virus. *RNA Biol.* 11, 10–17
8. Hooks, K.B. *et al.* (2016) Novel intronic RNA structures contribute to maintenance of phenotype in *Saccharomyces cerevisiae*. *Genetics* 203, 1469–1481
9. Wu, H.-W. *et al.* (2018) A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves. *New Phytol.* 219, 1480–1491
10. Li, Z. *et al.* (2016) Intron lariat RNA inhibits microRNA biogenesis by sequestering the dicing complex in *Arabidopsis*. *PLoS Genet.* 12, 1–25
11. Pek, J.W. *et al.* (2015) Stable intronic sequence RNAs have possible regulatory roles in *Drosophila melanogaster*. *J. Cell Biol.* 211, 243–251
12. Tay, M.L.I. and Pek, J.W. (2017) Maternally inherited stable intronic sequence RNA triggers a self-reinforcing feedback loop during development. *Curr. Biol.* 27, 1062–1067
13. Wong, J.T. *et al.* (2017) DIP1 modulates stem cell homeostasis in *Drosophila* through regulation of sisR-1. *Nat. Commun.* 8, 1–11
14. Lu, Z. *et al.* (2015) Metazoan tRNA introns generate stable circular RNAs *in vivo*. *RNA* 21, 1554–1565
15. St Laurent, G. *et al.* (2012) Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* 13, 1–23
16. Ng, J.S.S. *et al.* (2018) Generation of *Drosophila* sisRNAs by independent transcription from cognate introns. *iScience* 4, 68–75
17. Jiang, M. *et al.* (2018) Self-recognition of an inducible host lncRNA by RIG-I feedback restricts innate immune response. *Cell* 173, 906–919
18. Tompkins, V.S. *et al.* (2018) Human regulatory proteins associate with non-coding RNAs from the EBV IR1 region. *BMC Res. Notes* 11, 1–8
19. Boutz, P.L. *et al.* (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* 29, 63–80
20. Singh, I. *et al.* (2018) Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.* 9, 1–16
21. Williamson, L. *et al.* (2017) UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene. *Cell* 168, 843–855.e13
22. Zheng, S. *et al.* (2015) Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* 161, 762–773
23. Zhang, Y. *et al.* (2013) Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806
24. Talhouarne, G.J.S. and Gall, J.G. (2014) Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *RNA* 20, 1476–1487
25. Taggart, A.J. *et al.* (2017) Large-scale analysis of branchpoint usage across species and cell lines. *Genome Res.* 27, 639–649
26. Panda, A.C. *et al.* (2017) High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic Acids Res.* 45, e116–e116
27. Talhouarne, G.J.S. and Gall, J.G. (2018) Lariat intronic RNAs in the cytoplasm of vertebrate cells. *Proc. Natl. Acad. Sci. U. S. A.* 115, E7970–E7977
28. Li, Z. *et al.* (2015) Exon–intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 22, 256–264
29. Moss, W.N. (2014) Analyses of non-coding RNAs generated from the Epstein–Barr virus W repeat region. In *Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering 2014*, pp. 238–252, IWBBIO
30. Houseley, J. and Tollervey, D. (2009) The many pathways of RNA degradation. *Cell* 136, 763–776
31. Osman, I. and Pek, J.W. (2018) A sisRNA/miRNA axis prevents loss of germline stem cells during starvation in *Drosophila*. *Stem Cell Rep.* 11, 4–12
32. Guil, S. *et al.* (2012) Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat. Struct. Mol. Biol.* 19, 664–670
33. Armakola, M. *et al.* (2012) Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat. Genet.* 44, 1302–1309
34. Pek, J.W. (2018) Stable intronic sequence RNAs engage in feedback loops. *Trends Genet.* 34, 330–332
35. Ng, A.Y.E. *et al.* (2018) Germline stem cell heterogeneity supports homeostasis in *Drosophila*. *Stem Cell Rep.* 11, 13–21
36. Isomura, A. and Kageyama, R. (2014) Ultradian oscillations and pulses: coordinating cellular responses and cell fate decisions. *Development* 141, 3627–3636
37. Ebert, M.S. and Sharp, P.A. (2012) Roles for MicroRNAs in conferring robustness to biological processes. *Cell* 149, 505–524
38. Skalska, L. *et al.* (2017) Regulatory feedback from nascent RNA to chromatin and transcription. *Nat. Rev. Mol. Cell Biol.* 18, 331–337
39. Tomita, S. *et al.* (2015) A cluster of noncoding RNAs activates the ESR1 locus during breast cancer adaptation. *Nat. Commun.* 6, 1–15
40. Szymula, A. *et al.* (2018) Epstein–Barr virus nuclear antigen EBNA-LP is essential for transforming naïve B cells, and facilitates recruitment of transcription factors to the viral genome. *PLoS Pathog.* 14, 1–34
41. Ulitsky, I. and Bartel, D.P. (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46
42. Nissim, L. *et al.* (2017) Synthetic RNA-based immunomodulatory gene circuits for cancer immunotherapy. *Cell* 171, 1138–1150
43. Grünwald, D. and Singer, R.H. (2010) *In vivo* imaging of labelled endogenous B-actin mRNA during nucleocytoplasmic transport. *Nature* 467, 604–607
44. Paige, J.S. *et al.* (2011) RNA mimics of green fluorescent protein. *Science* 333, 642–646
45. Buxbaum, A.R. *et al.* (2014) Single B-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* 343, 419–422
46. Abudayyeh, O.O. *et al.* (2017) RNA targeting with CRISPR-Cas13. *Nature* 550, 280–284