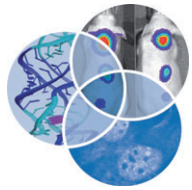
Advanced Review



Regulatory RNAs discovered in unexpected places

Jun Wei Pek[1]* and Katsutomo Okamura[1,2]*

Recent studies have discovered both small and long noncoding RNAs (ncRNAs) encoded in unexpected places. These ncRNA genes were surprises at the time of their discovery, but many quickly became well-accepted families of functional regulatory RNA species. Even after years of extensive gene annotation studies using high-throughput sequencing technologies, new types of ncRNA genes continue to be discovered in unexpected places. We highlight ncRNAs that have atypical structures and that are encoded in what are generally considered 'junk' sequences, such as spacers and introns. We also discuss current bottlenecks in the approaches for identifying novel ncRNAs and the possibility that many remain to be discovered. © 2015 Wiley Periodicals, Inc.

How to cite this article:
WIREs RNA 2015. doi: 10.1002/wrna.1309

INTRODUCTION

We once thought that the protein-coding genes in our genome could explain the diversity of life on Earth, the complexity of developmental processes, and the mechanisms by which organisms cope with stress and the ever-changing environment. However, we soon realized that the complex regulatory networks in our body are the main players in the fine-tuning of gene expression, the rapid response to environmental changes, and in maintaining memories of the past. A major player in these regulatory networks is regulatory noncoding RNA (ncRNA), both short and long.

Many ncRNAs were discovered in unexpected genomic regions, and many functional ncRNA species have unexpected structures. The discovery of microRNAs (miRNAs) is a prominent example.[1–3] Because of the unusual structures of their gene products with their short lengths (22 nt) and no protein-coding capability, this important family of genes was long overlooked despite the precedents in worms.[4,5] This history taught us that it is difficult to recognize important genes if they do not look like what we have known. At the same time, the rapid expansion of knowledge in miRNA biology in the last decade exemplifies the importance of complete gene annotation for understanding the molecular mechanisms behind biological processes.

Recent advances in the ncRNA field continue to surprise us, with the emergence of regulatory RNAs from the most unexpected places. In this review, we will cover small and long ncRNAs identified in introns, spacer sequences, and atypical structures. We also highlight the possibility that novel regulatory RNA molecules remain to be discovered due to current limitations in bioinformatics, experimental procedures, and suitable biological materials.

miRNAs FROM ATYPICAL STRUCTURES

Background

Most miRNAs are generated from short hairpin precursors via small RNA duplex intermediates produced by RNases that precisely cut at defined positions within the hairpin.[6] One duplex strand will be loaded into the effector complex containing the Argonaute proteins.[7] Taking advantage of characteristic patterns of small RNA reads produced by this processing mechanism, the identification of novel

*Correspondence to: junwei@tll.org.sg; okamurak@tll.org.sg

[1]Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore, Singapore

[2]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

Conflict of interest: The authors have declared no conflicts of interest for this article.

© 2015 Wiley Periodicals, Inc.

miRNA genes usually relies on small RNA sequencing evidence that indicates the precisely defined cleavage positions and typical small RNA duplex structures.[8]

Because of the advances in deep sequencing technologies and associated bioinformatics tools to identify miRNA genes, it is often assumed that all biologically important miRNA genes have been discovered in major model organisms and humans. However, new types of miRNAs continue to be discovered, disproving this general assumption. These include single-stranded RNA species directly loaded to Argonaute effectors without forming a duplex intermediate structure.[9] In addition, there may be classes of miRNA genes whose hairpin structures are difficult to predict, even though the miRNA reads are indeed produced as duplex species. The following section will focus on miRNA genes that highlight potential caveats of miRNA discovery tools that rely on predictable duplex structures.

Intron-Split miRNAs

Recent studies have uncovered miRNAs generated from hairpins that are split by introns. In plants, such intron-split miRNAs may form a class of miRNA genes.[10,11] Sunkar et al. identified several miRNAs occurring in the exons of annotated protein-coding mRNAs. This includes *MIR444*, a miRNA that is broadly conserved in monocots. The *MIR444* hairpin structure could not be recognized in the genome sequence because the hairpin was split by two introns (Figure 1(a)). Mature *MIR444* could be detected by Northern blotting in rice and maize, and potential target cleavage events were also verified by 5′ RACE assays. These results indicated that biologically active regulatory RNA molecules could be produced from intron-split miRNAs.

Later, *MIR444* and its paralogs were found to form a family termed nat-miRNAs (natural antisense transcript-miRNAs).[11] *MIR444* genes target a family of transcription factors containing a MADS box, and the host transcripts of the miRNAs were generated as antisense transcripts of target MADS box genes (Figure 1(a)). Without splicing, if a transcript from one DNA strand has a hairpin structure, the antisense strand that spans the same region should also have a similar hairpin structure. However, due to the presence of introns splitting the nat-miRNA hairpins, the sense transcripts of MADS box proteins do not contain miRNA-like hairpin structures, thereby avoiding cleavage by the miRNA processing machineries. Instead, the sense protein-coding mRNA would have a target sequence completely complementary to the mature miRNA sequence (Figure 1(a)).

Although such intron-split miRNAs have not been discovered in metazoan genomes, the possibility still exists. Elegant experiments were performed in *Caenorhabditis elegans* as a proof of concept.[12] In this study, they made use of a genetic mutant of *lin-4*, a miRNA mutant that exhibits heterochronic phenotypes.[13] This phenotype could be rescued by the injection of DNA constructs containing the wild-type *lin-4* gene sequence (Figure 1(b)). By modifying this rescue plasmid, a variety of rescue constructs containing introns in the hairpin region were generated and used to ask whether such intron-containing miRNA hairpins (inc-miRs) could be efficiently processed into functional mature species. Most inc-miR constructs could efficiently rescue the mutant phenotype, even those containing insertions in the mature miRNA region. Consistent with the idea that this process was splicing dependent, the removal of the consensus group II intron structure (gu….ag) from an intron-containing rescue construct resulted in a severe decrease in rescue activity. Although this was still a completely artificial system, the study highlighted the possibility that active, mature miRNAs could originate from intron-split hairpins whose structures were difficult to predict based on the reference genome sequences. Therefore, it is possible that animal inc-miRs were overlooked in previous studies.

TRANSCRIBED SPACERS IN CANONICAL RNA PRECURSORS WITH REGULATORY ROLES

Background

A challenge in ncRNA gene annotation is to distinguish functional ncRNAs and processing byproducts/partial degradation products of other abundant RNA species. Sequencing reads originating from abundant ncRNAs are often discarded by considering them as 'background reads'. Nevertheless, accumulating evidence argues that specific fragments generated from mature and precursor tRNAs have trans-regulatory activity under certain conditions.[14] In addition, recent findings uncovered that trans-regulatory RNA species are generated from spacers of rRNA or tRNA precursors in particular organisms, in contrast to the previous notion that those spacers were mere byproducts. In this section, we review recent studies that uncovered spacer species with gene regulatory activity.

© 2015 Wiley Periodicals, Inc.

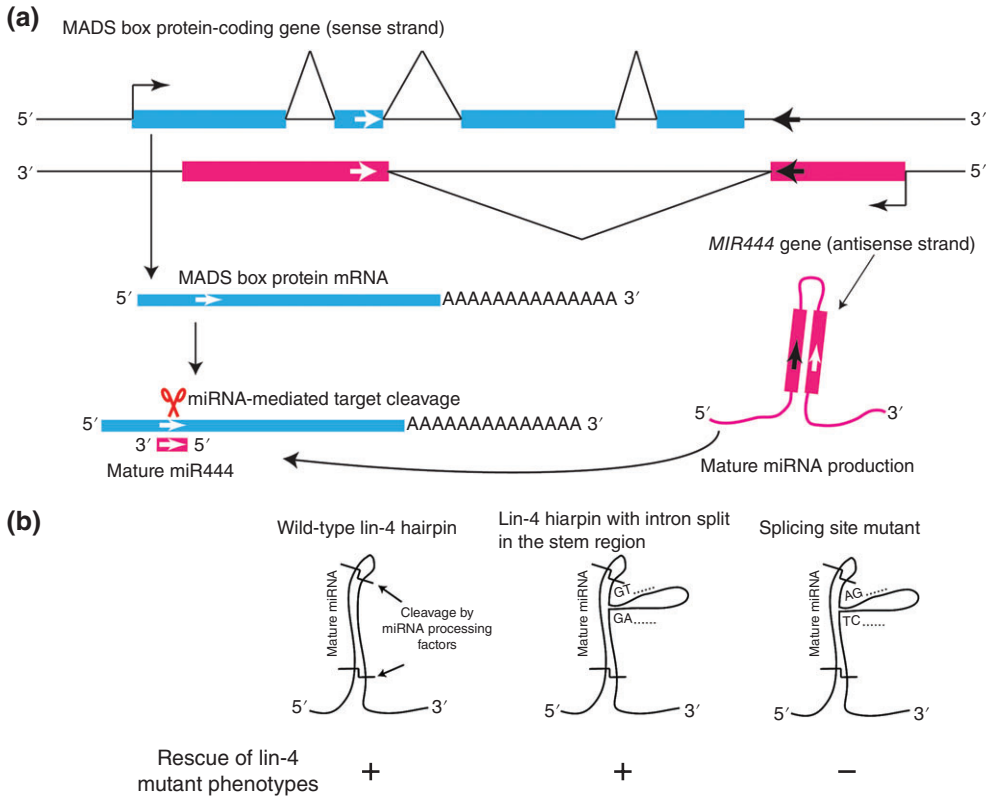


FIGURE 1 | Intron-split miRNAs. (a) nat-miRNAs (natural antisense transcript-miRNAs) in monocots. The gene model illustrates the structure of a MADS box protein/*MIR444* locus. The MADS box protein-coding gene and the *MIR444* gene are located on the upper and lower DNA strands, respectively. Two parts of the *MIR444* hairpin sequence are far apart in the genomic DNA sequence, but brought closer after the splicing of primary miRNA transcripts. This allows the spliced transcript to be cleaved by the miRNA processing machineries, resulting in the production of mature miR444. On the other hand, because the MADS protein mRNA does not include one part of the *MIR444* antisense sequence, this transcript does not form a miRNA-like hairpin. However, because the mature miRNA product is produced from the region that is perfectly complementary to a part of the MADS protein mRNA (white arrows), the expression of *MIR444* results in the downregulation of the MADS protein via direct cleavage of the mRNA mediated by Argonaute proteins. (b) Artificial inc-miRs (intron-containing miRNAs) in *Caenorhabditis elegans*. The heterochronic phenotypes in the *lin-4* mutant were rescued when mutant animals were injected with a plasmid encoding wild-type *lin-4*. Artificial mutant constructs containing intron insertions in the *lin-4* stem region also rescued the phenotype, indicating that miRNAs could be produced even with an intron insertion. When the consensus sequences essential for splicing (GU……..AG) were mutated, the rescue activity was diminished, suggesting that the splicing of inserted introns is essential for efficient production of mature *lin-4* miRNA.

A miRNA from the ITS of Pre-rRNA in *Drosophila melanogaster*

The internal transcribed spacer (ITS) sequence of the *Drosophila* ribosomal RNA precursor (pre-rRNA) was recently found to encode an evolutionarily conserved miRNA (Figure 2(a)).[15] Small RNA analysis using fly ovaries revealed that precisely defined small RNAs were generated from both arms of a hairpin located in the ITS1 region. These reads were enriched in the purified Argonaute complexes, indicating that this is a previously unknown miRNA. This hairpin was renamed *mir-ITS1/mir-10404*. *mir-ITS1/mir-10404* was highly expressed in a variety of tissues, and its expression level in fly ovaries was ranked at ~20th–30th among the 240 miRNA genes. Despite its robust expression, this miRNA was overlooked in previous studies, potentially because the small RNA sequence can be mapped to many places in the genome. It is standard practice to remove such reads mapping to multiple positions in the reference genome prior to analysis.

In contrast to the rapid evolution of other ITS regions, the *mir-ITS1/mir-10404* hairpin sequence is highly conserved, even in fly species that were separated from the *Drosophila* ancestor ~100 million years ago.[16,17] The seed sequence, a 7-nt sequence residing at the 5′ end of the miRNA that is most important for miRNA target regulation, is perfectly conserved, while there are some nucleotide substitutions in other regions. These observations suggested

© 2015 Wiley Periodicals, Inc.

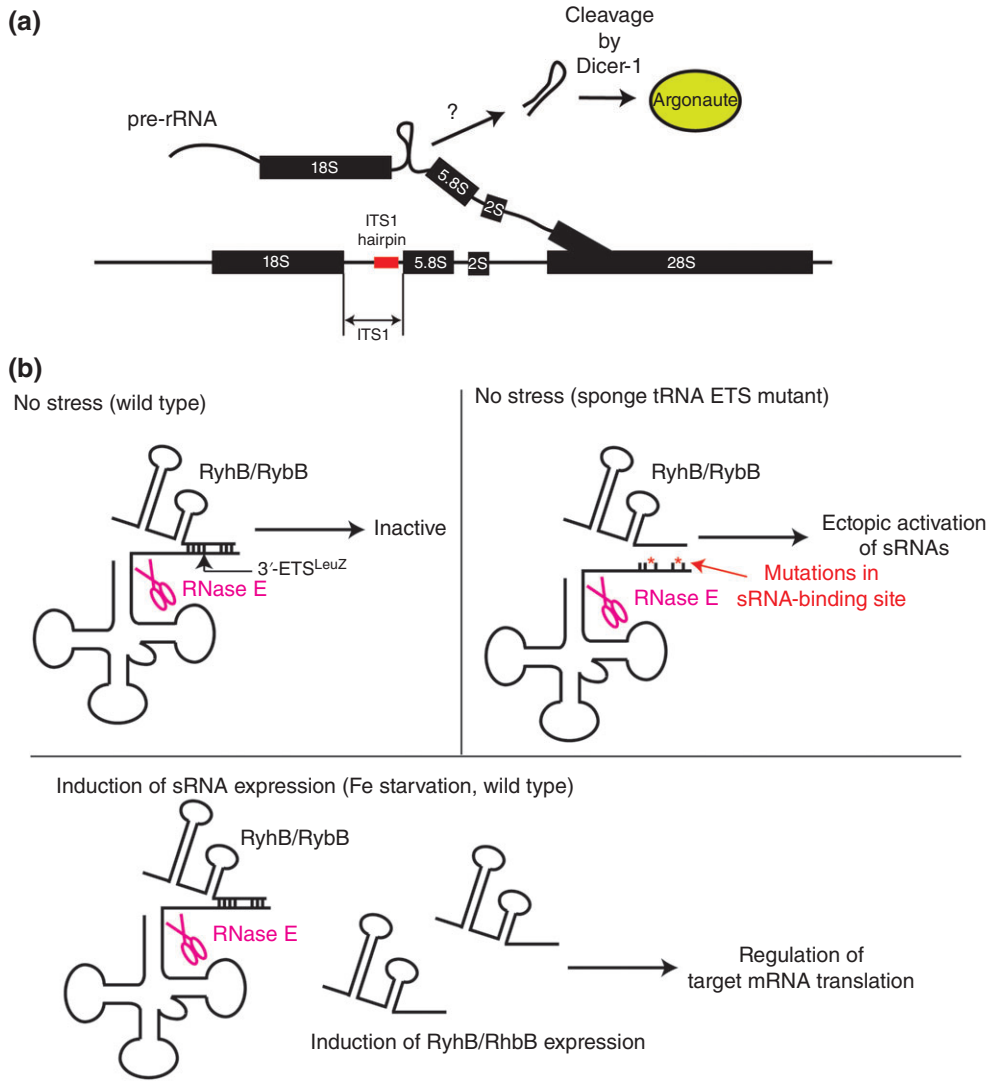


FIGURE 2 | Spacer sequences with regulatory activity. (a) An miRNA produced from the fly rRNA ITS1 (internal transcribed spacer 1). A transcribing rDNA unit is shown. The fly rRNA precursor is transcribed as polycistronic RNA containing 18S, 5.8S, 2S, and 28S rRNAs. Transcribed spacer sequences between the mature rRNA sequences are known as ITS. The conserved miRNA hairpin is found in the ITS1 region located between 18S and 5.8S rRNA sequences. The hairpin is processed into mature species by a Drosha-independent Dicer-1-dependent mechanism, and mature miRNA products are loaded to Argonaute effector complexes. The enzyme producing pre-miRNA hairpins from pre-rRNAs is currently unknown. (b) A tRNA spacer acting as an sRNA (small regulatory RNA) sponge in *E. coli*. External transcribed spacers (ETS) of tRNAs are cleaved from tRNA precursors by RNase E during tRNA maturation. In wild-type bacteria, the iron-starvation-responsive sRNAs RyhB and RybB bind to a 3′-ETS of a Leu tRNA (3′-ETS^{leuZ}). This binding ensures the complete repression of RyhB/RybB activity under no stress conditions. When the sRNA-binding site on 3′-ETS^{leuZ} is mutated, the sRNAs cannot bind 3′-ETS^{leuZ}, leading to the ectopic activation of sRNA activity under no stress conditions. Under iron starvation, the expression of RyhB and RybB is elevated and 3′-ETS^{leuZ} can no longer repress the activity of these sRNAs. Therefore, these sRNAs can bind target mRNAs to regulate their expression levels.

the functional importance of *mir-ITS1/mir-10404* as a regulatory RNA gene, and it will be important to study its functions and identify its important targets.

This study reminded us that bioinformatics filters such as the removal of repetitive sequences can cause loss of information regarding important genes. On the other hand, care has to be taken when annotating small regulatory RNA loci within such extremely actively transcribed genes. Abundant transcripts always produce small RNA fragments, presumably through partial degradation. For example, in the case of pre-rRNA-derived miRNA gene, many previously annotated miRNA genes in mammals lack supporting evidence for precise processing even in large collections of small RNA libraries, suggesting

© 2015 Wiley Periodicals, Inc.

that they were erroneous annotations.[18] As a result, most mammalian miRNAs overlapping with rDNA were recently removed from miRBase.[8] Therefore, requiring a higher standard for the annotation of small regulatory RNA genes overlapping with highly abundant RNAs is important. Preferably, those miRNA loci should be supported by experimental evidence such as enrichment in the effector RNP complexes or reporter assays detecting the regulatory activity of mature miRNAs.

A sRNA Sponge from the 3′-ETS of tRNA in *Escherichia coli*

A recent study in *E. coli* identified tRNA 3′-external transcribed spacers (3′-ETS) as a molecular sponge regulating the activities of small regulatory RNAs (sRNAs).[19] The study aimed to identify targets of the sRNAs RyhB and RybB, which are induced during iron starvation and regulate a number of genes related to iron homeostasis. To identify potential targets, the RNA molecules that copurified with the complexes containing RyhB/RybB were cloned. By this approach, many tRNA sequences, including the 3′-ETS of the leuZ tRNA gene (3′-ETS^{leuZ}), were identified as candidate target RNAs.

Functional analyses of the interaction between the Leu tRNA ETS and sRNAs suggested that a major role for 3′-ETS^{leuZ} is to act as a molecular sponge of sRNA molecules to ensure complete repression of sRNA activity under iron-rich conditions (Figure 2(b)). The mutant strain carrying mutations in the RyhB/RybB target site in 3′-ETS^{leuZ} exhibited misregulation of mRNAs controlled by RyhB or RybB (Figure 2(b)). In addition, the 3′-ETS^{leuZ} mutant showed a significant delay in cell growth in a RyhB/RybB-dependent manner, demonstrating the physiological importance of the sRNA sponge activity of 3′-ETS^{leuZ}.

Sequence conservation analysis revealed that many ITS or 3′-ETS sequences including 3′-ETS^{leuZ} were evolutionarily constrained, suggesting a widespread functionality of tRNA spacers. In fact, additional tRNA spacer–sRNA pairs were biochemically verified. An interesting speculation is that robust and constitutive expression of tRNA genes under non-stressed conditions may have made tRNA spacers an ideal place for sponge genes because such sponge molecules would have to accumulate at high levels to repress sRNA activity.

This study again highlights the importance of unbiased screens and subsequent functional assays to ensure the regulatory activity of spacer sequences.

STABLE INTRONIC SEQUENCE RNAs

Background

Since the discovery of introns in 1977,[20,21] it has been shown that most coding sequences in the eukaryotic genome are interrupted by intervening sequences (introns). Early experiments revealed that after splicing, intronic sequences are degraded within seconds or minutes in the nucleus.[22] Therefore, intronic RNAs are generally regarded as junk, having no functions. Later, it was found that some intronic RNAs can be processed into ncRNAs that have biological activities in the cells, such as small nucleolar RNAs (snoRNAs), small Cajal body-associated RNAs (scaRNAs) and miRNAs.[23–31] In addition to these major classes of ncRNAs, only a few examples of intronic RNAs with unusually high stability were described in the past 28 years. No clear biological functions were assigned to these intronic RNAs, and they were considered to be exceptional cases without considerable interest.

Recent developments in deep sequencing and methods for RNA enrichment and isolation have led to the discovery of the widespread occurrence of stable intronic sequence RNAs (sisRNAs) (Figure 3). This surprising observation immediately poses two important questions. First, how are sisRNAs produced? Second, what is the biological significance of sisRNAs? Over the past 3 years, progress has been made in understanding sisRNAs in various model organisms.

sisRNAs from *Xenopus* Oocytes

The first genome-wide identification of sisRNAs was performed in the *Xenopus tropicalis* oocyte germinal

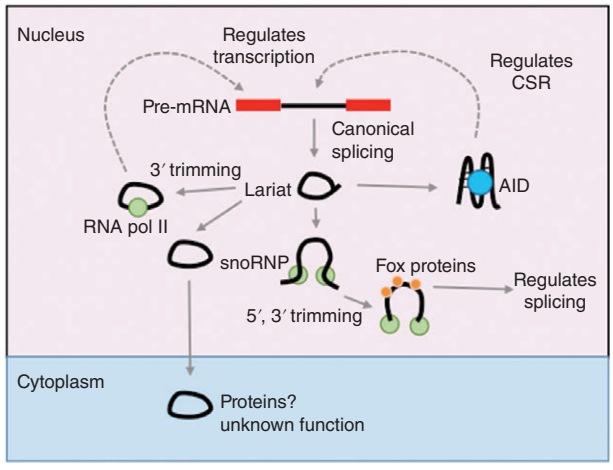


FIGURE 3 | Biogenesis and functions of sisRNAs. Diagram showing the biogenesis of sisRNAs. Exons and introns are in red and black, respectively. Circles depict the proteins that the RNAs bind.

© 2015 Wiley Periodicals, Inc.

vesicle (GV, or nucleus).[32] The detection of sisRNAs was only possible after obtaining pure nuclear RNA. GVs were dissected and nuclear envelopes manually removed to avoid contamination from highly abundant cytoplasmic mRNAs. Deep sequencing of GV RNAs revealed that more than 90% of the genes transcribed during oogenesis produce sisRNAs (Table 1). These sisRNAs were stable for at least 2 days, therefore defining a previously undescribed class of ncRNAs. Interestingly, in some cases one intron can give rise to more than one discrete fragment of sisRNA, suggesting different regulatory functions. It is unclear if these sisRNA fragments are processed from a single spliced-out intron or via recursive splicing. These sisRNAs exist as both circular (lariat without tail) and linear molecules, suggesting more than one pathway of sisRNA biogenesis (Figure 3).[40] Further analysis of nuclear lariat sisRNAs will give insights into whether recursive splicing is the mechanism for generating discrete sisRNA fragments from a single intron.

Why were sisRNAs not previously discovered in deep sequencing experiments of nuclear RNAs in other model systems? It is possible that sisRNAs were present in other model systems but were generally ignored because intronic reads were often regarded as coming from pre-mRNAs. Alternatively, the abundance of sisRNAs may be too low to be detected. The *Xenopus* oocyte constitutes a unique system to search for sisRNAs. Oogenesis takes months to

TABLE 1 | Summary of Discovered sisRNAs

Species	Description and Types	Tissues	Functions	References
Human, mouse	T cell receptor-β locus, nuclear, cytoplasmic, lariat	T cells	ND	Refs 33,34
Human, rat	*Pem* homeobox gene, nuclear, cytoplasmic	Cell lines	ND	Ref 35
Human	Genome-wide, nuclear, linear, flanked by snoRNA sequences	Cell lines	Binds splicing regulators and alters splicing patterns	Ref 36
	Genome-wide, nuclear, lariat		Positively regulates Pol II transcription	Ref 37
Mouse	Nuclear, linear, forms G-quadruplex structures	CH12 B lymphoma cell line	Binds and targets AID to S regions critical for CSR	Ref 38
Xenopus laevis	Simian virus 40, nuclear, lariat, 140 nt	Oocytes	ND	Ref 39
	Genome-wide, cytoplasmic, lariat	Oocytes	ND	Ref 40
Xenopus tropicalis	Genome-wide, nuclear, linear, and lariat	Oocytes, embryos	ND	Refs 32,40
	Genome-wide, cytoplasmic, lariat	Oocytes, embryos	ND	Ref 40
Drosophila melanogaster	*delta* locus	Embryos	ND	Ref 41
	tRNA locus	Larvae, pupae, adults	ND	Ref 42
Herpes simplex virus	Latency-associated transcript, nuclear, lariat	Human, mouse, monkey cells	Inhibits transactivation of gene expression by infected-cell polypeptide 0; blocks apoptosis of infected neuronal cells; induces heterochromatin formation	Refs 43–45
Epstein–Barr virus	Nuclear, 81 nt	Human B lymphocytes	ND	Ref 46
Cytomegalovirus (human and murine)	5 kb (human), 7.2 kb (murine), nuclear, lariat	Human and mouse cells	Murine 7.2 kb sisRNA facilitates the progression from the acute to the persistent phase during an infection	Refs 47,48

AID, activation-induced cytidine deaminase; CSR, class switch recombination; ND, not determined; sisRNAs, stable intronic sequence RNAs.

© 2015 Wiley Periodicals, Inc.

complete, and RNAs produced by the GV are stored over a long period of time. Therefore, the GV accumulates enough sisRNAs to be detected by deep sequencing. It is also important to consider that although the G2 chromosomes in the oocyte GV are actively transcribing, the absolute amount of pre-mRNAs is not sufficient to be detected by deep sequencing.[32] Taking these factors into consideration, one can unambiguously conclude that the intronic reads are derived from sisRNAs, not pre-mRNAs.

Outside the nucleus, sisRNAs were also found in the cytoplasm of *Xenopus tropicalis* oocytes.[40] Cytoplasmic sisRNAs are derived from a smaller percentage of introns (~5%) and mostly from short introns between 200 and 500 nt in length. One interesting aspect of cytoplasmic sisRNAs is that they are all circular molecules (Figure 3). How these sisRNAs are protected from the lariat debranching enzyme and exported into the cytoplasm remains an interesting question.

It was known from simian virus 40 (SV40) DNA injection experiments that sisRNAs existed in *Xenopus laevis* oocytes.[39] The ~140-nt SV40 sisRNAs accumulated as lariats in the nuclei. Recent deep sequencing experiments identified nuclear and cytoplasmic sisRNAs in *Xenopus laevis* similar to those in *Xenopus tropicalis*.[40] However, no sequence conservation or positional conservation in orthologous genes was found, suggesting the rapid evolution of sisRNA sequences.

sisRNAs in Other Organisms

The unique environment of *Xenopus* oocytes was useful for researchers to distinguish sisRNAs and fragments of pre-mRNAs, but this does not mean that sisRNAs exist only in oocytes. In fact, similar sisRNAs are broadly observed in other systems (Table 1). The earliest sisRNA discovered in mammalian cells was from the T cell receptor-β locus in mouse and human.[33,34] This sisRNA was shown to localize to both the nucleus and the cytoplasm as a lariat. Later, another sisRNA was discovered from the *Pem* homeobox gene in human and rat cell lines.[35] Similarly, this sisRNA was found in both the nucleus and the cytoplasm.

Recently, deep sequencing analysis of nonpolyadenylated or polyA(−) RNAs (Box 1) from HeLa cells and human embryonic stem cells revealed a number of linear sisRNAs that exhibit interesting biogenesis and stabilization mechanisms. These snoRNA-related lncRNAs (sno-lncRNAs) have either box C/D or box H/ACA sequences at both ends (Figure 3).[36] They are thought to be processed similarly to snoRNAs by lariat debranching and exoribonuclease-mediated trimming. The box C/D or box H/ACA sequences bind to snoRNP components to promote stability by blocking further degradation.

Five abundant sno-lncRNAs map to a genomic region specifically deleted in Prader–Willi Syndrome (PWS). Box C/D or box H/ACA snoRNAs and scaRNAs localize to the nucleoli and Cajal bodies, respectively. Interestingly, the PWS region sno-lncRNAs do not localize to the nucleoli or Cajal bodies; rather, they localize to the sites of transcription on the paternal chromosomes, suggesting that they do not function as canonical snoRNAs or scaRNAs. Instead, sno-lncRNAs bind to the splicing factor Fox2, suggesting that they act as 'molecular sinks' to regulate splicing.

By analyzing polyA(−) RNAs in HeLa and human embryonic stem cells, a class of circular sisRNAs called circular intronic RNAs (ciRNAs) was identified.[37] ciRNAs were proposed to be formed after splicing, and the lariats escaped debranching and further trimming at the 3′ end by exoribonucleases. It was proposed that a C-rich motif near the branch point bound to a G-rich motif near the 5′ splice site, thus forming secondary structures that inhibit the activity of lariat debranching enzyme. Studies of a ciRNA (*ci-ankrd52*) from the *ANKRD52* gene showed that it localizes to the site of transcription and is involved in promoting its host gene expression. *Ci-ankrd52* interacts with phosphorylated RNA polymerase II and modulates the rate of transcription of *ankrd52*.

A recent study identified the role of a sisRNA in regulating class switch recombination (CSR) in mice (Figure 3).[38] Transcription of the immunoglobulin switch (S) regions produces primary transcripts, which are spliced to produce intronic switch transcripts that are further debranched to form mature sisRNAs. These sisRNAs form G-quadruplex structures, which bind to activation-induced cytidine deaminase (AID). They served as guide RNAs by recruiting AID to the complementary S region DNA locus for efficient CSR. This study provides evidence for an important physiological function of sisRNAs in mammals.

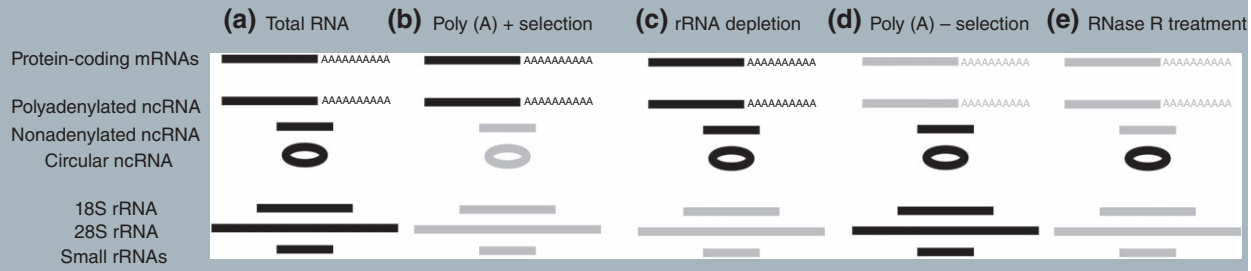
sisRNAs have also been found in invertebrates. The long introns of the *delta* gene were found to produce several sisRNAs that localize near the sites of transcription in the nucleus.[41] It was unclear whether these sisRNAs were linear or in the form of lariats. Interestingly, a recent study identified circular sisRNAs from the tRNA introns (tRNA intronic circular or tricRNAs) in *Drosophila*.[42] Therefore, it is highly

© 2015 Wiley Periodicals, Inc.

BOX 1

VARIOUS RNA FRACTIONATION METHODS.

(A) Total RNA samples are generally assumed to contain all cellular RNA molecules.
(B) PolyA(+) selection uses immobilized oligo-dT oligonucleotides and enrich for poly(A)-containing RNAs. This will enrich for mRNAs and polyadenylated ncRNAs.
(C) For rRNA depletion, immobilized oligonucleotides complementary to rRNA sequences are used. The fraction that does not hybridize with the oligonucleotides is used as the rRNA-depleted fraction.
(D) PolyA(–) selection uses the unbound fraction of polyA(+) selection. For deep sequencing experiments, this is usually combined with rRNA depletion or cDNA normalization to enrich for regulatory RNAs.
(E) RNase R is a robust 3′ → 5′ exonuclease that efficiently degrades linear RNA species while leaving circular RNA molecules intact. Therefore, RNase R-treated samples enrich circular RNA molecules. This is often combined with rRNA depletion to ensure the effective removal of rRNA.



likely that sisRNA production can be observed in a broad range of organisms. Comprehensive annotation and functional tests will be needed to understand the importance of sisRNAs.

Viruses

SisRNAs encoded in the viral genome had been recently reviewed.[49] Because viruses have small genomes, the fact that they encode sisRNAs strongly suggested their important regulatory roles (Table 1). The first viral sisRNA was discovered in the herpes simplex virus (HSV).[43–45] This sisRNA was also known as the latency-associated transcript (LAT), which accumulates in the nucleus as a lariat. It inhibits transactivation by infected-cell polypeptide 0 (ICP0), blocks apoptosis of infected neuronal cells, and promotes heterochromatin formation.[43,50–52]

In human and murine cytomegaloviruses (CMV), orthologous 5 and 7.2 kb sisRNAs were identified.[47,48] Both sisRNAs localize to the nucleus and most likely exist in the form of lariats. The murine 7.2 kb sisRNA is thought to promote the progression from the acute to the persistent phase during infection.

Recently, a linear sisRNA of ~81 nt was identified by deep sequencing nuclear RNA from cultured human B lymphocytes infected with Epstein–Barr virus.[46] Interestingly, this sisRNA appears to be a linear molecule and is relatively smaller than those found in HSV and CMV.

CIRCULAR sisRNAs AND VIROIDS

In a broad sense, viroids may be considered a form of functional circular RNAs. Viroids are the smallest infectious agents known, with circular single-stranded RNAs ranging from 250 to 400 nt in size. It had long been hypothesized that viroids originated from host genetic material. Because of their circular RNA nature, viroids are thought to have originated from spliced-out intron lariats.[53,54] Later, it was shown that a viroid complementary RNA could base pair with the 5′ end of U1 snRNA though sequence complementary.[55] It was also shown that some viroids contain a 16 nt sequence consensus also present in Group I introns. Furthermore, viroids also fold into similar secondary structures homologous to Group I introns.[56] These observations suggested that viroids either originated from 'escaped' introns or that both viroids and introns evolved from a common ancestor.[57] Thus far, viroids are found only in plants, with hepatitis D virus as the closest example to a viroid in animals.[58] The hepatitis D virus is a ~1700-nt-long circular RNA that folds into an unbranched rod-like structure.

© 2015 Wiley Periodicals, Inc.

If viroids are related to introns, then why have no viroids or viroid-like molecules been found in animals? The discovery of circular sisRNAs may provide valuable insights into this question. The cytoplasmic circular sisRNAs found in *Xenopus tropicalis* are mostly 200–500 nt long, similar to viroids (250–400 nt).[40] It will be interesting to examine the secondary structures of circular sisRNAs to determine if they fold into rod-like structures resembling certain viroids. It is tempting to speculate that some circular sisRNAs may function as endogenous viroids as opposed to infectious viroids. This idea will be similar to comparing endogenous retrotransposons with retroviruses.

CIRCULAR RNAs (circRNAs)

Background

Circular RNAs (circRNAs) were discovered more than 20 years ago, with the mouse *SRY*, human *ets-1* and *DCC* as notable examples.[59–63] However, much like the history of miRNA studies, the generality of circRNA formation was not recognized until recently. The recent revival of interest in circRNAs partly came from the detailed characterization of the *CDR1* antisense transcript.[64] While searching for a possible nuclear role for miR-671 in human cells, *CDR1as* was found to be a target. Interestingly, further analysis of *CDR1as* revealed that it is a circRNA. It was soon clear that *CDR1as* belonged to a huge class of circRNAs with regulatory functions (Table 2).

Traditional deep sequencing experiments using poly(A)-selected RNAs completely missed this class of noncanonical RNAs (Box 1). Recent deep sequencing of RNase R-treated and polyA(–) RNAs together with bioinformatics pipelines designed for the detection of back-spliced circRNAs have identified thousands of circRNAs in human, mouse, round worm (*C. elegans*), fruit fly (*Drosophila*), fungi,

TABLE 2 | Summary of Discovered circRNAs

Species	Description and Types	Tissues	Functions	References
Human, rodents	*DCC* gene	Cell lines	ND	Ref 61
Human	*ets-1* gene	Cell lines	ND	Ref 60
Human	*ANRIL* gene, correlates with *INK4/ARF* expression	Cell lines and human peripheral blood T lymphocytes	ND	Ref 62
Human	Cytochrome P-450 2C18 gene	Epidermis	ND	Ref 63
Rat	*Androgen-binding protein* gene	Testis	ND	Ref 63
Mouse	*SRY* gene	Testis	miRNA sponge	Ref 59
Human, mouse	ciRS-7/CDR1as	Brain	miRNA sponge	Refs 65,66
	Genome-wide	Brain	ND	Refs 66,67
Human	EIciRNAs, retained introns	Cell lines	Binds to RNA polymerase II and U1 snRNP; promotes transcription of parental genes	Ref 68
Human, mouse, *Caenorhabditis elegans*	Genome-wide	Various tissues and cell lines	ND	Ref 65
Human	Genome-wide	Human mammary epithelial cells before and after epithelial–mesenchymal transition	ND	Ref 69
	Genome-wide	Human embryonic stem cells	ND	Ref 70
Drosophila melanogaster	Genome-wide	Various tissues and cell lines	ND	Refs 71,72
	Genome-wide	Head and S2 cells	ND	Ref 73
C. elegans	Genome-wide	Various tissues	ND	Ref 72

miRNAs, microRNAs; ND, not determined.

© 2015 Wiley Periodicals, Inc.

plants (*Arabidopsis*), and protists.[65–67,69–74] Detailed investigations of a few circRNAs have revealed interesting functional insights. Furthermore, much has progressed on the biogenesis mechanism of circRNAs. A major challenge remains the establishment of the biological significance of circRNAs in regulating gene expression.

Biogenesis

Studies in various model organisms revealed common principles for the production of circRNAs by backsplicing (Figure 4). The circRNA production is promoted by complementary sequences or Alu repeats within the flanking introns.[68,70,73,75] However, roles of complementarity between the flanking introns remain unclear in *Drosophila*. Although assays using a panel of model circRNA minigenes containing truncations in the flanking introns suggested the importance of base pairing between the flanking introns, genome-wide analysis of circRNAs failed to detect a correlation between the level of circularization and the pairing formed by the flanking introns.[71,73] A recent study revealed that in *Schizosaccharomyces pombe*, circular RNAs are generated via an exon-containing precursor, defining an alternative mode of circular RNA biogenesis in genes that lack flanking complementary intronic sequences.[76]

In *Drosophila*, circRNA production is a cotranscriptional event, and the splicing factor Muscleblind (Mbl) promotes circularization by binding to the Mbl-binding sites in the flanking introns.[73] Interestingly, the involvement of Mbl appears to be conserved in humans. Another interesting observation from *Drosophila* is that circRNAs accumulate in neural tissues and are preferentially generated from neural genes.[71]

In *C. elegans*, the introns flanking circRNAs are enriched in RNA editing sites or hyper-editing events (Figure 4).[75] Knockdown of ADAR1, a double-stranded RNA editing enzyme, promotes the formation of circRNAs, further suggesting that intronic regions are edited to prevent complementary base pairing and therefore circularization. The involvement of ADAR1 in the regulation of circRNA biogenesis appears to be conserved in humans and mice.[67]

A recent study reported that circRNA expression is highly regulated during human epithelial–mesenchymal transition (EMT).[69] This event is regulated by the alternative splicing factor Quaking (QKI), which binds to QKI-binding sites flanking the circRNAs (Figure 4). Another study reported that circRNA expression is also regulated during neuronal differentiation in mice, further suggesting a role for circRNAs in controlling cellular processes.[67]

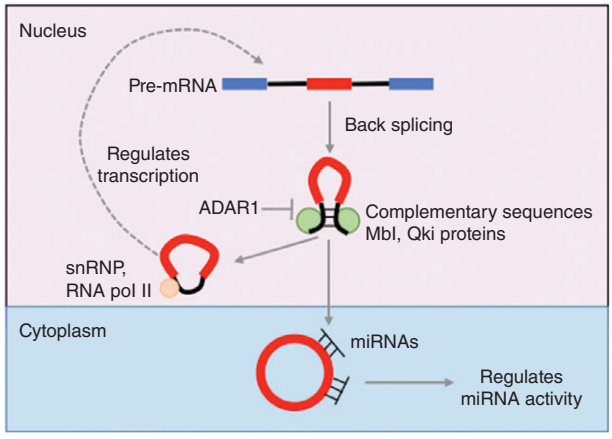


FIGURE 4 | Biogenesis and functions of circRNAs. Diagram showing the biogenesis of circRNAs. Exons and introns are in red/blue and black, respectively. Circles depict the proteins that the RNAs bind.

Functions of circRNAs

In principle, circRNAs are resistant to exoribonucleases. This feature makes circRNAs ideal sponge molecules to absorb regulatory RNAs that recruit exonuclease complexes, such as miRNAs.[77] Indeed, they can function as effective sponges for miRNAs in the cells (Figure 4).

Studies in human and mouse brain had found a circRNA (ciRS-7/CDR1as) that functions as a sponge for an miRNA.[65,66] ciRS-7/CDR1as harbors 73–74 seed target sites for miR-7. Importantly, these sites are conserved in eutherian mammals. ciRS-7/CDR1as binds to miR-7 and AGO2 as a complex and colocalizes to P-bodies in the cytoplasm in a miR-7-dependent manner. ciRS-7/CDR1as limits the activity of miR-7 in cell culture and zebrafish, thereby functioning as a sponge. Interestingly, another well-known circRNA, the testis-specific SRY, contains 16 seed target sites for miR-138 and appears to also function as a sponge.[66] In addition, examination of the fly circRNA sequences identified conserved seed target sites of abundant and conserved miRNAs, suggesting potential regulatory relationships between circRNAs and miRNAs in *Drosophila*.[71]

A recent study uncovered another interesting function of circRNA in regulating gene expression. A class of circRNAs with retained introns (EIciRNAs) was found to be associated with RNA polymerase II in human cells.[78] These EIciRNAs localize to the nucleus and bind to the U1 snRNP, possibly via the

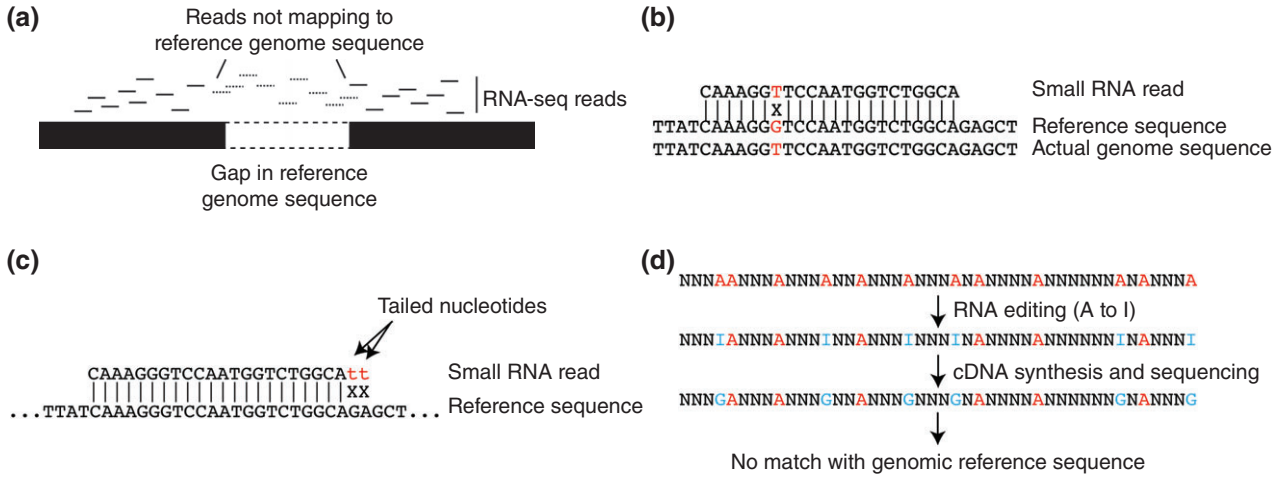
© 2015 Wiley Periodicals, Inc.

5′ splice sites. It is hypothesized that the EIciRNAs recruit U1 snRNP to the promoters of the parental genes to promote transcription in *cis*. This mechanism suggests a possible role for EIciRNAs in mediating positive feedback and fine-tuning gene expression.

OBSTACLES AND FUTURE PERSPECTIVES

Although high-throughput sequencing technologies have markedly broadened our knowledge of ncRNAs, the continued discoveries introduced above suggest that important ncRNA species from unexpected loci or with unexpected chemical structures could still be overlooked. Here, we review the limitations of widely used molecular biology and bioinformatics techniques and discuss what might remain to be discovered.

Bioinformatics analyses for the identification of novel ncRNA genes almost always depend on assembled genomic sequences as reference sequences. However, the assumption that assembled genome sequences represent real genome sequences is not true. Because shotgun sequencing, a technology widely used to sequence eukaryotic genomes, has a weakness in assembling repetitive sequences, parts of genomes are often missing (Figure 5(a)). For example, the well-studied Y-linked gene encoding the Su(Ste) piRNA that is essential for male fertility was not present in the assembled Y chromosome sequence until the latest fly genome assembly was released very recently.[79–81]



(e)

Chemical structures at termini	Examples	Behavior in small RNA libraries
5′ and 3′ ends compatible with linker ligation P-NNNNNNNNNNNNNNNNNNNNNNNNN-OH	miRNA, piRNA endo-siRNA (mouse, fly and some worm siRNA)	Enriched
5′ end incompatible with linker ligation PPP-NNNNNNNNNNNNNN.....	Some worm endo-siRNA	Depleted
3′ end incompatible with linker ligation .....NNNNNNNNNNNNNNNNNN>P	piR-IL, 5′-SHOT-RNA	

FIGURE 5 | Limitations in RNA-seq analyses. (a, b) Examples of artificial removal of RNA-seq reads by inaccurate reference genome sequences. Reference genomic sequences may have deletions (a) or single-nucleotide errors/variants (b). These may cause the loss of useful RNA-seq reads. (c, d) Examples of RNA-seq reads with posttranscriptionally modified RNA species. RNA can undergo tailing (untemplated nucleotide additions at 3′ ends, shown in (c)) or editing (adenosine deamination resulting in conversion to inosine, shown in (d)). Reads corresponding to modified RNA species are often removed during genome mapping because these reads do not perfectly match with the reference genome sequence. (e) Enrichment of small RNAs with 5′-monophosphate and 3′-hydroxyl groups. Because most commonly used protocols for small RNA library construction depend on the ligation of 5′- and 3′-linkers, the resulting libraries will be enriched in small RNAs with 5′-monophosphate and 3′-hydroxyl groups. This includes most known small regulatory RNA families. However, there are known RNA species lacking compatible structures at the 5′ or 3′ terminus, such as some worm endo-siRNAs (with 5′ triphosphate) and piR-ILs (with 2′–3′ cyclic-phosphate). These species are depleted in regular small RNA libraries.

© 2015 Wiley Periodicals, Inc.

The miRNA gene annotation studies often highlight the incompleteness of genome assemblies. For example, the highly conserved and biologically/clinically important miRNA *mir-155* was not present in older versions of the rat genome assembly, although the presence of this gene in rats was experimentally verified.[82] Similar artificial losses of conserved miRNA genes in the genome sequence have commonly been observed in comparative genomics studies. A recent study of miRNA evolution in *Drosophilids* identified >20 conserved miRNA genes that were present only in the trace DNA sequencing data and not in the assembled genome sequences of the 12 *Drosophilids*.[83] In addition, PCR assays detected four additional miRNA genes that were even absent in the trace data. Missing genome sequences are not the only problems. The same study also identified single-nucleotide errors/polymorphisms within the miRNA stem regions that could be 'corrected' using the trace sequence data. As miRNA gene annotation studies usually require small RNA reads to be mapped perfectly to the reference genome, even a single nucleotide difference could cause a substantial loss of genome mapping reads (Figure 5(b)).

Even with a correct reference sequence, RNA sequences could differ from the genome sequence due to posttranscriptional modifications of RNA molecules (Figure 5(c) and (d)).[84,85] Such modifications include 3′ nucleotide additions (also called 'tailing') and adenosine deamination (also called 'editing'). Tailing and editing occur efficiently on certain RNA species. Tailing of maternal miRNA (with adenosines) and mirtrons (with uridines) occurs efficiently in insects and vertebrates, with some species tailed at nearly 100%.[85,86] RNA editing also occurs very efficiently. Early RNA of the mouse polyoma virus is highly edited, with ~50% of adenosines modified to inosine at the late stage of infection.[87] Sequencing reads of highly edited RNA molecules would be discarded as 'unmapped reads' and would not be found in bioinformatics analyses unless such modified molecules are specifically analyzed.

In addition, RNA-seq library construction protocols often deplete RNA molecules with certain chemical structures to enrich known functional RNA classes. Most commonly used small RNA sequencing protocols are highly selective for RNAs containing 5′-monophosphate, 3′-hydroxyl groups (Figure 5(e)). The efficient enrichment of miRNAs and related RNA classes by such methods was extremely useful for the identification of small regulatory RNAs. However, functional RNA species could be lost if they do not have chemical structures that are compatible with the library construction method. Worm endogenous siRNAs (endo-siRNAs), a significant class of worm-specific small RNAs that regulate genes and transposons, were missed in earlier studies due to their atypical chemical structure carrying a 5′ triphosphate group.[88] The intermediate precursors of PIWI-interacting RNAs (piRNA-Intermediate-Like molecules: piR-ILs) and 5′-SHOT-RNAs (Sex HOrmone-dependent TRNA-derived RNAs) are thought to have their 3′ ends blocked by phosphate groups.[89,90] Small RNAs with atypical structures in other species may remain undiscovered. There is also the 'black hole of RNA biology', which refers to RNA molecules of 50–300 nt.[91] Deep sequencing analysis has simply not been extensively applied to RNA molecules within this size range.

Therefore, the use of various cloning methods is vitally important to completely catalog all functional RNAs. This is not restricted to small RNA studies. Although the majority of RNA-seq libraries are generated from polyA(+)-enriched RNAs, polyA (+) selection depletes a number of ncRNA species as discussed above (Box 1). Therefore, recent studies have started employing rRNA depletion to identify a wider range of transcripts, including the isolation of polyA(−) RNAs and RNase R-resistant RNAs, resulting in discoveries of new regulatory RNA classes to examine specific classes of ncRNAs.[36,37,40,70]

The selection of biological materials is also important. The discovery of sisRNAs in the *Xenopus* oocyte nucleus provides a good example of the importance of using a suitable biological material to study a phenomenon. The fact that sisRNAs are low abundance and likely to be confused as pre-mRNAs in deep sequencing experiments make the *Xenopus* oocyte nucleus a unique system for the identification of sisRNAs. The *Xenopus* oocyte is a long-lived cell that accumulates stable RNAs for several months, providing a natural environment to enrich for stable and low abundance RNAs. Are there other biological systems that are transcriptionally quiescent and capable of storing RNAs over long periods of time? The *Drosophila* oocyte appears to be one such place. The male gametophyte of the four leaf clover *Marsilea vestita* stores an interesting class of intron-retaining transcripts (IRTs) for future development.[92] One may also consider certain cell types of hibernating animals to identify stable RNA molecules in quiescence cells.

CONCLUSION

After examining a myriad of ncRNAs, we can ask why there are so many ncRNAs and if they really are

© 2015 Wiley Periodicals, Inc.

functional. We propose to examine the question with an open mind. We are now at a stage of rapid discovery or identification of ncRNAs. An important task in the future will be to better understand the 'hidden language' of ncRNAs. Do they have a common language, or are they as diverse as they appear? Equally exciting and probably unexpected will be the elucidation of the impact of ncRNA activities on organismal phenotypes.

ACKNOWLEDGMENTS

Research in KO's group was supported by the National Research Foundation (NRF2011NRF-NRFF001-042). JWP is supported by the Temasek Life Sciences Laboratory.

REFERENCES

1. Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001, 294:862–864.
2. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001, 294:858–862.
3. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science* 2001, 294:853–858.
4. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993, 75:843–854.
5. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 1993, 75:855–862.
6. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 2014, 15:509–524.
7. Meister G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* 2013, 14:447–459.
8. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014, 42:D68–D73.
9. Chak LL, Okamura K. Argonaute-dependent small RNAs derived from single-stranded, non-structured precursors. *Front Genet* 2014, 5:172.
10. Sunkar R, Zhu JK. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* 2004, 16:2001–2019.
11. Lu C, Jeong DH, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, et al. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc Natl Acad Sci USA* 2008, 105:4951–4956.
12. Zhang H, Maniar JM, Fire AZ. 'Inc-miRs': functional intron-interrupted miRNA genes. *Genes Dev* 2011, 25:1589–1594.
13. Ambros V, Horvitz HR. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* 1984, 226:409–416.
14. Pederson T. Regulatory RNAs derived from transfer RNA? *RNA* 2010, 16:1865–1869.
15. Chak LL, Mohammed J, Lai EC, Tucker-Kellogg G, Okamura K. A deeply conserved, noncanonical miRNA hosted by ribosomal DNA. *RNA* 2015, 21:375–384.
16. Beverley SM, Wilson AC. Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. *J Mol Evol* 1984, 21:1–13.
17. Schlotterer C, Hauser MT, von Haeseler A, Tautz D. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Mol Biol Evol* 1994, 11:513–522.
18. Castellano L, Stebbing J. Deep sequencing of small RNAs identifies canonical and non-canonical miRNA and endogenous siRNAs in mammalian somatic tissues. *Nucleic Acids Res* 2013, 41:3339–3351.
19. Lalaouna D, Carrier MC, Semsey S, Brouard JS, Wang J, Wade JT, Masse E. A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol Cell* 2015, 58:393–405.
20. Berget SM, Moore C, Sharp PA. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* 1977, 74:3171–3175.
21. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 1977, 12:1–8.
22. Sharp PA, Konarksa MM, Grabowski PJ, Lamond AI, Marciniak R, Seiler SR. Splicing of messenger RNA precursors. *Cold Spring Harb Symp Quant Biol* 1987, 52:277–285.
23. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 2007, 8:209–220.

© 2015 Wiley Periodicals, Inc.

24. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. *Mol Cell* 2007, 28:328–336.
25. Brown JW, Marshall DF, Echeverria M. Intronic noncoding RNAs and splicing. *Trends Plant Sci* 2008, 13:335–342.
26. Kim YK, Kim VN. Processing of intronic microRNAs. *EMBO J* 2007, 26:775–783.
27. Leverette RD, Andrews MT, Maxwell ES. Mouse U14 snRNA is a processed intron of the cognate hsc70 heat shock pre-messenger RNA. *Cell* 1992, 71:1215–1221.
28. Liu J, Maxwell ES. Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res* 1990, 18:6565–6571.
29. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 2007, 130:89–100.
30. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. *Nature* 2007, 448:83–86.
31. Tycowski KT, Shu MD, Steitz JA. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes Dev* 1993, 7:1176–1190.
32. Gardner EJ, Nizami ZF, Talbot CC Jr, Gall JG. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev* 2012, 26:2550–2559.
33. Qian L, Vu MN, Carter M, Wilkinson MF. A spliced intron accumulates as a lariat in the nucleus of T cells. *Nucleic Acids Res* 1992, 20:5345–5350.
34. Clement JQ, Qian L, Kaplinsky N, Wilkinson MF. The stability and fate of a spliced intron from vertebrate cells. *RNA* 1999, 5:206–220.
35. Clement JQ, Maiti S, Wilkinson MF. Localization and stability of introns spliced from the Pem homeobox gene. *J Biol Chem* 2001, 276:16919–16930.
36. Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL. Long noncoding RNAs with snoRNA ends. *Mol Cell* 2012, 48:219–230.
37. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. Circular intronic long noncoding RNAs. *Mol Cell* 2013, 51:792–806.
38. Zheng S, Vuong BQ, Vaidyanathan B, Lin JY, Huang FT, Chaudhuri J. Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* 2015, 161:762–773.
39. Michaeli T, Pan ZQ, Prives C. An excised SV40 intron accumulates and is stable in *Xenopus laevis* oocytes. *Genes Dev* 1988, 2:1012–1020.
40. Talhouarne GJ, Gall JG. Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *RNA* 2014, 20:1476–1487.
41. Kopczynski CC, Muskavitch MA. Introns excised from the Delta primary transcript are localized near sites of Delta transcription. *J Cell Biol* 1992, 119:503–512.
42. Lu Z, Filonov GS, Noto JJ, Schmidt CA, Hatkevich TL, Wen Y, Jaffrey SR, Matera AG. Metazoan tRNA introns generate stable circular RNAs in vivo. *RNA* 2015, 21:1554–1565.
43. Farrell MJ, Dobson AT, Feldman LT. Herpes simplex virus latency-associated transcript is a stable intron. *Proc Natl Acad Sci USA* 1991, 88:790–794.
44. Wu TT, Su YH, Block TM, Taylor JM. Evidence that two latency-associated transcripts of herpes simplex virus type 1 are nonlinear. *J Virol* 1996, 70:5962–5967.
45. Zabolotny JM, Krummenacher C, Fraser NW. The herpes simplex virus type 1 2.0-kilobase latency-associated transcript is a stable intron which branches at a guanosine. *J Virol* 1997, 71:4199–4208.
46. Moss WN, Steitz JA. Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a novel stable intronic sequence RNA. *BMC Genomics* 2013, 14:543.
47. Kulesza CA, Shenk T. Human cytomegalovirus 5-kilobase immediate-early RNA is a stable intron. *J Virol* 2004, 78:13182–13189.
48. Kulesza CA, Shenk T. Murine cytomegalovirus encodes a stable intron that facilitates persistent replication in the mouse. *Proc Natl Acad Sci USA* 2006, 103:18302–18307.
49. Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, Steitz JA. Viral noncoding RNAs: more surprises. *Genes Dev* 2015, 29:567–584.
50. Cliffe AR, Garber DA, Knipe DM. Transcription of the herpes simplex virus latency-associated transcript promotes the formation of facultative heterochromatin on lytic promoters. *J Virol* 2009, 83:8182–8190.
51. Inman M, Perng GC, Henderson G, Ghiasi H, Nesburn AB, Wechsler SL, Jones C. Region of herpes simplex virus type 1 latency-associated transcript sufficient for wild-type spontaneous reactivation promotes cell survival in tissue culture. *J Virol* 2001, 75:3636–3646.
52. Perng GC, Jones C, Ciacci-Zanella J, Stone M, Henderson G, Yukht A, Slanina SM, Hofman FM, Ghiasi H, Nesburn AB, et al. Virus-induced neuronal apoptosis blocked by the herpes simplex virus latency-associated transcript. *Science* 2000, 287:1500–1503.
53. Diener TO. Viroids: structure and function. *Science* 1979, 205:859–866.
54. Crick F. Split genes and RNA splicing. *Science* 1979, 204:264–271.
55. Diener TO. Are viroids escaped introns? *Proc Natl Acad Sci USA* 1981, 78:5014–5015.

© 2015 Wiley Periodicals, Inc.

56. Dinter-Gottlieb G. Viroids and virusoids are related to group I introns. *Proc Natl Acad Sci USA* 1986, 83:6250–6254.

57. Lewin R. Viroids may be escaped introns. *Science* 1986, 233:1384.

58. Taylor J, Pelchat M. Origin of hepatitis delta virus. *Future Microbiol* 2010, 5:393–402.

59. Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* 1993, 73:1019–1030.

60. Cocquerelle C, Mascrez B, Hetuin D, Bailleul B. Mis-splicing yields circular RNA molecules. *FASEB J* 1993, 7:155–160.

61. Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B. Scrambled exons. *Cell* 1991, 64:607–613.

62. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* 2010, 6:e1001233.

63. Zaphiropoulos PG. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol Cell Biol* 1997, 17:2985–2993.

64. Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ, Kjems J. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J* 2011, 30:4414–4422.

65. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013, 495:333–338.

66. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. *Nature* 2013, 495:384–388.

67. Rybak-Wolf A, Stottmeister C, Glazar P, Jens M, Pino N, Giusti S, Hanan M, Behm M, Bartok O, Ashwal-Fluss R, et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 2015, 58:870–885.

68. Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* 2014, 28:2233–2247.

69. Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ. The RNA binding protein quaking regulates formation of circRNAs. *Cell* 2015, 160:1125–1134.

70. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell* 2014, 159:134–147.

71. Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, Lai EC. Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* 2014, 9:1966–1980.

72. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet* 2013, 9:e1003777.

73. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014, 56:55–66.

74. Wang PL, Bao Y, Yee MC, Barrett SP, Hogan GJ, Olsen MN, Dinneny JR, Brown PO, Salzman J. Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* 2014, 9:e90859.

75. Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon EY, Landthaler M, Dieterich C, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* 2015, 10:170–177.

76. Barrett SP, Wang PL, Salzman J. Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *Elife* 2015, 4:e07540.

77. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* 2011, 12:99–110.

78. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L, et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 2015, 22:256–264.

79. Livak KJ. Detailed structure of the *Drosophila melanogaster* stellate genes and their transcripts. *Genetics* 1990, 124:303–316.

80. Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol* 2001, 11:1017–1027.

81. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res* 2015, 25:445–458.

82. Uva P, Da Sacco L, Del Corno M, Baldassarre A, Sestili P, Orsini M, Palma A, Gessani S, Masotti A. Rat mir-155 generated from the lncRNA Bic is 'hidden' in the alternate genomic assembly and reveals the existence of novel mammalian miRNAs and clusters. *RNA* 2013, 19:365–379.

© 2015 Wiley Periodicals, Inc.

83. Mohammed J, Flynt AS, Siepel A, Lai EC. The impact of age, biogenesis, and genomic clustering on *Drosophila* microRNA evolution. *RNA* 2013, 19:1295–1308.

84. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, et al. MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res* 2013, 41: D262–D267.

85. Lee M, Kim B, Kim VN. Emerging roles of RNA modification: m(6)A and U-tail. *Cell* 2014, 158:980–987.

86. Westholm JO, Ladewig E, Okamura K, Robine N, Lai EC. Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. *RNA* 2012, 18:177–192.

87. Kumar M, Carmichael GG. Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc Natl Acad Sci USA* 1997, 94:3542–3547.

88. Billi AC, Fischer SE, Kim JK. Endogenous RNAi pathways in *C. elegans*. *WormBook* 2014:1–49.

89. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* 2010, 24:2493–2498.

90. Honda S, Loher P, Shigematsu M, Palazzo JP, Suzuki R, Imoto I, Rigoutsos I, Kirino Y. Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proc Natl Acad Sci USA* 2015, 112:E3816–E3825.

91. Steitz J. RNA-RNA base-pairing: theme and variations. *RNA* 2015, 21:476–477.

92. Boothby TC, Zipper RS, van der Weele CM, Wolniak SM. Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. *Dev Cell* 2013, 24:517–529.

© 2015 Wiley Periodicals, Inc.