Trends in Genetics

CellPress

Review

The Diversity of Long Noncoding RNAs and Their Generation

Huang Wu,[1] Li Yang,[2,3,*] and Ling-Ling Chen[1,3,*]

Long noncoding RNAs (lncRNAs) are emerging as potential key regulators in gene expression networks and exhibit a surprising range of shapes and sizes. Several distinct classes of lncRNAs are transcribed from different DNA elements, including promoters, enhancers, and intergenic regions in eukaryotic genomes. Additionally, others are derived from long primary transcripts with noncanonical RNA processing pathways, generating new RNA species with unexpected formats. These lncRNAs can be processed by several mechanisms, including ribonuclease P (RNase P) cleavage to generate mature 3′ ends, capping by small nucleolar RNA (snoRNA)–protein (snoRNP) complexes at their ends, or the formation of circular structures. Here we review current knowledge on lncRNAs and highlight the most recent discoveries of the underlying mechanisms related to their formation.

Eukaryotic RNA Transcription and Processing Yield a Diverse Catalog of lncRNAs

Recent large-scale RNA profiling efforts have revealed that >75% of the human genome is actively transcribed to yield a highly complex network of protein-coding transcripts (or mRNAs) and noncoding RNAs (ncRNAs) [1,2]. In eukaryotic cells, DNA transcription and RNA processing are crucial steps for the biogenesis and function of all RNA species. For example, it is well established that nascent precursor mRNA (pre-mRNA) processing is tightly connected to RNA polymerase II (Pol II) transcription [3]. On transcription, numerous pre-mRNA processing events including capping, splicing, cleavage/polyadenylation (C/P), export, and surveillance are seamlessly integrated to ensure the functionality of mRNA maturation [4]. 7-Methyl guanosine (m^7G) capping at the 5′ end occurs during the initiation phase of Pol II transcription. 3′ End maturation of nearly all Pol II-transcribed RNAs is connected to Pol II termination and involves the cleavage of nascent transcripts followed by the addition of poly(A) tails to their ends. The m^7G cap and 3′ poly(A) are hallmark structures necessary for the stabilization and function of eukaryotic mRNAs.

While protein-coding genes occupy only a small portion of the mammalian genome, transcriptomic analyses have unveiled the widespread occurrence of lncRNAs. lncRNAs comprise a wide variety of ncRNA species of size greater than 200 nucleotides (nt) that lack significant protein-coding capacity [5]. Emerging lines of evidence have shown that lncRNAs are key regulators of gene expression at both the transcriptional and the post-transcriptional level in diverse cellular contexts and biological processes [6,7]. In general, lncRNAs can be subdivided into several classes based on their positional relationship to protein-coding genes and different mechanisms of processing (Figure 1, Key Figure). On the one hand, different classes of lncRNA transcripts [i.e., promoter upstream transcripts (PROMPTs), enhancer RNAs (eRNAs), long intervening/intergenic ncRNAs (lincRNAs), and natural antisense transcripts (NATs)] have been shown to be transcribed

Trends

Eukaryotic DNA transcription and RNA processing yield a diverse catalog of long noncoding RNAs (lncRNAs) that are longer than 200 nucleotides and lack significant protein-coding potential.

lncRNAs transcribed from promoters and enhancers are usually targeted by nuclear exosomes and have short half-lives.

Although they have a 7-methyl guanosine (m^7G) cap and 3′ poly(A) at their ends, the mRNA-like long intervening/intergenic ncRNAs (lincRNAs) have patterns of transcription and processing distinct from those of mRNAs.

New types of linear lncRNA species are stabilized by various mechanisms, including the processing of 3′ ends by endoribonucleases, of 5′ ends by small nucleolar RNA–protein (snoRNP) caps, or of both ends by snoRNP protection.

Circular RNAs represent yet another new type of lncRNA that is processed from back-spliced exons or spliced intron lariats of RNA polymerase II-transcribed RNA precursors.

[1]State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences; University of Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China
[2]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences,

from promoter upstream regions, **enhancers** (see Glossary), intergenic regions, and the opposite strand of protein-coding genes, respectively, in eukaryotic genomes (Figure 1). On the other hand, many other lncRNAs are generated from long primary transcripts with unusual RNA processing pathways, resulting in new RNA species with unexpected structures (Figure 1). For instance, rather than using canonical 5′-end m^7G capping or 3′-end poly (A) tailing for maturation, stabilization of lncRNAs can be achieved by several noncanonical mechanisms, including RNase P cleavage to generate a mature 3′ end [8,9], capping by **snoRNPs** at both ends [10–12] or the 5′ end [13], or forming circular structures to protect them from degradation [14–19]. In this review we highlight the most recent discoveries relating to lncRNA diversity and the mechanisms of their biogenesis.

Chinese Academy of Sciences; University of Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China
[3]School of Life Science and Technology, ShanghaiTech University, 100 Haike Road, Shanghai 201210, China

*Correspondence:
liyang@picb.ac.cn (L. Yang) and
linglingchen@sibcb.ac.cn (L.-L. Chen).

Short-Lived lncRNAs Produced from Upstream Regions of Genes and Enhancers

Promoters and enhancers are essential DNA elements in the control of gene expression networks. Pol II transcription at promoter upstream regions and enhancers produces short-lived medium-length lncRNAs, usually ranging from 200 to 2000 nt [20,21]. These RNAs are targeted by the nuclear **RNA exosome** and have rapid turnover rates, challenging their functional significance in gene regulation.

PROMPTs

PROMPTs are transcribed in the antisense orientation, approximately 0.5–2.5 kb upstream of the active transcription start sites (TSSs) of most protein-coding genes in mammals [22,23] (Figure 1). They were originally discovered by tiling microarray using cells that had exonucleolytic RNA exosome components deleted to eliminate RNA degradation [23]. Similar transcripts are called cryptic unstable transcripts (CUTs) in *Saccharomyces cerevisiae* [24]. Of note, CUTs and PROMPTs are different from TSS-associated (Tssa-) [25] and transcription initiation (ti-) [26] RNAs, which are subsets of the RNA population of 20–90 nt in length and thought to be byproducts of Pol II stalling and backtracking.

Regions of DNA containing PROMPT transcription units are occupied by Pol II complexes containing serine 2-phosphorylated (S2P) C-terminal domains (CTDs), mimicking the associated genic region [20]. The resulting PROMPTs are heterologous in length (about 200–600 nt) and carry 5′-cap structures and 3′ adenosine tails [20]. These RNAs are largely retained in the nucleus and undergo rapid degradation by the RNA nuclear exosome targeting (NEXT) complex (degradation is from 3′ to 5′) [20,27]. Whether PROMPTs are functional remains mysterious. It was reported that PROMPT levels could be altered under stress conditions and their accumulation by exosome depletion was found to influence transcription factor binding to promoters, suggesting that these short-lived transcripts may have some regulatory role [28]. Besides lacking apparent functional significance, the rapid degradation of transcribed PROMPTs, but not of their neighboring promoter-downstream mRNAs, has been linked to the choice of promoter directionality [29].

eRNAs

eRNAs are usually less than 2000 nt in length and bidirectionally transcribed from enhancers by Pol II, with the two directions producing roughly equivalent levels of RNA [21,30,31] (Figure 1). Several reports have suggested that eRNAs have enhancer-like function in gene regulatory networks by controlling promoter and enhancer interactions and the topology of higher-order chromatin structure [32–34]. However, it has also been reported that knockdown of eRNAs did not inhibit the function of enhancers in multiple cases [35]. Thus, the question of whether most eRNAs are functional remains elusive.

eRNAs lack poly(A) tails and several studies have provided insights into their biogenesis. Integrator – a complex that has a core catalytic RNA endonuclease activity and is known to be required for 3′-end processing of non-polyadenylated small nuclear RNA genes – is involved in the 3′-end cleavage of eRNA primary transcripts [36]. Depletion of Integrator resulted in the accumulation of primary eRNA transcripts bound to Pol II [36]. Similar to PROMPTs, eRNAs are targets of the exosome, which appears to occur when they are released from Pol II [27]. Interestingly, while depletion of exosome components led to increases of both PROMPTs and eRNAs, eRNA increases appeared to be much higher [30]. In addition, **individual-nucleotide-resolution crosslinking and immunoprecipitation (iCLIP)** experiments have shown that the exosome RNA-binding protein (RBP) RBM7 is broadly associated with Pol II-derived RNAs, including pre-mRNAs as well as PROMPTs and eRNAs [27].

It should be noted that enhancers/promoters have interchangeable roles depending on tissue specificity. Enhancers have been described as resembling promoters for the production of cell-type-specific transcripts [37,38]. Promoters in one tissue were also predicted to be enhancers in other tissues [38]. These features of promoters and enhancers may in part explain the similarities seen in the processing of PROMPTs and eRNAs. In this context, both DNA elements can be regarded as sites of transcriptional initiation that are characterized by the types of transcripts they produce. Interestingly, RNAs transcribed from promoter-proximal and −distal enhancers remain bound to chromatin and contribute to stable transcription factor [such as Yin Yang 1 (YY1)] occupancy at these sites to maintain gene expression programs [39]. Some gene (both coding and noncoding) promoters have also been proposed to function as enhancers regulating neighboring gene expression; transcription and transcripts of these genes thus contribute to this regulation by recruiting activating factors or remodeling nucleosomes [40].

Synthesis and Turnover of lincRNAs

lincRNAs are transcribed by Pol II from intergenic regions between two genes and represent the best-studied subclass of lncRNAs (Figure 1). The functional significance of lincRNAs has been well summarized in recent reviews (for examples see [6,7]). Most annotated lincRNAs contain multiple exons and have typical mRNA-like features, with a 5′ m^7G cap and a 3′ poly(A) tail [41,42]. These similarities suggest that the processing of lincRNAs can presumably mimic that of mRNAs, leading to the assumption that mature lincRNAs may behave similarly to mRNAs in cells. However, this is not the case, and lincRNAs have their own characteristics. lincRNAs lack robust protein-coding potential; they are less evolutionarily conserved and less abundant, and exhibit more tissue-specific expression [5,41]; they exhibit greater nuclear localization than their mRNA counterparts [1]; and their functions are highly associated with their specific subcellular localization patterns [6]. Of note, similar mRNA-like lncRNAs transcribed from the opposite strand of protein-coding genes are called NATs [43] (Figure 1).

Early genome-wide analyses indicated that lincRNA loci appear similar to protein-coding genes at the chromatin level. This led to the systematic discovery of lncRNAs by monitoring the patterns of histone 3 Lys 4 trimethylation (H3K4me3) in the promoter region followed by histone 3 Lys 36 trimethylation (H3K36me3) across the actively transcribed intergenic regions in mammalian cells [42,44]. Although lincRNAs contain fewer exons than mRNAs and often have weak cryptic splicing and polyadenylation signals [41,45], these observations could not adequately address the differences between lincRNA and mRNAs.

Two recent studies have begun to dissect patterns of lincRNA transcription and processing distinct from those of mRNAs [46,47]. One study compared features related to pre- and post-transcriptional regulation of lincRNAs with those of mRNAs with similar expression levels in several human cell lines. It was found that lincRNAs differ from mRNAs in a couple of aspects [46]. First, lincRNAs in general have fewer histone marks and transcription factors bound to their

Glossary

4sUDRB-seq: captures newly transcribed RNAs, based on the reversible inhibition of transcription with 5,6-dichloro-1-β-d-ribofuranosylbenzimidazole (DRB) and pulse labeling with the uridine analog 4sU after DRB removal and high-throughput sequencing. It has been used to measure genome-wide transcription elongation rates and for the characterization of nascent RNAs [96].

Cajal body (CB): a type of membraneless nuclear subcompartment in eukaryotic cells. CBs contain specific protein and RNA components and play important roles in RNA-related metabolic processes including transcription, splicing, ribosome biogenesis, and telomere maintenance.

Enhancers: regulatory DNA sequences that are capable of binding master transcription factors/mediators and often form long-range chromatin loops with their target genes to activate temporally and tissue-specific gene expression independent of their proximity or orientation to their target genes.

Individual-nucleotide-resolution cross-linking and immunoprecipitation (iCLIP): a method using UV light to covalently crosslink proteins and RNA molecules and immunoprecipitation to identify protein–RNA interactions.

Native elongation transcript sequencing in mammalian cells (mNET-seq): generates single-nucleotide-resolution genome-wide profiles of nascent RNA and cotranscriptional RNA processing associated with different phosphorylation states of the CTD of the largest subunit of Pol II [47].

Promoters: regions of DNA that initiate the transcription of genes and are located near the TSSs of genes upstream on the same strand of DNA.

RNA exosome: an evolutionarily conserved RNA degradation complex with both 3′ → 5′ exonucleolytic and endonucleolytic activity.

Small nucleolar RNA (snoRNA)–protein (snoRNP): ribonucleoprotein complexed with snoRNA; required for snoRNA stabilization and function.

XRN2: an evolutionarily conserved 5′ → 3′ exoribonuclease that is predominantly localized in the

promoters than mRNAs. However, H3K9me3, a histone modification typically associated with transcriptional repression, is somehow more enriched at promoters of active lincRNA loci than at those of active mRNAs. Such H3K9me3-marked lincRNA genes are more tissue specific. Second, lincRNAs are less efficiently spliced, probably owing to the weaker internal splicing signals and the lower U2AF65 binding in lincRNAs than in mRNAs. Of note, U2AF65 is one important factor in splice site (ss) determination in pre-mRNA splicing [48]. Interestingly, although the abundance of lincRNAs is generally lower than that of mRNAs, their stabilities are comparable for expression-matched groups.

nucleus and recognizes single-stranded RNA with a 5′-terminal monophosphate to degrade it processively to mononucleotides.

Key Figure

A Diverse Catalog of Long Noncoding RNAs (lncRNAs) Yielded from Eukaryotic RNA Transcription and Processing

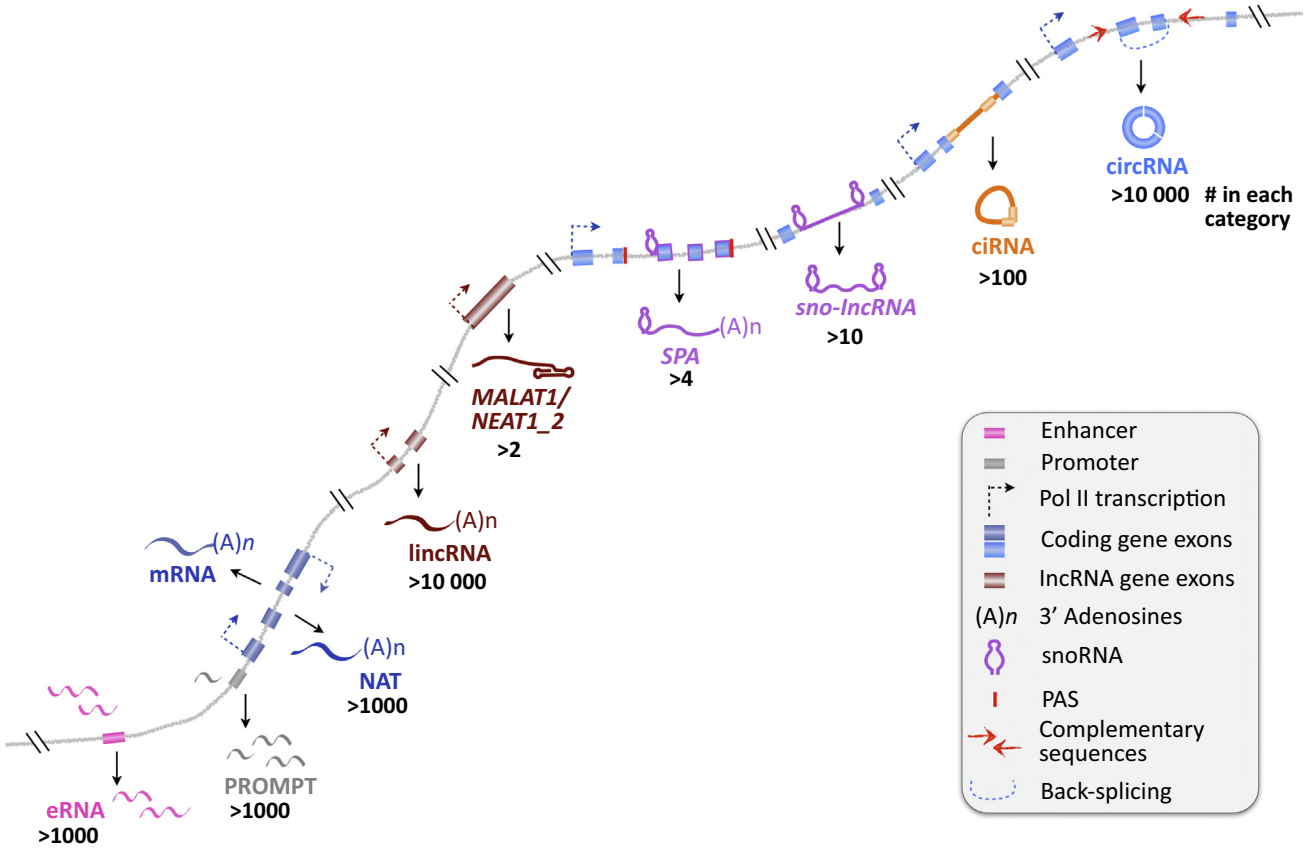


Figure 1. A schematic drawing to illustrate the diversity of lncRNAs in mammalian genomes. ciRNA, circular intronic RNA derived from intron lariats; circRNA, circular RNA produced from back-splicing of exons; eRNA, enhancer RNA; lincRNA, large intervening/intergenic noncoding RNA; NAT, natural antisense transcript; PROMPT, promoter upstream transcript; sno-lncRNA, small nucleolar RNA (snoRNA)-ended lncRNA; SPA, 5′ snoRNA-ended and 3′-polyadenylated lncRNA; metastasis-associated lung adenocarcinoma transcript 1 (MALAT1)/Nuclear enriched abundant transcript 1, the long isoform (NEAT1_2), lncRNA with the 3′ end alternatively processed by ribonuclease P (RNase P). The number of lncRNAs identified in each category in mammals is shown beneath

In another study lincRNA synthesis and turnover was characterized based on a range of nascent transcriptomic approaches, including **native elongation transcript sequencing in mammalian cells (mNET-seq)** [47]. It was found that lincRNAs and pre-mRNAs are transcribed by different Pol II phospho-CTD isoforms. It is known that pre-mRNAs are transcribed by defined Pol II isoforms that associate with the spliceosome through a serine 5 P (S5P) CTD, and mRNA 3′ ends are generated cotranscriptionally by CPSF73 cleavage as part of C/P processing, which in turn promotes Pol II termination [3]. This is in contrast to most lincRNAs, which are mainly transcribed by deregulated Pol II, weakly spliced and polyadenylated, and simultaneously degraded by the nuclear exosome on chromatin. Interestingly, the termination of many lincRNA genes appears to be CPSF73 independent. For example, the threonine 4-phosphorylated (T4P) CTD mark that correlates with protein-coding gene termination is distributed more evenly across the gene body of lincRNAs. This is similar to *S. cerevisiae*, where the cellular fate of lncRNA and mRNA transcripts is largely determined during 3′ end formation before the acquisition of export competence [49].

Together these studies have identified unique patterns of transcription and processing for human lincRNAs. These features can somehow explain why many lincRNAs remain bound to chromatin and are expressed at low levels. Of note, however, functional lincRNAs must escape from this targeted nuclear surveillance process [47] to accumulate to high levels in specific cell types.

New lncRNA Species Generated by Unique Biogenesis Pathways

While a large proportion of lncRNAs look like mRNAs, various lncRNAs are processed from long primary transcripts to yield mature lncRNAs without 5′-cap structures or 3′ adenosine tails (Figure 1). The generation of these lncRNAs is highly associated with eukaryotic RNA processing.

lncRNAs with 3′ Ends Alternatively Processed by Endoribonucleases

The metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) and multiple endocrine neoplasia β (Menβ) (also called Nuclear enriched abundant transcript 1, the long isoform NEAT1_2) lncRNAs each accumulate to high levels in nuclear bodies called splicing speckles or paraspeckles, where they play critical roles in cancer progression and the formation of nuclear paraspeckles, respectively [50]. They are processed at their 3′ ends not by canonical C/P but by the recognition and cleavage of tRNA-like structures by RNase P [8,9] (Figure 2A). RNase P is an endoribonuclease and is best known for its function in tRNA maturation [51].

RNase P cleavage leads to the formation of mature 3′ ends of lncRNAs, which are subsequently protected by a conserved, stable U-A·U triple-helical RNA structure (· denotes the Hoogsteen face and - denotes the Watson–Crick face) [52,53]. A similar triple-helical structure, called a nuclear retention element or element for nuclear expression (ENE), has also been found at the 3′ end of the PAN lncRNA, which is expressed by Kaposi's sarcoma-associated herpesvirus (KSHV), and in RNAs from other viruses [54,55]. However, the viral ENE is not formed by RNase P processing.

In addition to RNase P-mediated 3′ processing of lncRNAs, several lncRNA transcripts containing miRNAs (lnc-pri-miRNAs) use the cleavage by the endonuclease Microprocessor to terminate transcription, preventing transcriptional interference with downstream genes and generating some unstable lncRNAs without 3′-end poly(A) tails [56].

Excised Intron-Derived snoRNA-Ended lncRNAs

Excised introns can produce stable RNAs, although it is generally believed that most introns or intron fragments are unstable [57,58]. For instance, the great majority of snoRNAs are encoded

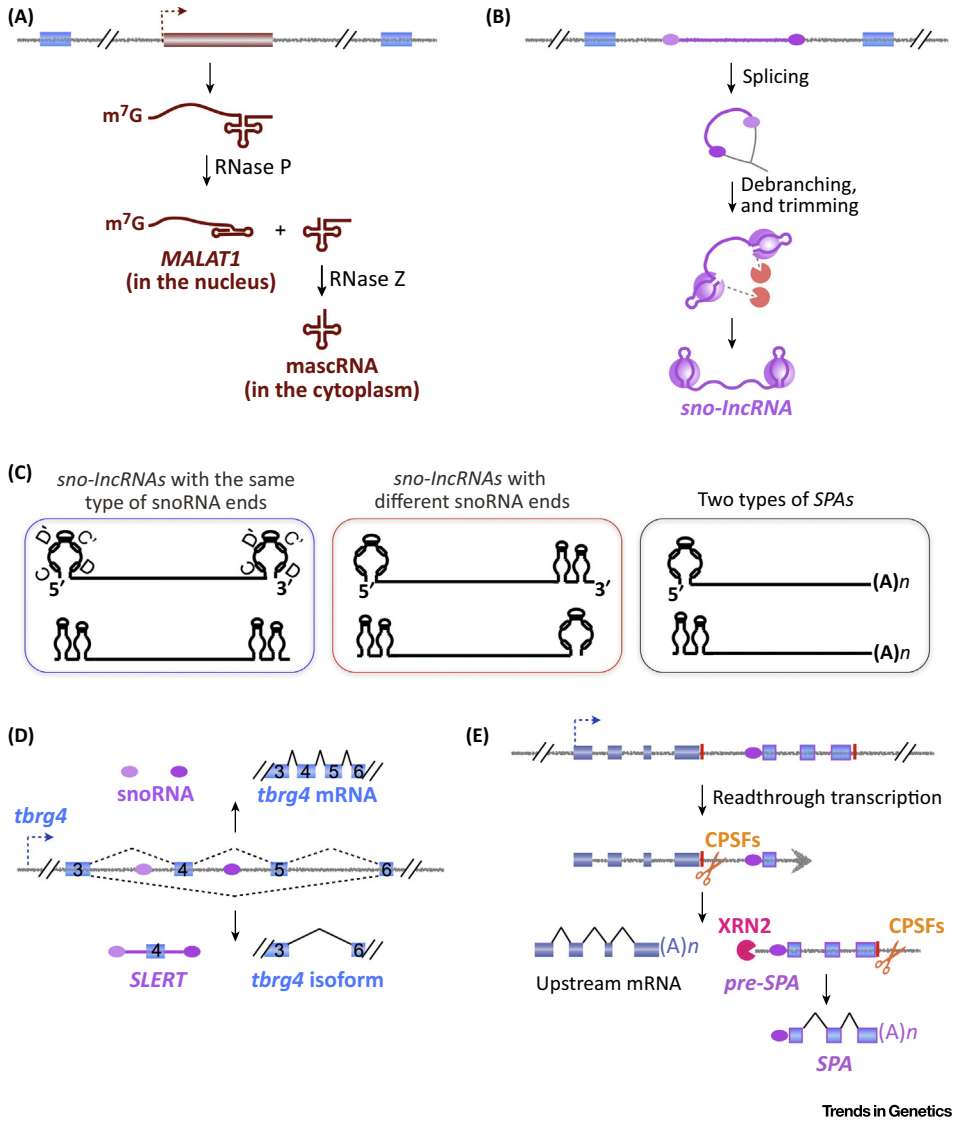


Figure 2. New Long Noncoding RNA (lncRNA) Species Generated from Unusual Processing Pathways. (A) Ribonuclease P (RNase P) processing of the 3′ end of MALAT1. The nascent metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) transcript forms a tRNA-like structure at its 3′ end, which can be recognized and cleaved by RNase P to generate stable MALAT1 with a U-A·U triple-helical structure at the 3′ end. The 3′-end product was further cleaved by RNase Z to form mascRNA, which is ~ 60 nucleotides (nt) in length with unknown function [8,9,52,53]. (B) Processing of small nucleolar RNA (snoRNA)-ended lncRNAs (sno-lncRNAs). sno-lncRNAs are formed when one intron contains two snoRNA genes. During splicing the sequences between the snoRNAs are not degraded, leading to the accumulation of lncRNAs flanked by snoRNA sequences but lacking 5′ caps and 3′ poly(A) tails [10]. (C) The diversity of lncRNAs related to snoRNAs (sno-processed lncRNAs). Four types of sno-lncRNA have been found in mammalian genomes and their ends are both capped by a Box C/D or a Box H/ACA snoRNA protein (snoRNP) complex (blue box) or each capped by one Box C/D or one Box H/ACA snoRNP (red box). Two types of 5′ snoRNA-ended and 3′-polyadenylated lncRNA (SPA) presumably exist in mammalian genomes (black box) [10,12,13,50]. (D) Species- or cell-type-specific expression of sno-lncRNA. One example is shown to illustrate that alternative splicing (AS) leads to two snoRNAs embedded within one intron and therefore sno-lncRNA formation. Of note, SLERT is a Box H/ACA-ended sno-lncRNA [11,12]. (E) Processing of SPA. SPA is derived from readthrough transcripts and its processing is associated with the kinetic competition of XRN2 and Pol II downstream of polyadenylation signals. Following cleavage/polyadenylation of its upstream gene, the downstream uncapped pre-SPA is trimmed by XRN2 until this exonuclease reaches the cotranscriptionally assembled snoRNP. This snoRNP prevents further degradation, generates a novel 5′ end, and allows continuous Pol II elongation [13]. Abbreviation: CPSFs, cleavage/polyadenylation-specific factors

within introns [59]. snoRNAs are a family of conserved nuclear RNAs (about 70–200 nt) that are usually concentrated in **Cajal bodies (CBs)** or nucleoli. They serve as guide RNAs with complementarity to specific target sequences to modify small nuclear RNAs (snRNAs) or rRNAs or participate in pre-rRNA processing during ribosome subunit maturation. They are processed from excised and debranched introns by exonucleolytic trimming and perform their functions in complex with specific protein components by forming snoRNPs [50]. There are several hundred known snoRNAs, which can be grouped into two classes: Box C/D and Box H/ACA.

The exploration of non-polyadenylated and rRNA-depleted 'poly(A)−/ribo−' transcripts in human cells has revealed that many excised introns longer than 200 nt accumulate to high levels [60], leading to the question of how these excised introns escape from exonucleolytic trimming after splicing. Detailed studies have uncovered sno-lncRNAs, which are formed when one intron contains two snoRNA genes. This occurs when the sequences between the snoRNAs are not degraded during splicing, leading to the accumulation of lncRNAs flanked by snoRNA sequences but lacking 5′ caps and 3′ poly(A) tails [10] (Figure 2B). sno-lncRNAs can be capped by both Box C/D [10] and Box H/ACA [12] snoRNPs or, in some cases, capped by one Box C/D or one Box H/ACA snoRNP at each end (Figure 2C). Remarkably, the expression of sno-lncRNA is species or cell type specific, largely as the result of species- or cell-type-specific alternative splicing (AS). In these cases, AS results in two snoRNAs embedded within one intron, leading to sno-lncRNA formation [11] (Figure 2D).

Being protected by snoRNPs at both ends, sno-lncRNAs were found to have longer half-lives than mRNAs with matched expression levels. Interestingly, it was suggested that sno-lncRNAs could be further processed to produce individual snoRNAs [10–12], but how they are processed is unclear. The NEXT complex defines an early exosome targeting pathway acting on newly synthesized RNA, including snoRNAs located in pre-mRNA introns [27], implying that the nuclear surveillance pathway may play a role in sno-lncRNA degradation. Alternatively, some unknown endonuclease may bind to the non-snoRNA sequences of sno-lncRNAs to initiate their degradation.

The abundance of sno-lncRNAs argues that they are unlikely to be precursors of snoRNAs [10–12]. For instance, the genomic region encoding several of the most-abundant Box C/D-ended sno-lncRNAs is specifically deleted in Prader–Willi syndrome (PWS), a neurodevelopmental genetic disorder [10,61]. The minimal microdeletion associated with PWS has been mapped to the 3′ UTR of SNURF-SNRPN at the imprinted region of 15q11–q13 [62–65], but the molecular cause of PWS has remained elusive [61]. Importantly, this critical PWS deletion region encodes multiple sno-lncRNAs of size ranging from 1000 to 3000 nt. These RNAs are expressed at high levels (comparable with some histone mRNAs) in normal human embryonic stem cells (hESCs), accumulate near their sites of synthesis, and interact with the splicing regulator RBFOX2 (see below for details). Intriguingly, although these sno-lncRNAs contain snoRNA ends they do not colocalize with nucleoli or CBs [10], indicating that an unknown mechanism is associated with their unique subcellular localization.

In addition to Box C/D snoRNA-ended PWS-region sno-lncRNAs, SLERT is a Box H/ACA snoRNA-ended lncRNA and has recently been characterized as playing an essential role in rRNA biogenesis (Figure 2D) [12]. SLERT is translocated to the nucleolus depending on its snoRNA ends. In the nucleolus it directly binds to the DEAD-box RNA helicase DDX21 via a 143-nt non-snoRNA sequence. Strikingly, DDX21 molecules in the nucleolus form ~400-nm ring-shaped structures surrounding all RNA Pol I complexes to suppress pre-rRNA transcription. SLERT binding alters the conformation of DDX21 and modulates DDX21 rings to remove their suppression on Pol I, thereby providing a mechanism to control the differential expression

of rDNA clusters [12]. Importantly, it has been established that dysregulated rRNA synthesis by Pol I is associated with uncontrolled cell proliferation. Consistent with this notion, SLERT depletion impairs pre-rRNA transcription and rRNA production, leading to decreased tumorigenesis [12].

These studies together have shown that excised introns can produce lncRNAs that are stabilized by snoRNPs. These sno-lncRNAs impact important cellular processes and functions and their misregulation is highly associated with human diseases. Due to their unique biogenesis pathway, cell-type-specific expression pattern, and functional potential, it will be of great interest to annotate additional sno-lncRNAs and investigate their roles in health and disease contexts.

5′ snoRNA-Ended and 3′-Polyadenylated lncRNAs (SPAs) Are Derived from Polycistronic Transcripts and Coupled to Nascent RNA 3′-End Processing

Distinct from the abovementioned excised, intron-derived sno-lncRNAs, processing of polycistronic transcripts can produce yet another new species of lncRNA that are 5′ snoRNP ended and 3′ polyadenylated, called SPAs [13] (Figure 2E). Two SPAs were reported to be generated from the minimal PWS deletion region, the 3′ UTR of SNURF-SNRPN at the imprinted region of 15q11–q13 [10,13]. PWS-region SPAs contain multiple exons and are 35 000 and 16 000 nt in length, respectively. Using *in vitro* recapitulation assays and modulating Pol II elongation rates *in vivo*, it was shown that the processing of SPAs is highly associated with fast Pol II transcription elongation and the kinetic competition of **XRN2** and Pol II at its upstream SNURF-SNRPN gene termination. Followed by C/P of SNURF-SNRPN, the downstream uncapped pre-SPA is degraded by XRN2 until it reaches the cotranscriptionally assembled snoRNP located several kilobases downstream, which ensures SPA 5′ end formation and continuous Pol II elongation (Figure 2E) [13]. As both Box C/D and Box H/ACA snoRNAs are capable of mediating SPA formation [13], it is possible to identify additional SPAs and polycistronic transcripts of this type in various cells (Figure 2D).

The two PWS-region SPAs also locate to sites of their transcription [13]. Together with five sno-lncRNAs [10], these seven lncRNAs form a $1 \sim 2\text{-}\mu m^3$ nuclear accumulation [13]. These lncRNAs can sequester multiple RBPs including TDP43, RBFOX2, and hnRNP M, which are well known to be involved in many aspects of the regulation of mRNA metabolism. The generation of a human PWS cellular model by depletion of all of these lncRNA genes results in genome-wide altered binding patterns for RBPs and AS. Interestingly, some genes with altered cassette exons in these PWS cell lines were associated with synaptosome and neurotrophin signaling pathways [13]. Thus, these results have implicated PWS-region sno-lncRNAs and SPAs in the molecular pathogenesis of PWS, by acting to sequester multiple RBPs. Future studies will be needed to dissect the distinct roles of sno-lncRNAs and SPAs in the molecular and disease phenotype at this locus by identifying additional proteins, DNAs, and RNAs that are associated with individual lncRNAs in hESCs and neurons.

Circular RNAs (circRNAs) Processed from Pol II-Transcribed RNA Precursors

Besides the aforementioned linear lncRNAs with 5′-to-3′ polarity, circRNAs are produced by distinct mechanisms and feature covalently closed structures [66,67]. Two types of circRNA are reported to be produced from Pol II-transcribed RNA precursors, presumably by the spliceosomal machinery [58,66]. Owing to their non-polyadenylated loop structures, their widespread expression has not been detected by transcriptome profiling from polyadenylated RNAs but was only recently found by non-polyadenylated RNA-seq [10,14–19,60,68].

Circular Intronic RNAs (ciRNAs) Derived from Spliced Intron Lariats

ciRNAs are generated from spliced introns [18]. Specifically, they are produced from excised intron lariats that fail to be debranched after splicing, leading to a covalent circle with a 2′,5′-phosphodiester bond between the 5′ splice donor site and the branch point site (Figure 3A). Unlike lariat RNAs, which usually contain a variety of lengths of 3′ linear tails [69], ciRNAs do not contain 3′ linear appendages [18]. Both bioinformatic and experimental lines of evidence have shown that ciRNA formation depends on a consensus RNA motif containing a 7-nt GU-rich element near the 5′ ss and an 11-nt C-rich element near the branch point [18]. It remains unclear how these *cis* elements function to resist debranching and what other, *trans* factors are involved in this process.

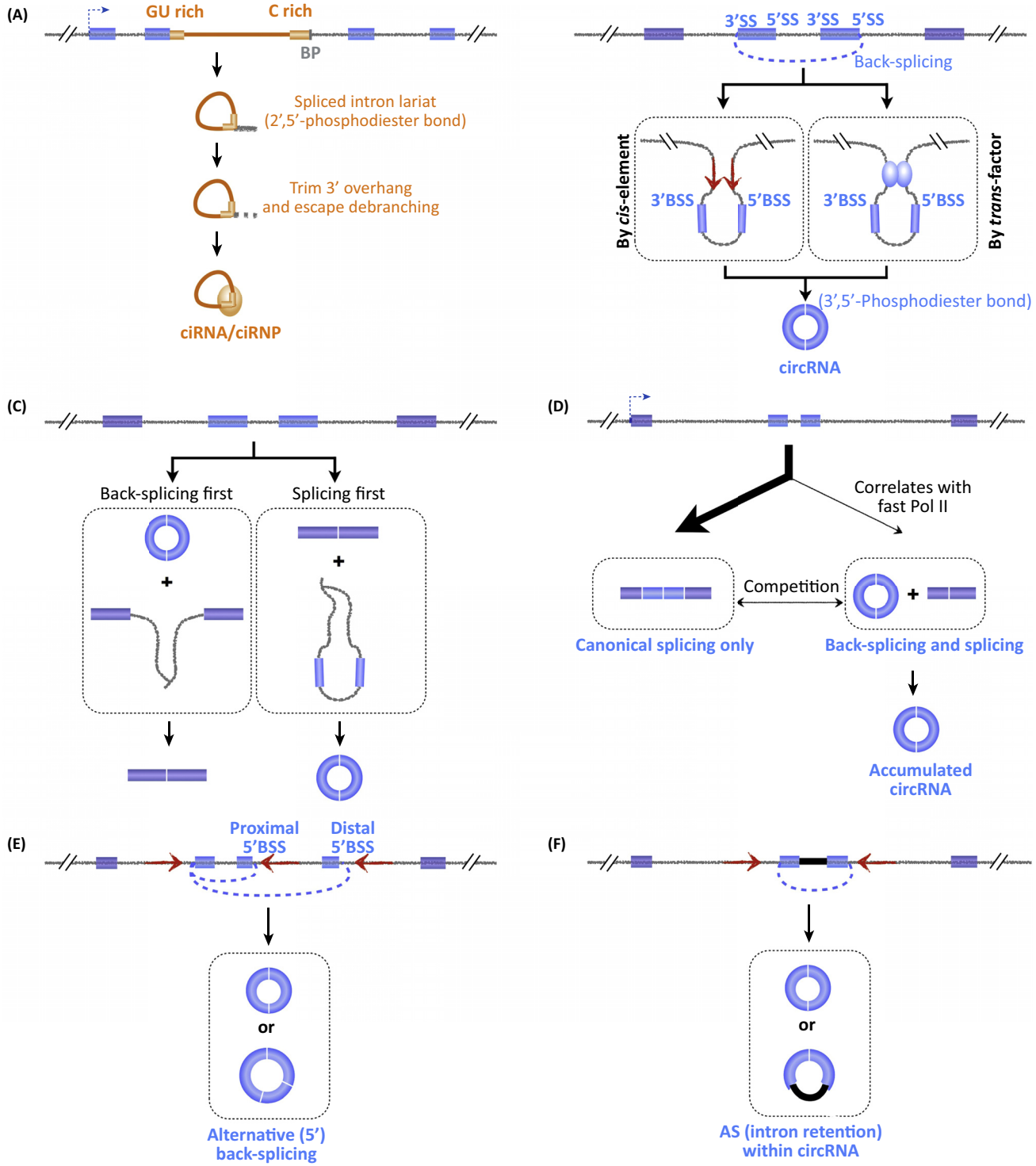
Spliced-lariat-derived human ciRNAs are preferentially localized in the nucleus, while their corresponding linear mRNAs are mainly located in the cytoplasm [18]. Some abundant ciRNAs play a *cis*-regulatory role in promoting transcription of their host genes by associating with the Pol II machinery [18]. Stable intronic sequence RNAs (sisRNAs) were also found in both the oocyte nucleus and cytoplasm of *Xenopus tropicalis* [70,71]. Whether these sisRNAs form similar circle structures or have similar roles in Pol II regulation awaits further investigation.

circRNAs Generated from Back-Spliced Exons

By taking advantage of non-polyadenylated transcriptomes and specific computational approaches that identify reads mapped to back-splice junction sites with a reversed genomic orientation [66,72,73], a large number of circRNAs from back-spliced exons have been recently rediscovered in various cell lines/tissues and across species [15–17,19,74–78]. Unlike canonical splicing, which ligates an upstream 5′ ss with a downstream 3′ ss to form a linear RNA, back-splicing connects a downstream 5′ ss with an upstream 3′ ss to yield a circRNA with a 3′,5′-phosphodiester bond.

Back-splicing is unfavorably catalyzed by the spliceosomal machinery [66,74,79,80] and regulated by both *cis* elements [19,81] and *trans* factors [74,82] (Figure 3B). Theoretically, orientation-opposite complementary sequences that juxtapose flanking introns of circularized exons can form RNA pairs to generally facilitate back-splicing by bringing the downstream donor and upstream acceptor sites close together. Protein factors that bind to pre-mRNAs to bridge flanking introns together could also enhance circRNA formation [74,82]. Of note, back-splicing can also be repressed by forming competitive RNA pairing within individual introns [19] and protein factors can downregulate circRNA biogenesis by melting paired RNA duplexes [76,77]. Thus, it would seem that the regulation of circRNA production in cells is more complicated than is currently appreciated. For instance, it has been reported that *cis* elements and *trans* factors could act in a combinatorial manner to regulate circRNA formation [83]. Some RBPs regulate circRNA biogenesis by directly associating with intronic RNA pairing, especially that formed by inverted repeated *Alu* sequences (IR*Alu*s) [84,85].

The majority of circRNAs are usually processed from internal exons of pre-mRNAs and normally contain multiple exons [19]. Since it is catalyzed by spliceosomal machineries [74,79,80,86], back-splicing can compete with canonical splicing [19,74,87]. As illustrated in Figure 3C, back-splicing for circRNA formation can occur before splicing, referred to as the ‘direct back-splicing’ model (left panel), or back-splicing occurs after canonical splicing, referred to as the ‘lariat-intermediate’ model (right panel), in which circRNA is produced from a long spliced-out intron containing exons for later back-splicing [14,15,17,73]. Although direct lines of biochemical evidence are still needed to evaluate which step occurs first, it is possible that these two steps may occur stochastically or synergistically. A recent study aiming to illustrate the link between circRNA processing and transcription using **4sUDRB-seq** [96] revealed that endogenous circRNA production from pre-mRNA back-splicing is slow, largely occurs post-



Trends in Genetics

Figure 3. The Biogenesis of Circular RNA (circRNA) and Its Complex Regulation. (A) Circular intronic RNA (ciRNA) is derived from excised introns and depends on consensus RNA sequences (orange bars) to avoid debranching of the lariat intron [18]. (B) circRNA produced by back-splicing circularization is catalyzed by the spliceosome machinery. Back-splicing is enhanced by *cis* orientation-opposite complementary sequences (red arrows) in introns flanking circularized exons (left panel) [19,79,81] or *trans* protein factors that can facilitate the positioning of distal back-splicing sites in close proximity (right panel) [74,82]. (C) Two proposed models for circRNA biogenesis. Left: 'Direct back-splicing' model, in which back-splicing occurs before splicing. Right: 'Lariat-intermediate' model, in which circRNA is

(Figure legend continued on the bottom of the next page.)

transcriptionally, and is associated with fast Pol II transcription [80]. The low efficiency of back-splicing might lead to the observation that most circRNAs are less abundant than their linear mRNA counterparts. However, some circRNAs were detected at levels as high as or even higher than their linear mRNA counterparts, possibly due to their stability and post-transcriptional accumulation (Figure 3D) [15,17,77,80].

The complex regulation of circRNA formation is further evidenced by the observation that one gene locus can produce multiple circRNAs [19] through alternative back-splice site selection and alternative ss selection [68]. The alternative (5′ or 3′) back-splice site selection is positively correlated with the existence of multiple RNA pairs that bracket different circle-forming exons [68]. Mechanically, an across-intron RNA pairing flanking proximal back-splice sites could lead to proximal back-splice site selection, whereas an across-intron RNA pairing that flanks the distal back-splice sites could lead to distal back-splice site selection (exemplified by alternative 5′ back-splicing in Figure 3E). Interestingly, alternative back-splicing is more common in human than in other examined non-primate species, largely due to the abundance of primate-specific *Alu* sequences in the human genome [68,75]. The majority of annotated human circRNAs comprise multiple exons [19] and all four types of AS have been identified in circRNAs [68] (exemplified by intron retention in Figure 3F).

Although by far the majority of circRNAs await functional annotation, recent studies have begun to reveal that some circRNAs regulate gene expression at multiple levels by titrating miRNAs [16,88], regulating Pol II transcription [89], or competing with host linear RNA splicing [19,74,87]. Furthermore, pseudogenes are also found to be retrotranscribed from circRNAs [90]. Moreover, the stability of circRNAs and their detection in human blood indicate that circRNAs may be used as disease biomarkers [91]. Very recently, several studies have shown that cytoplasmic circRNAs can be translated into small peptides [92–94], and we also uncovered the involvement of circRNAs in viral infection [84]. As circRNAs are generally expressed at low levels [66], we speculate that many circRNAs, but not one specific circRNA, may act as a group in response to the immune response [84]. Together these recent findings have suggested that circRNAs are not merely tolerated byproducts of eukaryotic transcriptomes, but their regulatory potential has not yet been fully explored.

Concluding Remarks

Recent advances in RNA-seq data mining has allowed the discovery of thousands of lncRNA molecules, which can be subdivided into a diverse catalog of lncRNAs according to their molecular features and processing pathways (Figure 1). The transcriptional activity of some loci including lincRNAs, eRNAs, or PROMPTs also appears to influence neighborhood gene expression [40,95], arguing that lncRNAs themselves or their transcriptional activity executes a regulatory function. However, there is no doubt that many lncRNAs have established, important roles in diverse cellular contexts and biological processes [6,7]. Nevertheless, the seemingly endless discovery of new RNA species has significantly enlarged the diversity of eukaryotic transcriptomes, provided new modes of gene regulation, and shed new light on molecular causes of human diseases. With unexpected integrated approaches, new types of lncRNA are likely to be discovered in the future.

Outstanding Questions

How are different lncRNA species processed, matured, and degraded? What rules are applied to distinguish each class of lncRNAs and to differentiate them from eukaryotic mRNA processing?

How are the molecular features and life cycles linked to the cellular localization and function of lncRNAs?

Are there new types of lncRNA yet to be discovered in eukaryotic transcriptomes? How do we discover additional new RNA species?

To what extent is the expression and function of different classes of lncRNAs conserved? Does the strength of their non-conservation contribute to any evolutionary advantage?

How are diverse lncRNA species structured and how do they act? Are there distinctions or commonalities in gene regulation?

What approaches can be applied to dissect their distinctions and commonalities, such as common structural domains for localization and/or function?

produced from a long spliced-out intron containing exons for later back-splicing [72,73]. (D) Processing of endogenous circRNA from precursor mRNA (pre-mRNA) back-splicing. In general, endogenous back-splicing is slow, competes with linear mRNA splicing, and is associated with fast RNA polymerase II (Pol II) transcription [80]. Some circRNAs can post-transcriptionally accumulate to high levels [15,17,77,80]. (E) Competition between RNA pairs flanking proximal or distal back-splice sites leads to alternative back-splice site selection. The orientation-opposite complementary sequences (red arrows) flanking proximal or distal back-splice sites lead to alternative 5′ back-splice site selection [68]. (F) A schematic diagram of one type of alternative splicing (AS), intron retention, within circRNAs. Blue bars, exons; grey lines, introns; blue broken lines, back-splicing [68]

Remarkably, given the vast number of lncRNAs, we still know little about what they do in cells. Future studies aimed at understanding the characteristics and life cycles of the different categories of lncRNAs will be helpful in underscoring their regulatory potential and mechanism of action in depth (see Outstanding Questions).

Acknowledgments

The authors apologize to colleagues whose work is not discussed due to space limitations. This work was supported by grants XDB19020104 from CAS, 2016YFA0100701 from MOST, and 91440202 and 91540115 from NSFC.

References

1. Djebali, S. *et al.* (2012) Landscape of transcription in human cells. *Nature* 489, 101–108
2. Hangauer, M.J. *et al.* (2013) Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* 9, e1003569
3. Moore, M.J. and Proudfoot, N.J. (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136, 688–700
4. Bentley, D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175
5. Ulitsky, I. and Bartel, D.P. (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46
6. Chen, L.L. (2016) Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* 41, 761–772
7. Quinn, J.J. and Chang, H.Y. (2016) Unique features of long noncoding RNA biogenesis and function. *Nat. Rev. Genet.* 17, 47–62
8. Wilusz, J.E. *et al.* (2008) 3′ End processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135, 919–932
9. Sunwoo, H. *et al.* (2009) MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* 19, 347–359
10. Yin, Q.F. *et al.* (2012) Long noncoding RNAs with snoRNA ends. *Mol. Cell* 48, 219–230
11. Zhang, X.O. *et al.* (2014) Species-specific alternative splicing leads to unique expression of sno-lncRNAs. *BMC Genomics* 15, 287
12. Xing, Y.H. *et al.* (2017) SLERT regulates DDX21 rings associated with Pol I transcription. *Cell* 169, 664–678.e16
13. Wu, H. *et al.* (2016) Unusual processing generates SPA lncRNAs that sequester multiple RNA binding proteins. *Mol. Cell* 64, 534–548
14. Salzman, J. *et al.* (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7, e30733
15. Jeck, W.R. *et al.* (2013) Circular RNAs are abundant, conserved, and associated with *ALU* repeats. *RNA* 19, 141–157
16. Memczak, S. *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338
17. Salzman, J. *et al.* (2013) Cell-type specific features of circular RNA expression. *PLoS Genet.* 9, e1003777
18. Zhang, Y. *et al.* (2013) Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806
19. Zhang, X.O. *et al.* (2014) Complementary sequence-mediated exon circularization. *Cell* 159, 134–147
20. Preker, P. *et al.* (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res.* 39, 7179–7193
21. Kim, T.K. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187
22. Balbin, O.A. *et al.* (2015) The landscape of antisense gene expression in human cancers. *Genome Res.* 25, 1068–1079
23. Preker, P. *et al.* (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854
24. Neil, H. *et al.* (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042
25. Seila, A.C. *et al.* (2008) Divergent transcription from active promoters. *Science* 322, 1849–1851
26. Taft, R.J. *et al.* (2009) Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* 41, 572–578
27. Lubas, M. *et al.* (2015) The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep.* 10, 178–192
28. Lloret-Llinares, M. *et al.* (2016) Relationships between PROMPT and gene expression. *RNA Biol.* 13, 6–14
29. Ntini, E. *et al.* (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* 20, 923–928
30. Andersson, R. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461
31. De Santa, F. *et al.* (2010) A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* 8, e1000384
32. Lam, M.T.Y. *et al.* (2013) Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* 498, 511–515
33. Li, W.B. *et al.* (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520
34. Melo, C.A. *et al.* (2013) eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* 49, 524–535
35. Hah, N. *et al.* (2013) Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* 23, 1210–1223
36. Lai, F. *et al.* (2015) Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399–403
37. Kowalczyk, M.S. *et al.* (2012) Intragenic enhancers act as alternative promoters. *Mol. Cell* 45, 447–458
38. Leung, D. *et al.* (2015) Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* 518, 350–354
39. Sigova, A.A. *et al.* (2015) Transcription factor trapping by RNA in gene regulatory elements. *Science* 350, 978–981
40. Engreitz, J.M. *et al.* (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452–455
41. Cabili, M.N. *et al.* (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Gene Dev.* 25, 1915–1927
42. Ulitsky, I. *et al.* (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537–1550
43. Katayama, S. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science* 309, 1564–1566
44. Khalil, A.M. *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11667–11672
45. Derrien, T. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789
46. Mele, M. *et al.* (2017) Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* 27, 27–37

47. Schlackow, M. *et al.* (2017) Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* 65, 25–38
48. Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 12, 340–345
49. Tuck, A.C. and Tollervey, D. (2013) A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* 154, 996–1009
50. Zhang, Y. *et al.* (2014) Life without A tail: new formats of long noncoding RNAs. *Int. J. Biochem. Cell Biol.* 54, 338–349
51. Gopalan, V. *et al.* (2002) RNase P: variations and uses. *J. Biol. Chem.* 277, 6759–6762
52. Brown, J.A. *et al.* (2012) Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19202–19207
53. Wilusz, J.E. *et al.* (2012) A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Gene Dev.* 26, 2392–2407
54. Mitton-Fry, R.M. *et al.* (2010) Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* 330, 1244–1247
55. Tycowski, K.T. *et al.* (2012) Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep.* 2, 26–32
56. Dhir, A. *et al.* (2015) Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat. Struct. Mol. Biol.* 22, 319–327
57. Rodriguez-Trelles, F. *et al.* (2006) Origins and evolution of spliceosomal introns. *Annu. Rev. Genet.* 40, 47–76
58. Yang, L. (2015) Splicing noncoding RNAs from the inside out. *Wiley Interdiscip. Rev. RNA* 6, 651–660
59. Filipowicz, W. and Pogacic, V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* 14, 319–327
60. Yang, L. *et al.* (2011) Genomewide characterization of non-polyadenylated RNAs. *Genome Res.* 12, R16
61. Cassidy, S.B. *et al.* (2012) Prader–Willi syndrome. *Genet. Med.* 14, 10–26
62. Bieth, E. *et al.* (2015) Highly restricted deletion of the SNORD116 region is implicated in Prader–Willi syndrome. *Eur. J. Hum. Genet.* 23, 252–255
63. de Smith, A.J. *et al.* (2009) A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum. Mol. Genet.* 18, 3257–3265
64. Duker, A.L. *et al.* (2010) Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D Box snoRNA cluster in Prader–Willi syndrome. *Eur. J. Hum. Genet.* 18, 1196–1201
65. Sahoo, T. *et al.* (2008) Prader–Willi phenotype caused by paternal deficiency for the HBII-85 C/D Box small nucleolar RNA cluster. *Nat. Genet.* 40, 719–721
66. Chen, L.L. (2016) The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 17, 205–211
67. Lasda, E. and Parker, R. (2014) Circular RNAs: diversity of form and function. *RNA* 20, 1829–1842
68. Zhang, X.O. *et al.* (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287
69. Qian, L. *et al.* (1992) A spliced intron accumulates as a lariat in the nucleus of T-cells. *Nucleic Acids Res.* 20, 5345–5350
70. Gardner, E.J. *et al.* (2012) Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Gene Dev.* 26, 2550–2559
71. Talhouarne, G.J. and Gall, J.G. (2014) Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *RNA* 20, 1476–1487
72. Chen, L.L. and Yang, L. (2015) Regulation of circRNA biogenesis. *RNA Biol.* 12, 381–388
73. Jeck, W.R. and Sharpless, N.E. (2014) Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32, 453–461
74. Ashwal-Fluss, R. *et al.* (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66
75. Dong, R. *et al.* (2016) Increased complexity of circRNA expression during species evolution. *RNA Biol.* Published online December 16, 2016. http://dx.doi.org/10.1080/15476286.2016.1269999
76. Ivanov, A. *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 10, 170–177
77. Rybak-Wolf, A. *et al.* (2015) Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885
78. Westholm, J.O. *et al.* (2014) Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.* 9, 1966–1980
79. Starke, S. *et al.* (2015) Exon circularization requires canonical splice signals. *Cell Rep.* 10, 103–111
80. Zhang, Y. *et al.* (2016) The biogenesis of nascent circular RNAs. *Cell Rep.* 15, 611–624
81. Liang, D.M. and Wilusz, J.E. (2014) Short intronic repeat sequences facilitate circular RNA production. *Gene Dev.* 28, 2233–2247
82. Conn, S.J. *et al.* (2015) The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134
83. Kramer, M.C. *et al.* (2015) Combinatorial control of *Drosophila* circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Gene Dev.* 29, 2168–2182
84. Li, X. (2017) Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol. Cell* http://dx.doi.org/10.1016/j.molcel.2017.05.023
85. Aktaş, T. *et al.* (2017) DHX9 suppresses RNA processing defects originating from the *Alu* invasion of the human genome. *Nature* 544, 115–119
86. Wang, Y. and Wang, Z.F. (2015) Efficient backsplicing produces translatable circular mRNAs. *RNA* 21, 172–179
87. Kelly, S. *et al.* (2015) Exon skipping is correlated with exon circularization. *J. Mol. Biol.* 427, 2414–2417
88. Hansen, T.B. *et al.* (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388
89. Li, Z.Y. *et al.* (2015) Exon–intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 22, 256–264
90. Dong, R. *et al.* (2016) circRNA-derived pseudogenes. *Cell Res.* 26, 747–750
91. Memczak, S. *et al.* (2015) Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* 10, e0141214
92. Legnini, I. *et al.* (2017) circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 66, 22–37
93. Pamudurti, N.R. *et al.* (2017) Translation of circRNAs. *Mol. Cell* 66, 9–21
94. Yang, Y. *et al.* (2017) Extensive translation of circular RNAs driven by N^6-methyladenosine. *Cell Res.* 27, 626–641
95. Liu, S.J. *et al.* (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* Published online December 15, 2016. http://dx.doi.org/10.1126/science.aah7111
96. Fuchs, G. *et al.* (2014) 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Res.* 15, R69