Trends in Cell Biology

CellPress

Review

*ALU*ternative Regulation for Gene Expression

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

Alu **elements belong to the primate-specific** ***SINE*** **family of retrotransposons and constitute almost 11% of the human genome.** ***Alu*****s are transcribed by RNA polymerase (Pol) III and are inserted back into the genome with the help of autonomous** ***LINE*** **retroelements. Since** ***Alu*** **elements are preferentially located near to or within gene-rich regions, they can affect gene expression by distinct mechanisms of action at both DNA and RNA levels. In this review we focus on recent advances of how** ***Alu*** **elements are pervasively involved in gene regulation. We discuss the impacts of** ***Alu*** **DNA sequences that are in close proximity to genes, Pol-III-transcribed free** ***Alu*** **RNAs, and Pol-II-transcribed** ***Alu*** **RNAs that are embedded within coding or noncoding RNA transcripts. The recent elucidation of** ***Alu*** **functions reveals previously underestimated roles of these selfish or junk DNA sequences in the human genome.**

Abundant *Alu* Elements in the Human Genome

The initial sequencing and analysis of the human genome revealed that ~45% of the genome is derived from transposable elements [1]. One of the most abundant of these mobile elements is the short interspersed nuclear repetitive DNA elements (SINEs). The majority of human *SINE*s belong to a single family known as *Alu* elements, which are ~300 nucleotides in length and constitute ~11% of the reference human genome, with >1 million copies [1]. The original *Alu* sequence is thought to have resulted from dimerization of two distinct 7SL RNA genes via a head-to-tail fusion in the primate lineage (Figure 1A) [2] (reviewed in [3–5]). Since then, *Alu* elements have been successfully amplified throughout the genome by retrotransposition of Pol-III-transcribed RNA intermediates (Figure 1B) [4,6]. Different from the long interspersed element 1 (L1) insertion rate, which is relatively constant among primates including humans, the rate of *Alu* insertion varies in a species- and lineage-specific manner [7] (reviewed in [8]). Thus, *Alu* polymorphism has been widely used for the study of human population genetics [9,10] and primate comparative genomics [7] (reviewed in [4]).

Although originally considered as 'selfish' or junk DNA sequences because most of them seem to be genetically inert except for their own retrotransposition [11,12], *Alu* elements can accelerate genome evolution by nonallelic recombination that results in duplications or deletions of DNA segments [6,13]. *Alu* elements can also lead to genetic diversity and genetic defects by disrupting coding regions or splicing events (reviewed in [14–16]). Furthermore, recent genome-wide analyses have clearly shown that *Alu* elements play significant roles in a variety of biological processes. These new findings suggest previously underestimated impacts of *Alu* sequences in the human genome.

The functional potential of *Alu*s depends on their genomic localization and sequence features. One distinct characteristic of *Alu* distribution in the human genome is that it is biased toward gene-rich regions (Figure 1C); probably because short *Alu* (~300 bp) insertions are easily

Trends

Primate-specific *Alu*s constitute 11% of the human genome, with >1 million copies, and their genomic distribution is biased toward gene-rich regions.

The functions of *Alu*s are highly associated with their sequence and structural features.

*Alu*s can regulate gene expression by serving as *cis* elements.

Pol-III-transcribed free *Alu*s mainly affect Pol II transcription and mRNA translation *in trans*.

Embedded *Alu*s within Pol-II-transcribed mRNAs can impact their host gene expression through the regulation of alternative splicing, and RNA stability and translation.

Nearly half of annotated *Alu*s are located in introns; RNA pairing formed by orientation-opposite *Alu*s across introns promotes circRNA biogenesis.

[1]State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Key Laboratory of Molecular Andrology, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China
[2]Key Laboratory of Computational Biology, CAS Center for Excellence in Brain Science and Intelligence Technology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[3]School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China

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

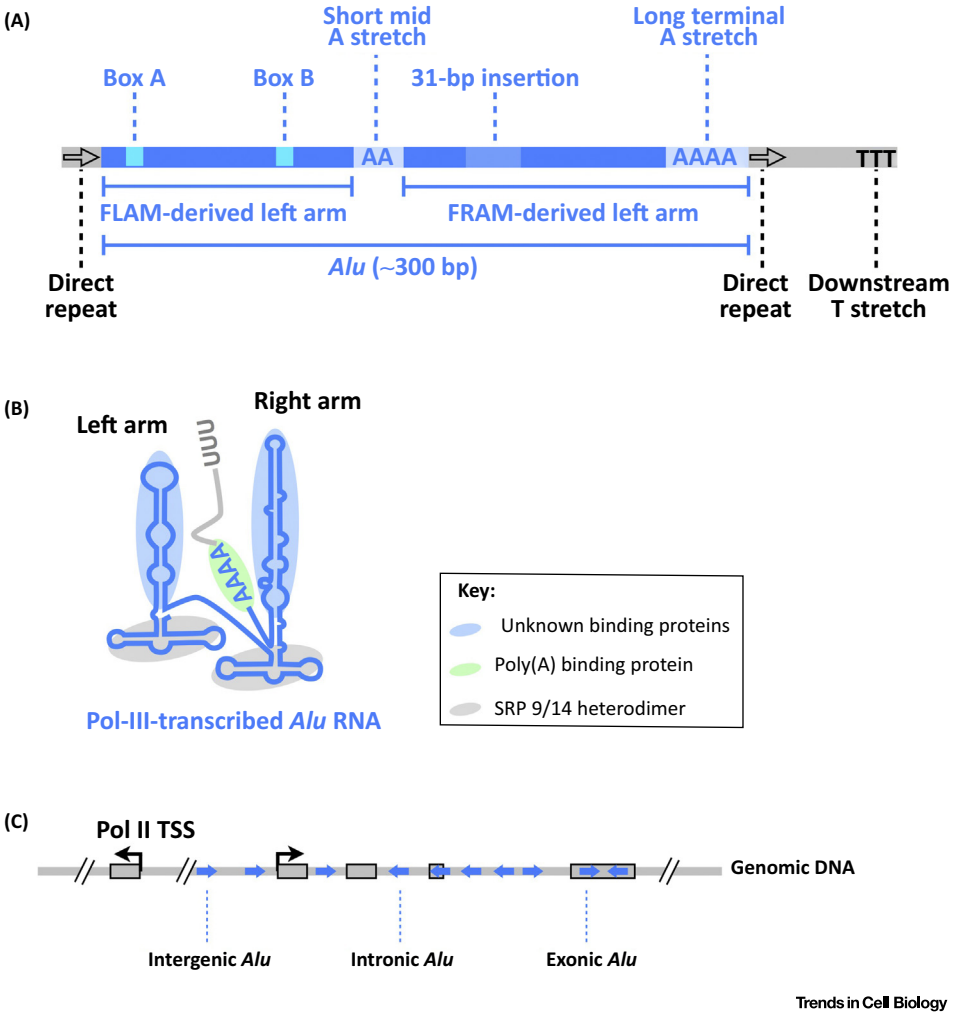

Figure 1. Features of *Alu* Sequences, Structures and Genomic Locations. (A) Sequence features of an *Alu* element and its genomic contexts. The dimeric *Alu* sequence is separated by a short A stretch in the middle and tailed by a long A stretch at the 3′ end. Boxes A and B in the left arm are derived from Pol III promoter, and a 31-bp insertion is located in the right arm. Upstream and downstream direct repeats indicate the insertion site of *Alu* element, and a T stretch is located downstream of the *Alu* element for transcription termination. FLAM, left fossil *Alu* monomer; FRAM, right fossil *Alu* monomer. (B) A typical structure of a Pol-III-transcribed *Alu* RNA and its binding protein factors (modified from [8]). The free *Alu* RNA transcript by Pol III is thought to form separate structures for left arm and right arm monomer units. SRP 9/14 heterodimer, poly(A) binding protein, and other unknown binding proteins are shown to bind to Pol-III-transcribed *Alu* RNA. (C) The genomic location of *Alu* elements is biased toward gene-rich regions. The majority of human *Alu* elements are close to (as intergenic *Alu*) or within (as either intronic or exonic *Alu*) gene bodies. Abbreviations: Pol, RNA polymerase; TSS, transcription start site.

tolerated when located in gene-rich regions [1,17]. In contrast, the long *L1* elements (~6000 bp) are under strong negative selection when located in genes, leading to their abundance in gene-poor regions [8]. The observation that younger *Alu* and *L1* elements exhibit fewer differences than those between older *Alu* and *L1* elements in their genomic locations suggests that *Alu* elements are mostly removed during evolution when they are not in gene-rich regions, through nonallelic homologous recombination [8]. Alternatively, the chromatin at gene-rich regions might be more accessible for *Alu* insertion to occur, leading to a positive selection on the *Alu*s in the gene-rich regions. However, the detailed mechanism of this possibility has yet to be proven.

The specific primary sequences have endowed great potential for *Alu*s to be involved in gene regulation. For instance, *Alu* elements are enriched with CpG dinucleotides, which are vulnerable to methylation at the C of CpGs. In this scenario, when inserted in proximity to gene bodies, *Alu* elements could be actively involved in gene expression regulation by introducing methylation at the DNA level. In addition, multiple splice donor and acceptor sites are located in *Alu* elements; mostly in the antisense strand [18]. When inserted in introns and cotranscribed with host genes, the transcribed *Alu* sequences can introduce a new splicing site selection [19], resulting in increased diversity of mRNA isoforms. Finally, transcribed *Alu* sequences can introduce additional *cis*-regulatory elements, such as new polyadenylation [p(A)] signals [20] or new adenine and uracil rich elements (AREs) [21], which are frequently involved in alternative polyadenylation or mRNA decay.

The regulatory roles of *Alu*s are also associated with the formation of secondary structures. Single *Alu* sequences, when transcribed independently or as part of other RNAs, can form specific secondary structures to regulate translation initiation [22–24]. Given the high sequence similarity among all *Alu* subfamilies [25], the existence of orientation-opposite *Alu* repeats within the same transcripts has a tendency to form inverted repeated *Alu* structures (IR*Alu*s) [6,8]. Depending on their genomic locations, IR*Alu*s have the potential to regulate mRNA nuclear retention or circular RNA (circRNA) formation that is produced from pre-mRNA back splicing of exons [26,27]. IR*Alu*s could also be formed in an intermolecular manner between long noncoding RNAs (lncRNAs) and mRNAs, which triggers mRNA degradation by recruiting specific protein factors [28]. These recent studies have revealed that *Alu* elements play important roles in gene expression due to their intrinsic sequence and structure features as well as their abundance in gene rich regions.

Transcription Regulation by *Alu*-Derived *cis*-Elements

The unique sequence features of *Alu* elements lead to gene regulation at the DNA level. *Alu* elements are enriched with CpG dinucleotides and contribute ~25% of CpG dinucleotides in the human genome [29]. These *Alu* DNA sequences thus provide a rich resource for DNA methylation that primarily occurs at the cytosine residues of CpGs in mammals. Although highly methylated in intronic and intergenic regions, *Alu* elements are typically methylated at lower levels when close to transcription start site (TSS) regions (Figure 2A, left panel) [29]. In addition, it appears that the methylation levels of *Alu* elements are variable among different tissues, implying a dynamic influence of *Alu* elements on transcription regulation [29]. *Alu* elements could also participate in transcription regulation by providing multiple transcription factor (TF) binding sites [30] when inserted in gene-rich regions. Thus, the CpG methylation of *Alu* elements and their interaction with TFs suggest a regulatory potential of proximal *Alu* elements as promoters (Figure 2A).

Alu elements have been speculated to evolve to new enhancers [23]. Upstream *Alu*s that are proximal to TSSs are reported to be enriched with epigenetic enhancer/enhancer-like H3K4me1 (histone H3 lysine 4 monomethylation) marks in a tissue-specific manner. These putative *Alu* enhancers are spatially close to nearby promoters through 3D long-range chromosome interactions (Figure 2B), thus potentially activating neighboring gene expression [31]. Further analysis of sequence conservation and multiple epigenetic characteristics on various subfamilies of *Alu* elements has revealed that the gain of epigenetic enhancer marks and TF binding motifs in *Alu* elements are positively correlated with their evolutionary age [31]. This observation is consistent with a proposed model that *Alu*s evolved first to protoenhancers and then to functional enhancers once fixed in the human genome [31]. Thus, some *Alu*s could evolve to enhancers and other *cis* elements to regulate gene expression at the transcriptional level; a process that is probably associated with evolutionary selection.

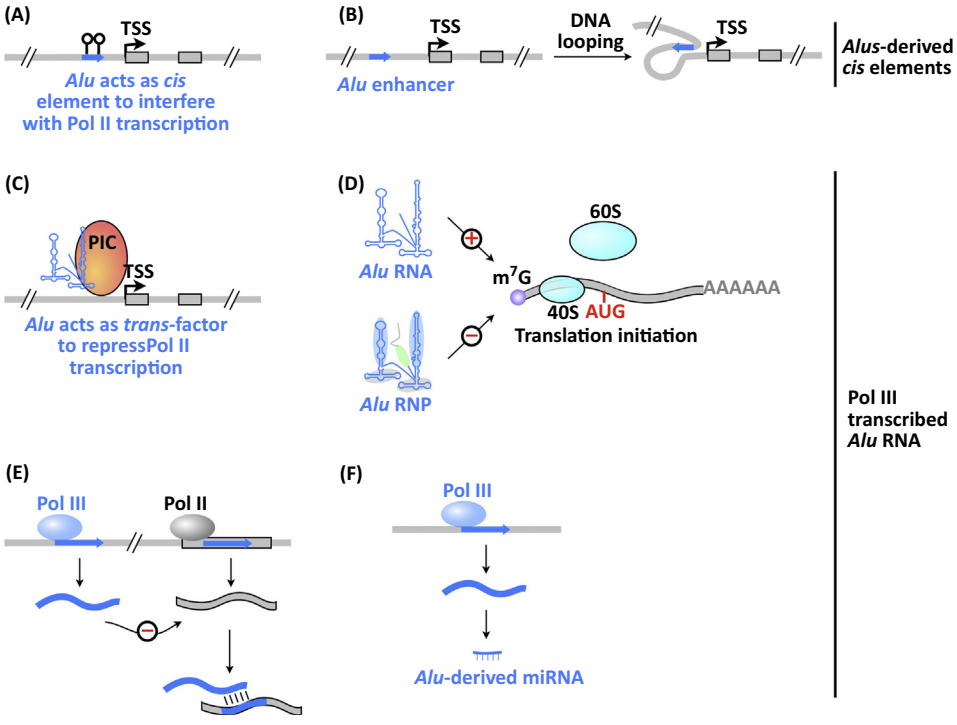

Figure 2. Gene Regulation by *Alu*-Derived *cis* Elements and Pol-III-Transcribed *Alu*s. (A,B) *Alu* functions at the DNA level. (A) *Alu* elements act as potential promoters to interfere with Pol II transcription. (B) Some newly evolved *Alu* enhancers can preferentially connected to nearby promoters through long-range chromosome interactions to activate gene expression. (C–F) The regulatory potential of Pol-III-transcribed free *Alu*s. (C) Pol-III-transcribed *Alu*s function as *trans*-modulators to repress Pol II transcription initiation. (D) Pol III-transcribed *Alu* RNAs and *Alu* RNP affect gene expression at the translational level with opposite effects. (E) Pol III-transcribed *Alu*s negatively regulate mRNA expression through intermolecular interaction. (F) Some primate-specific miRNA sequences are derived from Pol III-transcribed *Alu* sequences. Abbreviations: PIC, Pol II preinitiation complex; Pol, RNA polymerase; TSS, transcription start site.

Pol-III-Transcribed Free *Alu* RNAs in Gene Regulation

Alu elements can be transcribed by their own Pol III promoters to generate free *Alu* RNA transcripts. Due to inefficient Pol III promoter A and B boxes of *Alu* elements and other unfavorable epigenetic features, Pol-III-transcribed free *Alu* RNAs are usually expressed and accumulate at a low level (reviewed in [6,8]). However, the expression of free *Alu* RNAs is significantly increased under different stress stimulations, including viral infection, heat shock, and some cancerous situations [32] (reviewed in [6,8]).

Pol-III-transcribed *Alu* RNAs can affect gene expression at both the transcriptional (Figure 2C) and translational (Figure 2D) levels [6,8,23,33]. Human *Alu* RNAs and mouse B2 RNAs act in *trans* as a general Pol II repressor by binding to Pol II initiation complex during heat shock [33] (Figure 2C). Pol-III-transcribed *Alu* RNA boosts translation of cotransfected reporters both *in vivo* [34] and *in vitro* [22] (Figure 2D). However, purified synthetic *Alu* ribonucleoproteins (RNPs) that consist of *Alu* RNA associated with SRP9/14 proteins, which are known to regulate translation at the level of initiation, negatively affect translation *in vitro* [22] (Figure 2D). The discrepancy of Pol III *Alu* RNA and *Alu* RNP in translation regulation *in vivo* and *in vitro* could be due to conformational changes of *Alu* RNA with or without SRP9/14 binding [23]. A series of *Alu* sequence deletions showed that the regulatory activity of Pol-III-transcribed *Alu* RNAs on translation was located in the right arm [35], whereas the left arm exhibited a higher affinity for SRP9/14 heterodimers [36]. In addition, it has been observed that the Pol-III-transcribed

independent *Alu* RNA interacts with mRNA through sequence complementarity that negatively regulates gene expression [37] (Figure 2E). However, the detailed mechanism requires further investigation.

Some primate-specific miRNA sequences are located within *Alu* elements (Figure 2F) [38,39]. Lines of evidence have revealed that Pol III transcription of *Alu* RNAs is associated with and sufficient for the generation of these *Alu*-derived miRNAs [40,41]. Theoretically, the retrotransposition of the miRNA-producing *Alu*s could simultaneously expand their resident miRNAs in the human genome [42]. Duplication of a miRNA core cassette in chromosome 19 was reported to be facilitated by *Alu* retrotransposition [43], which led to dozens of these repeated miRNA cassettes in the chromosome 19 miRNA cluster (C19MC) [44]. The exclusive expression of C19MC miRNAs in germ line cells was found to target *Alu* sequences (presumably Pol-III-transcribed *Alu*s) and inhibit their duplication, leading to coevolution between *Alu* repeats and resident miRNAs [43].

Widespread Cotranscription of Embedded *Alu* RNAs in Pol II Transcripts and Their Functional Diversity

Alu elements are widespread within tens of thousands of gene bodies, leading to a broad transcription of embedded *Alu* RNAs as a part of long Pol-II-transcribed RNAs. It has been demonstrated that these embedded *Alu* RNA sequences play important regulatory roles in several biological processes by distinct mechanisms of action.

Alternative Splicing (AS) Regulated by Embedded *Alu* Sequences

One of the most well-understood functions of *Alu*s in gene regulation is AS. Insertions of mobile *Alu* elements in the human genome that cause human disease are thought to be largely due to *Alu*-related alternative splicing (reviewed in [14]). The differential usage of splice sites leads to AS, which significantly expands the proteomic complexity and functional diversity [45]. Nearly all human multiexonic genes undergo AS to yield more than one mature mRNA/protein [45]. One intrinsic feature of *Alu* elements is that they contain multiple potential splice donor and acceptor sites, suggesting their involvement in AS. A large number of intronic *Alu* elements can serve as a reservoir of alternative exons, and $\geq$5% of human alternative spliced exons are derived through *Alu* exonization [19] (Figure 3A). Most (19 of 23) potential splice sites are present on the minus strand of the *Alu* consensus sequence, suggesting that an opposed *Alu* insertion has a better chance to be exonized than one in the same direction of transcription of the host gene [23]. Accordingly, it has been reported that 85% of *Alu*-containing exons are derived from antisense *Alu* insertions [46].

Many *Alu*-containing exons are alternatively spliced [46,47]. As an AS event, *Alu* exonization can be regulated by auxiliary splicing enhancers/silencers, which can be also derived from *Alu* consensus sequences [48,49]. Even a single mutation can result in constitutive splicing of *Alu* exons that are originally alternatively spliced or silenced. Such mutations result in human diseases [18], such as autosomal recessive Alport syndrome [50] (reviewed in [19]). Also, inclusion of *Alu* exons in the coding region can introduce premature termination codons or induce a frameshift on the encoded protein [51]. Furthermore, it has been shown that *Alu* exonization in 5′ untranslated regions (UTRs) could alter the translation of some master transcriptional regulators, such as zinc finger transcription factors [47].

A cis-Regulatory Role for 5′-/3′-UTR *Alu* Sequences in Translation

Embedded *Alu* RNAs, when cotranscribed with their host genes in 5′-/3′-UTR regions, can regulate the translation of their host genes (Figure 3B). Different from Pol-III-transcribed *Alu* RNAs that act in *trans*, embedded *Alu* RNAs when located in UTR regions of host Pol II transcripts can function in *cis* to regulate translation initiation of their host mRNAs [23]. On the

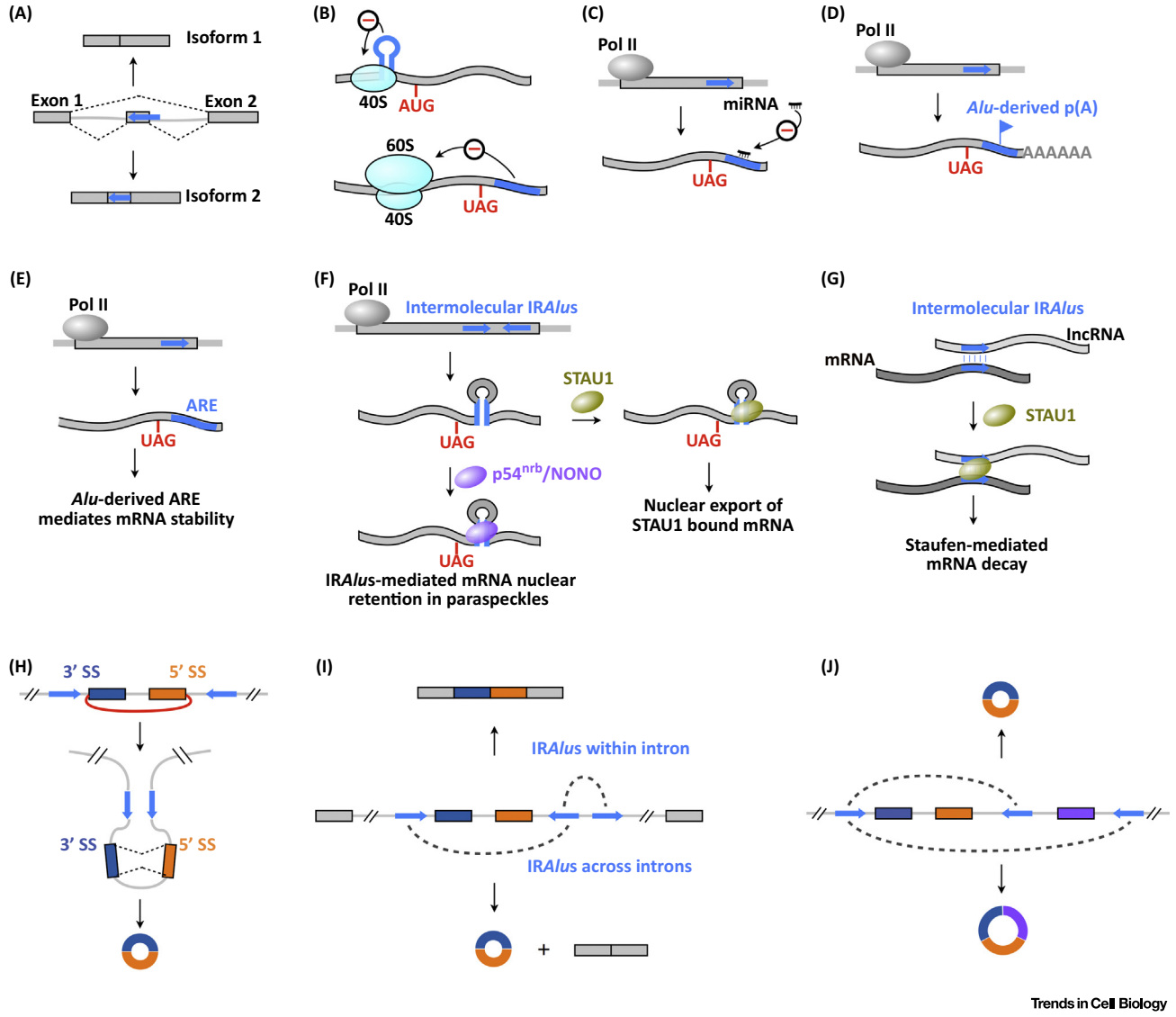

Figure 3. Functional Diversity of Pol-II-Transcribed *Alu*s (Embedded *Alu*s). (A) *Alu* exonization through alternative splicing. (B) Embedded *Alu* sequences regulate the translation of their host genes in *cis*. (C) Embedded *Alu* RNAs are involved in the miRNA regulatory network. (D) Embedded 3′-UTR *Alu* sequences contribute to the gain of new p(A) sites. (E) Embedded *Alu* sequence mediates host gene stability by providing new AREs. (F) IR*Alu*s formed within mRNA 3′ UTRs by orientation-opposite *Alu* repeats can bind to either $p54^{nrb}$/NONO or STAU1, leading to *IRAlus*-mediated mRNA nuclear retention or nuclear export of STAU1 bound mRNA, respectively. (G) The association of nonperfect intermolecular IR*Alu*s between lncRNAs and 3′ UTRs of mRNAs with STAUs induces Staufen-mediated mRNA decay. (H) RNA pairing formed by IR*Alu*s across introns that flank back-spliced exons enhances circRNA biogenesis. (I) RNA pairing formed by IR*Alu*s across flanking introns competes with IR*Alu*s within an individual intron in the same gene locus, leading to competition between back splicing for circRNAs and canonical splicing for linear RNAs. (J) The competition of IR*Alu*s across different pairs of flanking introns results in multiple circRNAs in the same gene locus through alternative back splicing. Note that the wide spread A-to-I RNA editing on *Alu* sequences have the potential to affect all *Alu* RNA-involved gene regulation described above. Abbreviations: ARE, adenine and uracil rich element; circRNA, circular RNA; IR*Alu*s, inverted repeated *Alu* elements; lncRNA, long noncoding RNA; Pol, RNA polymerase; STAU1, Staufen 1; UTR, untranslated region.

one hand, an *Alu* element in the longer 5′ UTR of the *Brca1* transcript, which is only expressed in breast cancer tissue, can form a stable secondary structure that inhibits 40S ribosomal subunit scanning of the leader sequence for the initiation codon in this specific *Brca1* transcript [24] (top, Figure 3B). On the other hand, embedded *Alu* at the 3′-UTR can act as a translation inhibitor [52] (bottom, Figure 3B). Additional cases of translation initiation inhibition by the

UTR-region *Alu* elements have been reported in other genes [53,54] (reviewed in [23]). Genome-wide analysis, together with mutation experiments, has also indicated that newly created 5′-UTR *Alu* exons could repress translational efficiency of primary open reading frames (ORFs) by creating or elongating small upstream ORFs [47].

Crosstalk between miRNA and 3′-UTR *Alu*s

Embedded *Alu*s in mRNA 3′ UTRs can interact with miRNA regulatory networks (Figure 3C). Several miRNAs exhibit typical sequence complementarities with *Alu* elements in their highly-conserved 3′ regions, indicating that 3′-UTR *Alu* elements, after being cotranscribed with host genes, could be targeted by specific miRNAs [39,55–57]. For instance, a primate-specific miR-661 was reported to target the 3′-UTR *Alu*s in *Mdm2* and *Mdm4* mRNAs, which regulates the tumor suppressor p53 [58]. Since both miR-661 and its targeted 3′-UTR *Alu*s in *Mdm2* and *Mdm4* genes are primate specific, miR-661 regulation of *Mdm2* and *Mdm4* and thus p53 suggests an additional regulatory layer in the p53 pathway during primate evolution [58].

As potential miRNA-targeted *Alu* elements could be present in several copies in the same 3′ UTRs [56], this could lead to a multiplicity of potential miRNA target sites in the same genes. However, miRNA-targeted sites within 3′-UTR *Alu* sequences are significantly less responsive to the binding of miRNA machinery than those outside *Alu* elements; possibly due to the tight secondary structure of *Alu*s and their frequent mutability by RNA editing modification [57].

Other Roles of 3′-UTR *Alu* RNAs

The existence of embedded *Alu* elements in 3′-UTR regions can affect gene expression by introducing other potential *cis*-regulatory elements. For example, *Alu* elements have A-rich tails at their 3′ ends (Figure 1A). Thus, when inserted into the 3′ UTRs of Pol-II-transcribed genes, the consensus A-rich sequence could introduce less-conserved p(A) sites after spontaneous mutations [20] (Figure 3D). Although *Alu*-derived p(A) signals are weak, and possibly not used owing to unfavorable adjacent context sequences [20], the insertion of *Alu* elements in 3′ UTRs nevertheless contributes to the gain of new p(A) sites during evolution [59]. Interestingly, the selection of *Alu*-derived p(A) sites in the upstream introns could result in truncated transcripts, which may have different fates compared to normal transcripts [60]. In addition, a significant portion of AREs in 3′ UTRs are associated with *Alu* elements [21] (Figure 3E). Since ARE-containing mRNAs are usually unstable, these *Alu*-derived AREs might affect gene expression by regulating mRNA stability.

3′-UTR IR*Alu*s-Mediated mRNA Nuclear Retention and Export

When close to each other, orientation-opposite *Alu*s can form IR*Alu*s within 3′-UTR regions. Hundreds of 3′-UTR IR*Alu*s have been identified in the human genome [61] and many of them have been shown to be highly susceptible to adenosine to inosine (A-to-I) RNA editing, leading to suppressed protein expression of their host genes [62]. Although a correlation between nuclear retention and 3′-UTR IR*Alu*s editing was observed [62], it has been proposed that the structure of long IR*Alu*s duplexes themselves may lead to the nuclear retention of such mRNAs, regardless of A-to-I RNA editing (reviewed in [26]). Indeed, recent studies have revealed that the double-stranded (ds) IR*Alu*s structure rather than editing is required for the nuclear retention of mRNAs containing 3′-UTR IR*Alu*s (mRNAs–IR*Alu*s) [63–65].

The mRNAs containing 3′-UTR IR*Alu*s are retained within nuclear paraspeckles; a nuclear membrane-less subdomain that contains both a structural lncRNA *NEAT1* and several characteristic protein components [66]. The mRNAs–IR*Alu*s are bound to paraspeckle protein component $p54^{nrb}$/NONO (Figure 3F, left) [62–64,66], resulting in silencing of such mRNAs by restraining them from export to cytoplasm for translation.

These retained mRNAs can escape nuclear retention under certain conditions. Coactivator-associated arginine methyltransferase (CARM)1 plays an important role in repressing the IR*Alu*s-mediated mRNA nuclear retention at two levels [63]. On the one hand, CARM1 methylates p54^{nrb}/NONO to reduce its binding capacity to mRNAs–IR*Alu*s; on the other hand, CARM1 suppresses *NEAT1* transcription as a negative transcription regulator to inhibit paraspeckle formation [63,65]. Other RNA-binding proteins (RBPs), such as Staufen 1 (STAU1), can associate with 3′-UTR IR*Alu*s to free them from binding with paraspeckle protein component p54^{nrb}/NONO, resulting in the nuclear export of corresponding mRNAs (Figure 3F, right) [16,64]. The binding of STAU1 with cytoplasmic mRNAs at their 3′-UTR IR*Alu*s could also inhibit their association with dsRNA-dependent protein kinase (PKR), leading to translational de-repression on mRNAs containing 3′-UTR IR*Alu*s [64]. Moreover, genes with 3′-UTR IR*Alu*s in humans or 3′-UTR IR*SINE*s in mice could generate short mRNA isoforms without such 3′-UTR RNA duplexes through alternative polyadenylation to avoid nuclear retention [62,67] (reviewed in [26]).

In addition to forming intramolecular IR*Alu*s within the 3′ UTRs, nonperfect intermolecular IR*Alu*s have been observed between lncRNAs and 3′ UTRs of mRNAs. Such intermolecular dsRNA pairing is associated with STAU proteins, which triggers Staufen-mediated mRNA decay (SMD) (Figure 3G) [28,68,69] (reviewed in [16]).

circRNA Formation Mediated by Intronic IR*Alu*s

Nearly half of *Alu* repeats are located in introns [27]. Since most of these intronic *Alu* sequences are removed by splicing, it had been thought that intronic *Alu* RNAs could play, at best, limited roles in gene expression through AS and RNA editing. However, recent studies have demonstrated that intronic IR*Alu*s are highly associated with the formation of thousands of circRNAs from back-spliced exons of precursor RNAs [27].

Different from normal splicing that ligates upstream (5′) splice donor sites with downstream (3′) splice acceptor sites, back splicing occurs in reverse and ligates a downstream splice donor site with an upstream splice acceptor site, resulting in covalently closed circRNA transcripts and alternatively spliced linear RNAs with skipped exons [27,70] (reviewed in [71–73]). Although positive correlation between exon skipping and the production of circRNAs that contain the skipped exons has been observed, not all skipped exons can produce circRNAs [27,70]. Back splicing is generally coupled and competed with canonical splicing [27,74]. Although requiring canonical spliceosomal machinery [74,75], back splicing is unfavorable and thus less efficiently catalyzed by the spliceosomal machinery, leading to less circRNA formation than its linear mRNA counterpart [76]. Remarkably, RNA pairing formed across introns that flank back-spliced exons can bring the distal splice sites into close proximity to facilitate circRNA biogenesis in humans [27] (Figure 3H). Correspondingly, eliminating RNA pairs significantly reduces circRNA formation [27,76–78].

RNA pairing across flanking introns is efficient, but may not be sufficient, to promote back splicing. Most of these RNA pairs are formed by intronic IR*Alu*s in humans [27,79]. Interestingly, RNA pairing formed by IR*Alu*s across flanking introns competes with IR*Alu*s within an individual intron in the same gene locus, leading to competition between back splicing for circRNAs and canonical splicing for linear RNAs [27] (Figure 3I). In addition, competition of IR*Alu*s across different pairs of flanking introns results in multiple circRNAs in the same gene locus through alternative back splicing [80] (Figure 2N).

Protein factors are also involved in back-splicing regulation, either positively or negatively. Quaking (QKI) and muscleblind-like (MBNL)1 proteins can bridge introns that flank circRNA-forming exons to induce back splicing [74,81]. ADAR1 was reported to suppress circRNA

biogenesis and ADAR1 knockdown could upregulate some circRNA expression [77]. The regulation of ADAR1 on circRNA biogenesis is possibly associated with A-to-I RNA editing of IR*Alu*s. In the wild-type condition, high A-to-I editing in an IR*Alu*s across flanking introns could diminish its pairing capacity and therefore inhibit back splicing; whereas in the condition of ADAR1 knockdown, the pairing of IR*Alu*s across flanking introns might be more stable with low A-to-I editing, thus favoring back splicing for circRNA production [77]. However, one cannot exclude the possibility that ADAR1 may regulate circRNA formation directly through its dsRNA binding activity, independent of RNA editing [82].

Broad Impact of Promiscuous A-to-I Editing on *Alu* RNA-Involved Regulation

A-to-I RNA editing is the most predominant form of RNA modification in higher eukaryotes, catalyzed by enzymes of the ADAR family (reviewed in [83,84]). In addition to site-specific editing in several pre-mRNAs, the majority of A-to-I RNA editing occurs promiscuously within dsRNAs of at least 100 base pairs in length; especially in IR*Alu* pairs formed by orientation-opposite embedded *Alu* RNAs [85,86]. Extensive deep-sequencing analyses have shown that virtually all A bases within IR*Alu*s can undergo A-to-I editing, but most at low levels (<1%) [87,88]. Currently, >2.5 million editing sites have been included in the RADAR database [89], among which, >95% are located in *Alu* regions [90].

A-to-I RNA editing has broad impacts on transcribed *Alu* RNAs at both primary sequence and secondary structure levels. On the one hand, since the resulting inosine functions as if it were guanosine (G), A-to-I RNA editing results in genetic recoding of RNA (I is read as G by the translation machinery) and affects AS selection and alternation of RNAi inhibition (reviewed in [26,91]). On the other hand, changes from A-U pairing to I-U mismatches caused by A-to-I RNA editing can alter the formation and stability of IR*Alu*s secondary structures, and thus have an influence on IR*Alu*s-mediated circRNA biogenesis and RNA nuclear retention. Other than these effects on gene regulation, IR*Alu*s may also function as a *cis*-regulator for site-selective A-to-I editing by recruiting ADAR enzymes to neighboring short hairpin regions for catalysis, which is consistent with the observation that site-selective A-to-I editing sites are often close to hyper-edited IR*Alu*s [88,92]. The *cis* regulation of primate-specific editing by IR*Alu*s within coding regions or UTRs of some transcription factors could alter the whole transcriptome in a primate-specific manner [92]. However, it is also possible that the binding of ADAR enzymes to *Alu*s/IR*Alu*s may function as ADAR sponges to deplete them from short hairpin structures involved in site-specific A-to-I RNA editing.

Nevertheless, although many *Alu*-related biological effects could be influenced by A-to-I editing, it is still possible that *Alu* editing may serve no particular purpose, and simply result from the abundance of *Alu* RNAs embedded in transcripts [26]. This is supported by the observation that most *Alu* editing occurs at a low level (<1%) [87]. Finally, another possibility is that the structure of *Alu*s/IR*Alu*s themselves is crucial for the observed impacts on gene regulation, especially at sites with low (<1%) A-to-I RNA editing.

Concluding Remarks

In addition to the astonishing number of >1 million dimeric *Alu* copies within the current human reference genome [1], *Alu* elements continue inserting into the modern human lineage [10], although at a low rate. It has been estimated that there is about one new *Alu* insertion per 21 human births [13], and thousands of new *Alu* insertions have been identified from a small number of individuals [9]. It is also worthwhile noting that the most current analyses on *Alu* impacts on biology are mainly focused on fixed *Alu* insertions in germ lines. However, *Alu* retrotransposon might be active in somatic tissues that continues to affect gene expression and even causes diseases, such as cancer, after birth (reviewed in [8]). Thus, it will be of interest to comprehensively scrutinize how *Alu* insertion reshapes our genome and transcriptome in

Outstanding Questions

How often are new *Alu* insertions found in individuals or populations?

What are the consequences of frequent insertion of *Alu*s in the primate brain?

What are the evolutionary advantages and disadvantages of primate-specific *Alu* insertion?

What additional roles can *Alu*s play in gene regulation?

What are the distinctions and commonalities between the different functions of *Alu*s?

different tissues and during the lifespan in a primate-specific manner. While the impacts of some *Alu* repeats on the human genome have been affirmatively revealed by recent studies, the influence of other less-characterized *Alu*s and their specific underlying mechanisms are still awaiting to be investigated. For instance, even a single point mutation in the *LINE/Alu* overlapped sequence of a human lncRNA could lead to lethal infantile encephalopathy [93]. Collectively, the widespread *Alu* elements largely increase the complexity of gene expression and the plasticity of the human genome (See Outstanding Questions).

Acknowledgments

We are grateful to Gordon Carmichael for critical reading of this manuscript. We apologize to authors whose work could not cite here owing to space/content limitations. Our work is supported by grants 2016YFA0100701 and 2014CB910601 from MOST , XDB19000000 from CAS, and 91540115 and 91440202 from NSFC.

References

1. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
2. Quentin, Y. (1992) Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic acids research* 20, 487–493
3. Mighell, A.J. *et al.* (1997) Alu sequences. *FEBS Lett.* 417, 1–5
4. Batzer, M.A. and Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nature reviews. Genetics* 3, 370–379
5. Kriegs, J.O. *et al.* (2007) Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.* 23, 158–161
6. Hasler, J. *et al.* (2007) Useful 'junk': Alu RNAs in the human transcriptome. *Cell. Mol. Life Sci.* 64, 1793–1800
7. Hormozdiari, F. *et al.* (2013) Rates and patterns of great ape retrotransposition. *Proc. Natl. Acad. Sci. USA* 110, 13457–13462
8. Deininger, P. (2011) Alu elements: know the SINEs. *Genome Biol.* 12, 236
9. Hormozdiari, F. *et al.* (2011) Alu repeat discovery and characterization within human genomes. *Genome Res.* 21, 840–849
10. Stewart, C. *et al.* (2011) A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* 7, e1002236
11. Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature* 284, 604–607
12. Orgel, L.E. *et al.* (1980) Selfish DNA. *Nature* 288, 645–646
13. Xing, J. *et al.* (2009) Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* 19, 1516–1526
14. Deininger, P.L. and Batzer, M.A. (1999) Alu repeats and human disease. *Mol. Genet. Metab.* 67, 183–193
15. Belancio, V.P. *et al.* (2008) Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 18, 343–358
16. Elbarbary, R.A. *et al.* (2016) Retrotransposons as regulators of gene expression. *Science* 351, aac7247
17. Versteeg, R. *et al.* (2003) The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* 13, 1998–2004
18. Lev-Maor, G. *et al.* (2003) The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. *Science* 300, 1288–1291
19. Kreahling, J. and Graveley, B.R. (2004) The origins and implications of Aluternative splicing. *Trends Genet.* 20, 1–4
20. Roy-Engel, A.M. *et al.* (2005) Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* 110, 365–371
21. An, H.J. *et al.* (2004) The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3′ untranslated regions. *BMC Genomics* 5, 97
22. Hasler, J. and Strub, K. (2006) Alu RNP and Alu RNA regulate translation initiation *in vitro*. *Nucleic Acids Res.* 34, 2374–2385
23. Hasler, J. and Strub, K. (2006) Alu elements as regulators of gene expression. *Nucleic Acids Res.* 34, 5491–5497
24. Sobczak, K. and Krzyzosiak, W.J. (2002) Structural determinants of BRCA1 translational regulation. *J. Biol. Chem* 277, 17349–17358
25. Sen, S.K. *et al.* (2006) Human genomic deletions mediated by recombination between Alu elements. *Am. J. Hum. Genet.* 79, 41–53
26. Chen, L.L. and Carmichael, G.G. (2008) Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle* 7, 3294–3301
27. Zhang, X.O. *et al.* (2014) Complementary sequence-mediated exon circularization. *Cell* 159, 134–147
28. Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 470, 284–288
29. Xie, H. *et al.* (2009) High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. *Nucleic Acids Res.* 37, 4331–4340
30. Polak, P. and Domany, E. (2006) Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7, 133
31. Su, M. *et al.* (2014) Evolution of Alu elements toward enhancers. *Cell Rep.* 7, 376–385
32. Tang, R.B. *et al.* (2005) Increased level of polymerase III transcribed Alu RNA in hepatocellular carcinoma tissue. *Mol. Carcinog.* 42, 93–96
33. Mariner, P.D. *et al.* (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* 29, 499–509
34. Chu, W.M. *et al.* (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell. Biol.* 18, 58–68
35. Rubin, C.M. *et al.* (2002) Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res.* 30, 3253–3261
36. Bovia, F. *et al.* (1997) The SRP9/14 subunit of the human signal recognition particle binds to a variety of Alu-like RNAs and with higher affinity than its mouse homolog. *Nucleic Acids Res.h* 25, 318–326
37. Pagano, A. *et al.* (2007) New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet.* 3, e1
38. Bentwich, I. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* 37, 766–770
39. Spengler, R.M. *et al.* (2014) Functional microRNAs and target sites are created by lineage-specific transposition. *Hum. Mol. Genet.* 23, 1783–1793
40. Borchert, G.M. *et al.* (2006) RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* 13, 1097–1101
41. Gu, T.J. *et al.* (2009) Alu-directed transcriptional regulation of some novel miRNAs. *BMC Genomics* 10, 563
42. Du, Z.Q. *et al.* (2013) Novel microRNA families expanded in the human genome. *BMC Genomics* 14, 98

43. Lehnert, S. *et al.* (2009) Evidence for co-evolution between human microRNAs and Alu-repeats. *PloS One* 4, e4456
44. Noguer-Dance, M. *et al.* (2010) The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Hum. Mol. Genet.* 19, 3566–3582
45. Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463
46. Sorek, R. *et al.* (2002) Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067
47. Shen, S. *et al.* (2011) Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl. Acad. Sci. USA* 108, 2837–2842
48. Lei, H. *et al.* (2005) Exonization of AluYa5 in the human ACE gene requires mutations in both 3′ and 5′ splice sites and is facilitated by a conserved splicing enhancer. *Nucleic Acids Res.* 33, 3897–3906
49. Lei, H. and Vorechovsky, I. (2005) Identification of splicing silencers and enhancers in sense Alus: a role for pseudoacceptors in splice site repression. *Mol. Cell. Biol.* 25, 6912–6920
50. Knebelmann, B. *et al.* (1995) Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum. Mol. Genet.* 4, 675–679
51. Nekrutenko, A. and Li, W.H. (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* 17, 619–621
52. Stuart, J.J. *et al.* (2000) The 3′ UTR of human MnSOD mRNA hybridizes to a small cytoplasmic RNA and inhibits gene expression. *Biochem. Biophys. Res. Commun.* 274, 641–648
53. Goodyer, C.G. *et al.* (2001) Alu elements in human growth hormone receptor gene 5′ untranslated region exons. *J. Mol. Endocrinol.* 27, 357–366
54. Landry, J.R. *et al.* (2001) Repetitive elements in the 5′ untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. *Genomics* 76, 110–116
55. Daskalova, E. *et al.* (2006) 3′UTR-located ALU elements: donors of potential miRNA target sites and mediators of network miRNA-based regulatory interactions. *Evol. Bioinform. Online* 2, 103–120
56. Smalheiser, N.R. and Torvik, V.I. (2006) Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.* 22, 532–536
57. Hoffman, Y. *et al.* (2013) The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery. *Bioinformatics* 29, 894–902
58. Hoffman, Y. *et al.* (2014) microRNAs and Alu elements in the p53-Mdm2-Mdm4 regulatory network. *J. Mol. Cell Biol.* 6, 192–197
59. Lee, J.Y. *et al.* (2008) Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3′-end of genes. *Nucleic Acids Res.* 36, 5581–5590
60. Chen, C. *et al.* (2009) Using Alu elements as polyadenylation sites: a case of retroposon exaptation. *Mol. Biol. Evol.* 26, 327–334
61. DeCerbo, J. and Carmichael, G.G. (2005) SINEs point to abundant editing in the human genome. *Genome Biol.* 6, 216
62. Chen, L.L. *et al.* (2008) Alu element-mediated gene silencing. *EMBO J.* 27, 1694–1705
63. Hu, S.B. *et al.* (2015) Protein arginine methyltransferase CARM1 attenuates the paraspeckle-mediated nuclear retention of mRNAs containing IRAlus. *Genes Dev.* 29, 630–645
64. Elbarbary, R.A. *et al.* (2013) STAU1 binding 3′ UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev.* 27, 1495–1510
65. Elbarbary, R.A. and Maquat, L.E. (2015) CARMing down the SINEs of anarchy: two paths to freedom from paraspeckle detention. *Genes Dev.* 29, 687–689
66. Chen, L.L. and Carmichael, G.G. (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* 35, 467–478
67. Prasanth, K.V. *et al.* (2005) Regulating gene expression through RNA nuclear retention. *Cell* 123, 249–263
68. Gleghorn, M.L. *et al.* (2013) Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat. Struct. Mol. Biol.* 20, 515–524
69. Park, E. *et al.* (2013) Staufen2 functions in Staufen1-mediated mRNA decay by binding to itself and its paralog and promoting UPF1 helicase but not ATPase activity. *Proc. Natl. Acad. Sci. USA* 110, 405–412
70. Kelly, S. *et al.* (2015) Exon skipping is correlated with exon circularization. *J Mol Biol.* 427, 2414–2417
71. Chen, L.L. (2016) The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 17, 205–211
72. Chen, L.L. and Yang, L. (2015) Regulation of circRNA biogenesis. *RNA Biol.* 12, 381–388
73. Yang, L. (2015) Splicing noncoding RNAs from the inside out. *Wiley Interdiscip. Rev. RNA* 6, 651–660
74. Ashwal-Fluss, R. *et al.* (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66
75. Starke, S. *et al.* (2015) Exon circularization requires canonical splice signals. *Cell Rep.* 10, 103–111
76. Zhang, Y. *et al.* (2016) The biogenesis of nascent circular RNAs. *Cell Rep.* 15, 611–624
77. Ivanov, A. *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 10, 170–177
78. Liang, D. and Wilusz, J.E. (2014) Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* 28, 2233–2247
79. Dong, R. *et al.* (2017) Increased complexity of circRNA expression during species evolution. *RNA Biol.* http://dx.doi.org/10.1080/15476286.2016.1269999
80. Zhang, X.O. *et al.* (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287
81. Conn, S.J. *et al.* (2015) The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134
82. Chen, T. *et al.* (2015) ADAR1 is required for differentiation and neural induction by regulating microRNA processing in a catalytically independent manner. *Cell Res.* 25, 459–476
83. Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846
84. Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79, 321–349
85. Ramaswami, G. *et al.* (2012) Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods* 9, 579–581
86. Zhu, S. *et al.* (2013) Prediction of constitutive A-to-I editing sites from human transcriptomes in the absence of genomic sequences. *BMC Genomics* 14, 206
87. Bazak, L. *et al.* (2014) A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* 24, 365–376
88. Bazak, L. *et al.* (2014) Genome-wide analysis of Alu editability. *Nucleic Acids Res.* 42, 6876–6884
89. Ramaswami, G. and Li, J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* 42, D109–D113
90. Ramaswami, G. and Li, J.B. (2016) Identification of human RNA editing sites: a historical perspective. *Methods* 107, 42–47
91. Nishikura, K. (2016) A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* 17, 83–96
92. Daniel, C. *et al.* (2014) Alu elements shape the primate transcriptome by *cis*-regulation of RNA editing. *Genome Biol.* 15, R28
93. Cartault, F. *et al.* (2012) Mutation in a primate-conserved retrotransposon reveals a noncoding RNA as a mediator of infantile encephalopathy. *Proc. Natl. Acad. Sci. USA* 109, 4980–4985