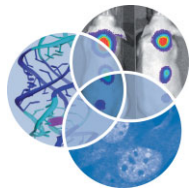
Advanced Review



SINEs

Dmitri A. Kramerov* and Nikita S. Vassetzky

Short interspersed elements (SINEs) are mobile genetic elements that invade the genomes of many eukaryotes. Since their discovery about 30 years ago, many gaps in our understanding of the biology and function of SINEs have been filled. This review summarizes the past and recent advances in the studies of SINEs. The structure and origin of SINEs as well as the processes involved in their amplification, transcription, RNA processing, reverse transcription, and integration of a SINE copy into the genome are considered. Then we focus on the significance of SINEs for the host genomes. While these genomic parasites can be deleterious to the cell, the long-term being in the genome has made SINEs a valuable source of genetic variation providing regulatory elements for gene expression, alternative splice sites, polyadenylation signals, and even functional RNA genes. *WIREs RNA* 2011 2 772–786 DOI: 10.1002/wrna.91

INTRODUCTION

Even before the introduction of cloning techniques, DNA reassociation kinetics experiments indicated repetitive nucleotide sequences interspersed in the genomes of most eukaryotes. These repetitive sequences included both long (kb) and short (hundreds of bp) elements. Cloning and sequencing of interspersed genomic sequences demonstrated that most of them are mobile genetic elements. Some of them, DNA transposons, retrotransposons with long terminal repeats (LTRs), and long interspersed elements (LINEs; Box 1) were 3–7 kb long, while the length of short interspersed elements (SINEs) did not exceed 600 bp. The first described SINEs were mouse B1 and B2[1,2] and human Alu.[3]

Eukaryotic genomes can contain more than a million copies of a SINE species. Most copies are not identical and their nucleotide sequences vary by up to 35%. Such sequences constitute a family, and genomes can harbor several SINE families. To date, more than 120 SINE families have been described; the bulk of them (66%) have the length from 150 to 300 bp (Figure 1). SINEs have been found in mammals of all studied orders (Table 1 summarizes SINEs in placentals); they have also been identified in the genomes of many fish and some reptilian species as well as in certain invertebrates including sea squirts, cephalopods, and insects; finally, a number

BOX 1

LINEs

LINEs are a class of mobile genetic elements amplified via retrotransposition.[4,5] In contrast to LTR retrotransposons, LINEs have no long terminal repeats. Sequences of LINEs from 3 to 7 kb usually contain two open reading frames (ORF). ORF1 encodes a protein with RNA binding and nucleic acid chaperone activities (similar to retroviral Gag). ORF2 encodes the reverse transcriptase (RT) and, commonly, RNase H and endonuclease. LINEs are transcribed by RNA polymerase II using a promoter that resides inside the LINE. The resulting RNA is reverse transcribed using the LINE-encoded activities, which introduces a new LINE copy in the genome. To date, dozens of LINE families falling into 17 clades have been described.[6–8] LINEs are widespread in eukaryotic genomes from invertebrates to human but are less common in unicellular eukaryotes.[9] LINE1 (L1) is one of the best studied LINEs.[10] L1 is found in the genomes of all mammals, where it can number in hundreds of thousands. Most genomic copies of LINEs are 5′-truncated, and only a minor fraction of copies (autonomous LINEs) encode functional proteins. Some LINE copies encoding no functional proteins can be transcribed and reverse transcribed, and these are called nonautonomous LINEs.

*Correspondence to: kramerov@eimb.ru

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

DOI: 10.1002/wrna.91

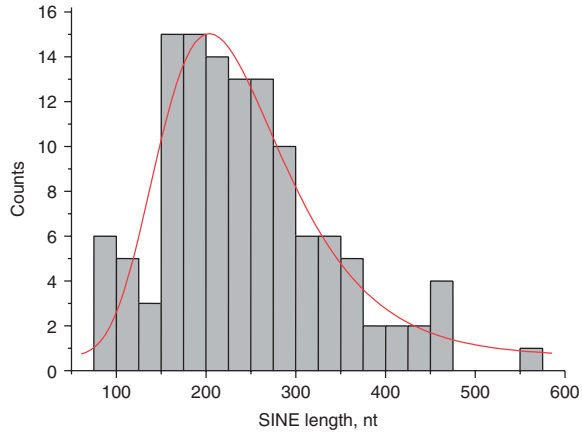


FIGURE 1 | Length distribution for 122 eukaryotic short interspersed elements (without tail).

of SINE families are known in flowering plants.[11,12] At the same time, no SINEs have been found in unicellular eukaryotes; they are also missing from the *Drosophila* genus.

In contrast to the rest of mobile elements transcribed by RNA polymerase II (pol II), SINEs rely on RNA polymerase III (pol III) transcription initiated at the internal promoter. SINEs code no proteins, and their reverse transcription is carried out by the enzyme encoded by a partner LINE. This process called retroposition gives rise to new genomic copies of SINEs (hence, they are also called short retroposons).

STRUCTURE, ORIGIN, AND EVOLUTION OF SINEs

A typical SINE consists of three parts: 5′-terminal 'head', 'body', and 3′-terminal 'tail' (Figure 2). The heads of all SINEs demonstrate a clear similarity with one of the three types of cellular RNAs synthesized by pol III: tRNA, 7SL RNA, or 5S rRNA. This similarity indicates that SINE heads originated from these cellular RNAs, most likely, through the intermediate stage of RNA retropseudogenes. tRNA-derived SINEs are particularly abundant (Figure 3), although the particular tRNA species of origin cannot be always identified due to numerous substitutions in the SINE evolution. 7SL RNA-derived SINEs known to date are limited to rodents, primates, and tree shrews.[13–15] Cellular 7SL RNA (300 nt) is found in all eukaryotes as a component of the signal recognition particle, which targets specific proteins to the endoplasmic reticulum. 7SL RNA-derived SINEs have a central deletion in the corresponding region covering about a half of the original RNA length. SINE families descending from 5S rRNA are also not abundant; they are found in fishes[16,17] and in a few mammals, fruit bats,[18] and springhare.[19]

All SINEs as well as the RNA genes that gave rise to them possess an internal pol III promoter. The promoter in tRNA- and 7SL RNA-derived SINEs includes 11-bp boxes A and B at a distance of 30–35 bp, while in 5S RNA-derived SINEs, the promoter is composed of boxes A, IE, and C (Figure 2). The internal pol III promoter is indispensable for SINEs because it is preserved in new SINE copies, thus making possible their transcription.

The body of SINEs is usually unique for each SINE family and its origin is largely unclear (Figure 2); this pattern is particularly common in mammals. At the same time, a part of the body can contain domains shared by distant SINE families. To date, four such domains have been described: CORE domain in vertebrates,[20] V-domain in fishes,[21] Deu-domain in deuterostomes,[17] and Ceph-domain in cephalopods.[22]

On the other hand, in a substantial fraction of SINEs, the 3′-terminal portion of the body (30–100 bp) originates from the 3′-terminal sequence of a partner LINE[23,24]; this pattern is found in most fish SINEs but also occurs in other groups including mammals (Figure 3). In most (but not all) LINEs, the 3′-terminal portion of their RNA is recognized by their RT, and SINE RNA also needs this sequence to be reverse transcribed by such enzymes.[25]

All SINEs have a 3′-terminal tail composed of repeated sequences of 1–8 nt or irregular A-rich sequence, while some plant SINEs have T-rich tails.

Yet, the structure of some SINE families can be simpler or more complex. For instance, some SINEs have no body (this applies to all known 7SL RNA-derived SINEs) and consist of a head and a tail, thus resembling RNA pseudogenes (Figure 2(e)). However, such simple SINEs have specific nucleotide substitutions distinguishing them from the RNA of origin, which indicates that they directly descend not from the RNA gene but from a SINE copy with such substitutions.

On the other hand, SINE copies can combine into dimers (or even trimers) that are further amplified in this complex form (Figure 2(i)). Such complex SINEs can combine representatives of the same (e.g., 7SL RNA-derived Alu in primates or tRNA-derived CYN in colugos) or different families. Heterodimeric SINE structures described to date include tRNA/tRNA, 7SL RNA/tRNA, and 5S rRNA/tRNA hybrids, while no 7SL RNA/5S rRNA hybrids are known (Table 1).

As distinct from other mobile genetic elements, SINE families emerged *de novo* many times in

TABLE 1 | Short Interspersed Elements (SINEs) of Placentals

SINE Family Name(s)	Length (w/o Tail)	Taxon	Cell RNA of Origin	Complexity	LINE [Clade]	Tail	Features
AfroLA	158	Afrotheria	tRNA	Monomer	Bov-B [RTE]	T	Bipartite LINE region
AfroSINE	230	Afrotheria	tRNA	Monomer	Bov-B [RTE]	TTG	Bipartite LINE region
Alu	282	Primates	7SL RNA	Dimer	L1 [L1]	A	
B1	135	Rodentia	7SL RNA	Monomer	L1 [L1]	A	Quasidimer
B1-dID	200	Sciuridae and Gliridae	7SL RNA/tRNA	Dimer	L1 [L1]	A	
B2	185	Muridae, Cricetidae, Spalacidae, and Rhizomiidae	tRNA	Monomer	L1 [L1]	A	T^+
B4 (RSINE2)	275	Mouse and rat	tRNA/7SL RNA	Dimer	L1 [L1]	A	
Bov-tA	210	Ruminantia	$tRNA^{Gly}$	Monomer	Bov-B [RTE]	AT	Bipartite LINE region
C	309	Lagomorpha	$tRNA^{Gly}$	Monomer	L1 [L1]	A	$(TC)_n\ T^+$
CAN	160	Carnivora	$tRNA^{Lys}$	Monomer	L1 [L1]	A	$(TC)_n\ T^+$
CHR-2	270	Cetacea, Hippopotamidae, and Ruminantia	$tRNA^{Glu}$	Monomer	L1 [L1]	A	T^+
CHRS (CHR-1; CHRS-S)	120	Cetacea, Hippopotamidae, Ruminantia, and Suidae	$tRNA^{Glu}$	Monomer	L1 [L1]	A	
CYN-I (t-SINE)	90	Dermoptera	$tRNA^{Ile}$	Monomer	L1 [L1]	A	Simple
CYN-II	160	Dermoptera	$tRNA^{Ile}$	Dimer	L1 [L1]	A	
CYN-III	220	Dermoptera	$tRNA^{Ile}$	Trimer	L1 [L1]	A	
DAS-I	90	Dasypodidae	$tRNA^{Ala}$	Monomer	L1 [L1]	A	
DAS-II	190	Dasypodidae	$tRNA^{Ala}$	Dimer	L1 [L1]	A	T^+ (DAS-IIb)
DAS-III	440	Dasypodidae	$tRNA^{Ala}$	Trimer	L1 [L1]	A	T^+
DIP	190	Dipodidae and Zapodidae	$tRNA^{Ala}$	Monomer	L1 [L1]	A	T^+
ERE-1	212	Perissodactyla	$tRNA^{Ser}$	Monomer	L1 [L1]	A	T^+
ERI-1	126	Erinaceidae	$tRNA^{Lys}$	Monomer	L1 [L1]	A	T^+
ERI-2	186	Erinaceidae	tRNA	Monomer	L1 [L1]	A	T^+
ID	75	Rodentia	$tRNA^{Ala}$	Monomer	L1 [L1]	A	Simple
IDL-Geo	192	Geomyidae and Heteromyidae	$tRNA^{Ala}$	Dimer	L1 [L1]	A	
MEG-RL	197	Megachiroptera	5S rRNA	Monomer	L1 [L1]	A	
MEG-RS	120	Megachiroptera	5S rRNA	Monomer	L1 [L1]	A	Simple
MEG-T2	252	Megachiroptera	tRNA	Monomer	L1 [L1]	A	
MEG-TR	206	Megachiroptera	tRNA/5S rRNA	Monomer	L1 [L1]	A	
MEN	259	Menetes and Callosciurus	$tRNA^{Ala}$/7SL RNA	Dimer	L1 [L1]	A	
MyrSINE	83	Myrmecophagidae	$tRNA^{Gly}$	Monomer	L1 [L1]	A	Simple

TABLE 1 | Continued

SINE Family Name(s)	Length (w/o Tail)	Taxon	Cell RNA of Origin	Complexity	LINE [Clade]	Tail	Features
Ped-1	256	Pedetidae	5S rRNA	Monomer	Bov-B [RTE]	CAA	Bipartite LINE region
Ped-2	185	Pedetidae	$tRNA^{Ala}$	Monomer	Bov-B [RTE]	CAA	Bipartite LINE region
PRE-1	246	Suidae and Tayassuidae	$tRNA^{Arg}$	Monomer	L1 [L1]	A	
Rhin-1	190	Rhinolophidae and Hipposideridae	$tRNA^{Ile}$	Monomer	L1 [L1]	A	T^{+}
SINE type II (GarnAlu)	260	Lorisiformes	tRNA/7SL RNA	Dimer	L1 [L1]	A	
SINE type III (Garnel1)	78	Lorisiformes	tRNA	Monomer	L1 [L1]	A	Simple
SOR	157	Soricidae	$tRNA^{Ala}$	Monomer	L1 [L1]	A	
TAL	237	Talpidae	tRNA	Monomer	L1 [L1]	A	T^{+}
Tu type I	179	Scandentia	tRNA/7SL RNA	Dimer	L1 [L1]	A	
Tu type II	278	Scandentia	tRNA/7SL RNA/7SL RNA	Trimer	L1 [L1]	A	
Tu type III	242	Scandentia	tRNA	Monomer	L1 [L1]	A	$(R)_n$–$(Y)_n$
VES	191	Vespertilionidae, Molossidae, Phyllostomidae, and Emballonuridae	RNA^{Tyr}	Monomer	L1 [L1]	A	$(TC)_n$ T^{+}
vic-1	117	Camelidae	$tRNA^{Ala}$	Monomer	L1 [L1]	A	Simple

evolution (e.g., at least 23 times during mammalian evolution). Progressive evolution made them more complex and the descendant SINE families were often more successful. SINEs are transmitted vertically from parent to offspring species, and their copies are maintained in the genomes over tens or even hundreds of millions of years, which made SINEs a convenient and reliable phylogenetic marker (Box 2).

BOX 2

SINEs AS A TOOL IN PHYLOGENETIC STUDIES

Although individual SINE copies can have a function in gene activity, the great majority of SINEs are not functional. Overall, SINEs are transmitted vertically but not horizontally, and the probability of independent emergence of the same SINE families in unrelated species is negligible. Being essentially neutral characters, SINEs are not subject to convergent evolution, which makes them excel over morphological characters and gene and protein sequences. Generally, there are two approaches to using SINEs in phylogenetic studies.

The *method of families* consists in the identification of species containing or lacking a SINE family in their genomes (by dot hybridization and/or PCR). All species having SINEs of a particular family are considered to be phylogenetically closer to each other than to other species. This approach was used to test the phylogenetic relationships between certain rodent families[26,27]; supported the relationship among rodents, primates, and tree shrews[14,17]; and confirmed the mammalian clade Afrotheria.[28]

The *method of insertions* tests the presence of a SINE copy in a particular genomic locus of different species.[29,30] This test involves PCR with primers targeted to the SINE flanking regions, and the PCR product is shorter if the locus contains no SINE. Species sharing a SINE copy in the locus are considered to be phylogenetically closer to each other than to other species. This approach was applied to elucidate the relationships between human populations,[31] major rodent lineages,[32] or salmon species[33] and to demonstrate the phylogenetic relationship between cetaceans and artiodactyls.[34]

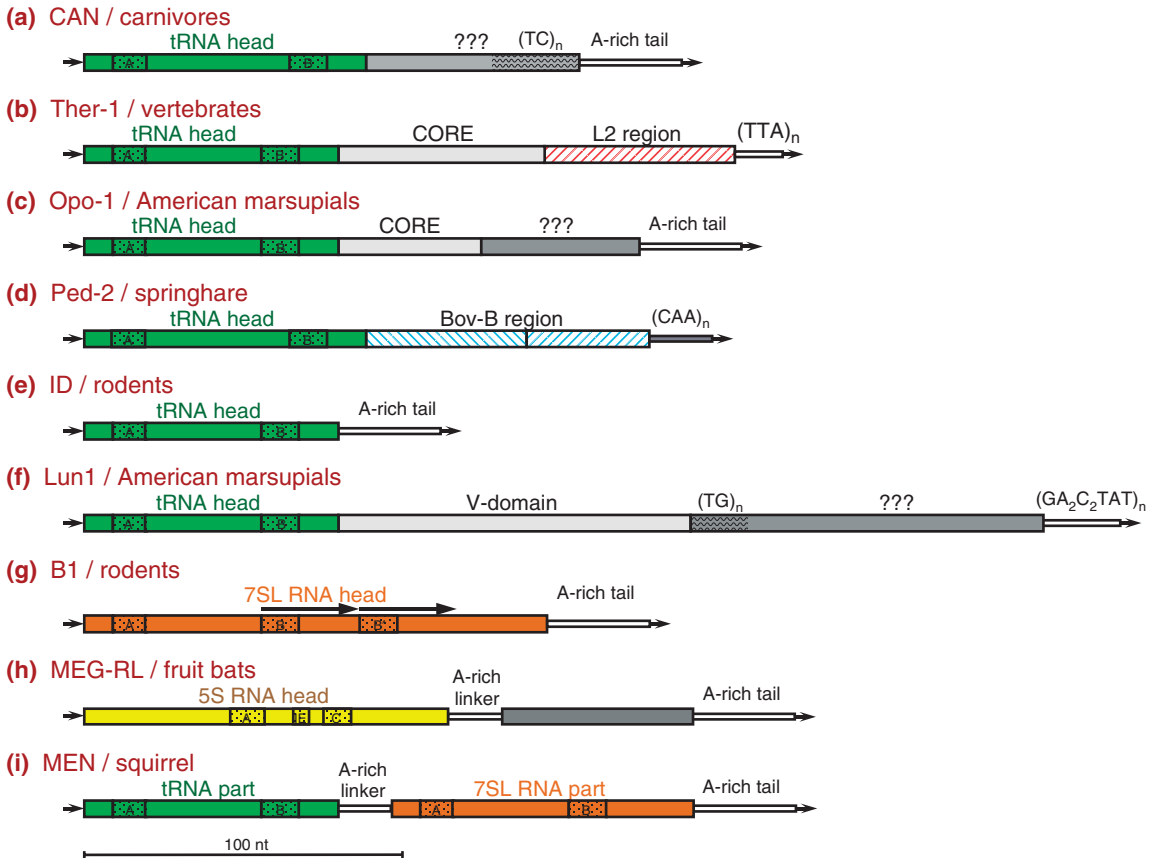


FIGURE 2 | Short interspersed element (SINE) structure examples. (a) CAN, tRNA-derived SINE with a unique region of unknown origin and a $(TC)_n$ stretch; (b) Ther-1, tRNA-derived SINE with a CORE domain and a LINE-derived region; (c) Opo-1, tRNA-derived SINE with a CORE domain and a unique region; (d) Ped-2, tRNA-derived SINE with a bipartite LINE-derived region; (e) ID, simple tRNA-derived SINE; (f) Lun-1, tRNA-derived SINE with a V-domain, a $(TG)_n$ stretch, and a unique region; (g) B1, 7SL RNA-derived SINE with an internal duplication; (h) MEG-RL, 5S rRNA-derived SINE with a unique region; (i) MEN, dimeric tRNA/7SL RNA-derived SINE. Boxes with dotted background correspond to pol III promoter regions; '???' corresponds to *body* parts of unknown origin; direct repeats including terminal target site duplications are indicated by arrows.

SINE TRANSCRIPTION

SINEs are abundant both in intergenic regions and within genes. They are particularly ample in introns, and most mammalian genes contain intronic SINE copies. Exonic copies of SINEs also occur, largely, in the 3′-untranslated regions (UTRs). Accordingly, numerous SINE copies are transcribed by pol II within pre-mRNAs, and nuclear RNA is rich in SINE sequences; however, their content in mRNA sharply decreases after splicing. Intronic copies of SINEs occur in both orientations; as a result, relatively remote copies in opposite orientations can form hairpin structures in pre-mRNA. Such double-stranded pre-mRNA regions made it possible to discover and describe the first SINE families.[1,35] SINE sequences can substantially contribute to the secondary structure of pre-mRNAs. Recent data indicate the involvement of such hairpins in splicing control of pre-mRNAs that contain them.[36] Pol II-synthesized SINE RNA seems to have no role in SINE retroposition, which relies on pol III transcripts of SINEs.

Many but by no means all SINE copies can be transcribed by pol III. Necessarily, transcribable SINE copies should have a functionally intact pol III promoter. The upstream genomic sequence also seems to have an impact on the transcriptional competence.[37,38] In the case of the tRNA- and 7SL RNA-derived SINEs, the transcription initiation starts with the binding of transcription factor for polymerase III C (TFIIIC) to the promoter boxes A and B. This complex recruits another pol III transcription factor, TFIIIB, which assembles pol III at the start site of transcription: the beginning of SINE sequence 10–12 nt upstream of box A. In the case of 5S rRNA-derived SINEs, the transcription initiation is primed by the binding of TFIIIA to the promoter box C, which recruits TFIIIC and initiation proceeds as described above.[39] Pol III transcription is stopped at

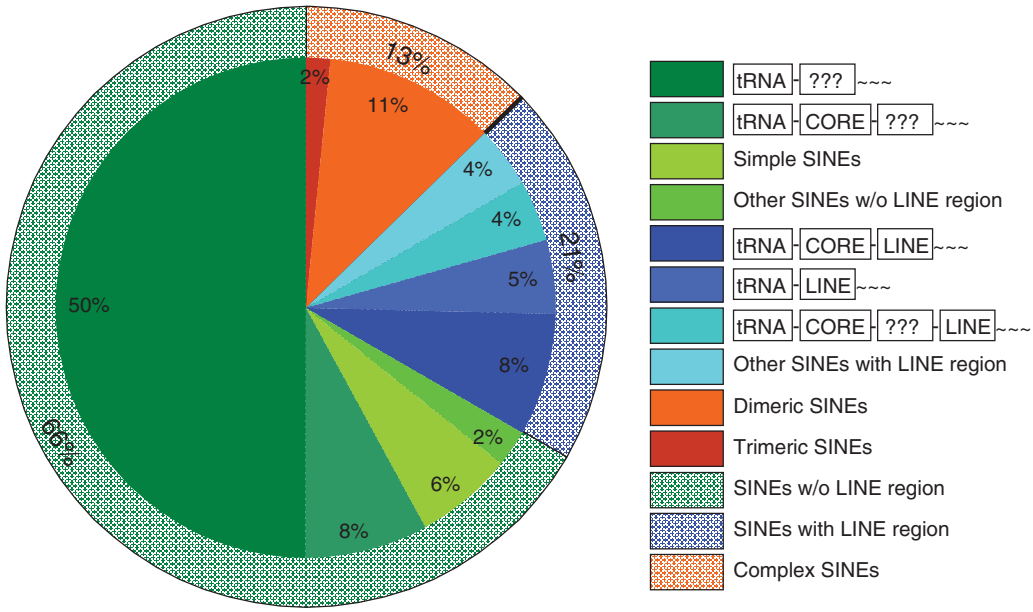


FIGURE 3 | Occurrence of short interspersed element structures. [tRNA] is tRNA-derived *head*; [???] corresponds to *body* parts of unknown origin; [CORE] is CORE, Deu, V, or Ceph domains; [LINE] is long interspersed element-derived *body* region; and '~~~' denotes the *tail*.

the terminator, a stretch of at least four Ts (TCTTT also acts as a terminator). Interestingly, sequences of most SINE families contain no terminator, and their transcription continues to a random T stretch downstream of the SINE sequence (Figure 4). The transcription of numerous SINE copies gives rise to transcripts of variable length usually not exceeding hundreds of nucleotides, because short T stretches are not uncommon in the genome. On the other hand, many (about a third in mammals) SINE families have terminators at their 3′ end. In this case, the transcript length corresponds to that of the transcribed SINE. Similar terminators can be found in a few plant and fish SINEs.[40–42]

The tissue specificity of pol III transcription of SINEs remains underexplored. In somatic tissues, it is either blocked (B1 and Alu) or greatly suppressed (B2). Heat shock and other stress stimuli can induce SINE transcription.[43] Relatively high levels of SINE transcripts were found in the murine testis, zygote, and embryos at the earliest stages of development.[44,45] Many malignant tumors also feature high levels of SINE transcription.[46,47] High levels of SINE transcription are often linked to DNA demethylation.

POST-TRANSCRIPTIONAL MODIFICATION OF SINE RNA

As mentioned above, most mammalian SINEs have an A-rich tail and some of them have a terminator at the 3′ end. The presence of a terminator proved to correlate with the presence of one or several AATAAA signals in the tail.[48] Elements with these signals are called T^+ SINEs, while those without them are T^- SINEs (Figure 4). Previous data pointed to the polyadenylation of pol III transcripts of mouse B2,[49] a typical T^+ SINE; recently it has been strictly confirmed using cell transfection with B2 copies carrying different modifications in the signals.[50] Thus, B2 transcription termination is followed by the nontemplated synthesis of poly(A) up to 300 nt in length, and similar to mRNA, B2 RNA polyadenylation depends on the polyadenylation signal AAUAAA. This finding was surprising because AAUAAA-dependent polyadenylation was thought to be limited to pol II transcripts. Apparently, transcripts of all T^+ SINE families can be polyadenylated (Figure 4). Amazingly, pol III transcripts of Alu, a T^- SINE, are not polyadenylated even after the polyadenylation and termination signals are introduced into its tail.[51]

Polyadenylation makes B2 RNA more stable in the cell.[50] This can be attributed to the protection by poly(A)-binding protein (PABP), which can protect mRNAs from degradation in the cell. A-rich RNA tails in T^- SINEs can also bind PABP,[52,53] which can be critical for their reverse transcription in a substantial fraction of SINE families.[54]

In contrast to tRNA transcripts that are subject to processing, no nucleotides are removed from the 5′ end of SINE pol III transcripts, which leaves a

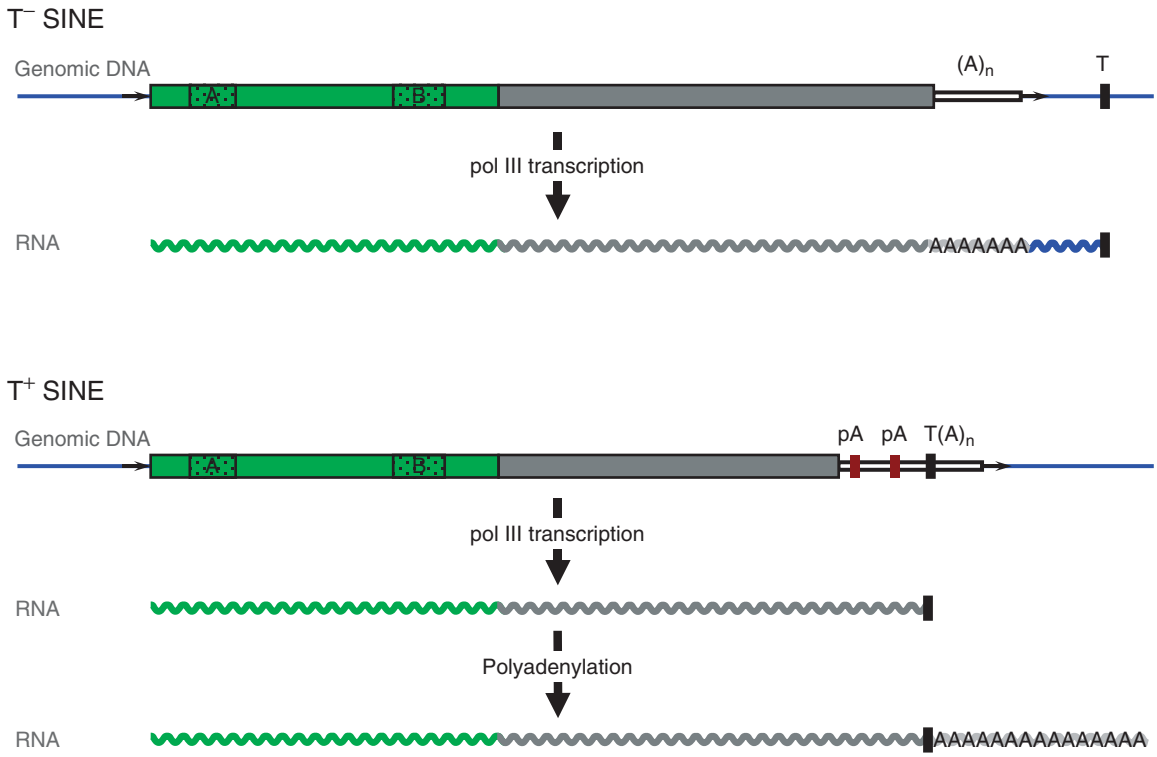


FIGURE 4 | Transcription of T^+ and T^- short interspersed elements. T, transcription termination signal; pA, polyadenylation signal.

5′-terminal triphosphate group (5′-pppN…). In B2 RNA, the γ-phosphate proved to be methylated (5′-CH_3pppN…) and this cap-like structure protected the group from phosphatase attack.[55] Apart from B2, such γ-monomethyl group was found in U6 and 7SK RNA as well as plant (but not animal) U3 RNAs (all transcribed by pol III), and a specific secondary structure is required for the methylation.[56] The modification is performed by methylphosphate capping enzyme[57] and it increases the lifetime of B2, U6, and 7SK RNAs in the cell.[58] On the other hand, pol III transcripts of another mouse SINE, B1, is not capped.[55] No such data are available for other SINE families.

Pre-tRNAs are subject to extensive modification by a number of enzymes. Although most SINEs descend from tRNAs, their RNA is not similarly modified. Early publications reported pseudouridines in salmon SmaI SINE and 5-methylcytidine in rat ID RNAs after *in vitro* transcription[59,60]; however, no modified nucleotides have been identified in ID after *in vivo* transcription.[61] Overall, the modification enzymes involved in tRNA maturation likely cannot recognize SINE transcripts due to a substantial sequence deviation of tRNA-derived SINEs from the corresponding tRNAs (25–35%) and radically different secondary structures of SINE RNAs and tRNAs.[62]

There is one more type of ribonucleotide modification affecting SINE transcripts. This is A-to-I RNA editing: the modification of adenosine to inosine in double-stranded RNA regions catalyzed by adenosine deaminases that act on RNA. More than 90% of human targets for A-to-I editing reside within Alu sequences in introns and 3′-UTRs (in pol II but not pol III transcripts). This can be due to the abundance of Alu copies (on average, one pre-mRNA contains 16 Alus), which allows double-stranded structures formed by neighboring Alus in opposite orientation. Amazingly, the rate of A-to-I editing is 30–40 times lower in mice than in humans, which can be attributed to a higher sequence divergence and shorter length of mouse SINEs (B1, B2, and ID) relative to human Alu, which lowers their propensity to form double-stranded RNA regions.[63]

The role of inosines within Alus and other noncoding regions of the pre-mRNA is still a mystery. It is not improbable that the editing of these noncoding regions has no biological significance. At the same time, recent data suggest that A-to-I editing of certain Alu copies can modulate the expression of genes that carry them (see below).

AMPLIFICATION OF SINEs

SINEs are nonautonomous mobile elements, because their reverse transcription relies on the enzyme of another type of mobile elements, LINEs (Box 1). LINE RT utilizes one of the two systems protecting it from processing foreign RNA templates: (1) specific sequence recognition of the RNA encoding the enzyme and (2) *cis*-preference, when the mRNA translated into RT is used by it as the template for reverse transcription. SINEs utilizing RTs of the first (stringent recognition) group have the sequence recognized by the RT at their 3′ end. It remains unclear how SINEs utilizing RTs of the second (relaxed recognition) group overcome the *cis*-preference, but all of them have poly(A) or A-rich tails as the recognized sequence. Such SINEs are particularly abundant in mammals, where they are mobilized by the RT of LINE1 (L1) of the relaxed recognition group.

After pol III transcription, SINE RNA binds the LINE RT (most likely in the cytoplasm) and the complex is transported back to the nucleus (Figure 5). The RT cleaves one of the genomic DNA strands using its endonuclease activity. The cleavage site is not very specific (e.g., 3′-AATTTT in the case of L1 RT). The resulting 3′ end of the genomic DNA serves as a primer for the reverse transcription of SINE RNA. Later, the RT cleaves the other genomic DNA strand (usually, 8–16 nt away from the first break), jumps to the resulting 3′ end of the genomic DNA, and uses it as a primer for the synthesis of the second strand of SINE DNA and an extra fragment of the genomic DNA (target site duplication, TSD). SINE RNA is displaced from the duplex with or without the RNase H activity of RT. Finally, the cellular DNA repair system fills the gaps in DNA.

Although the details of this retrotransposition mechanism called target-primed reverse transcription have been elucidated for LINEs, its applicability to SINE amplification has been specifically confirmed for eel UnaSINE1,[25] human Alu,[64] and mouse B1 and B2[54] (minor distinctions may exist; for instance, LINE *gag*-like protein is required for the reverse transcription of L1 RNA but not Alu RNA[64]).

SIGNIFICANCE OF SINEs

SINEs as Genomic Parasites

The concept of mobile genetic elements including SINEs as genomic parasites was proposed long ago.[65,66] Subsequent studies confirmed that SINEs do resemble parasites. They can amplify in the genome to huge numbers (for instance, Alus amount to more than 10% of the human genome). SINE copies integrated into coding or regulatory sequences can

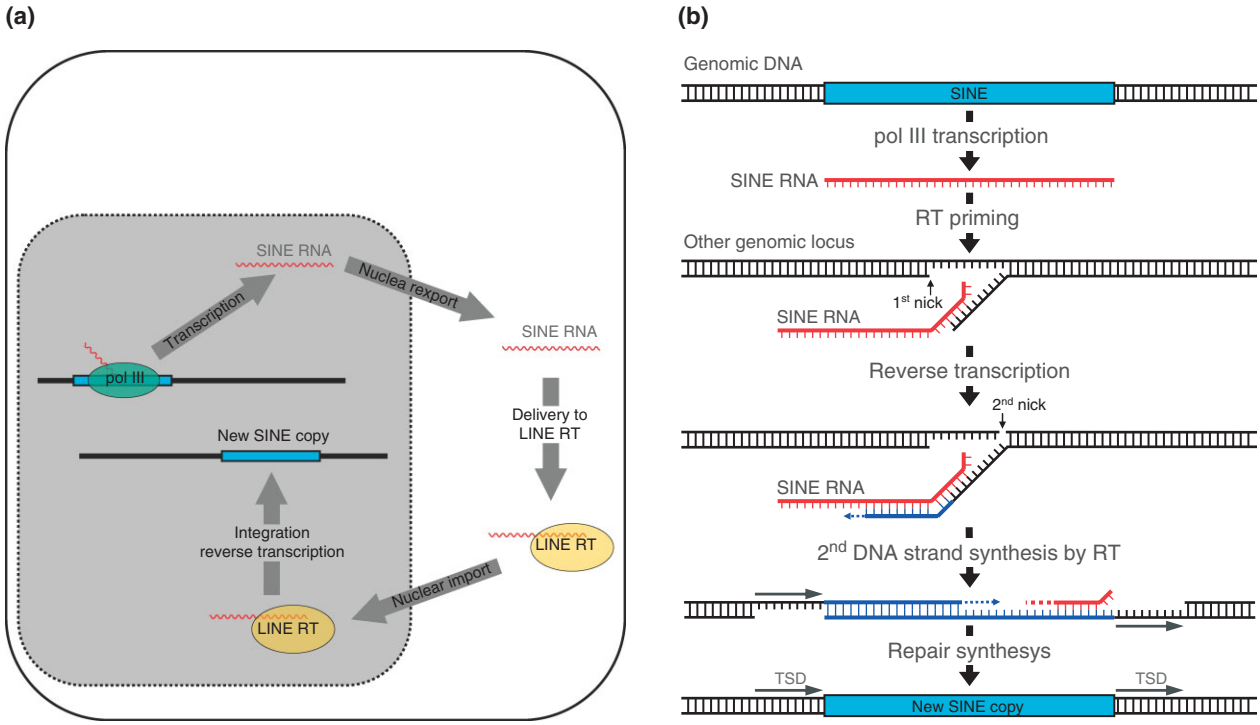


FIGURE 5 | Short interspersed element (SINE) life cycle. (a) Schematic amplification cycle. (b) Target-primed reverse transcription of SINEs. LINE, long interspersed element; pol III, RNA polymerase III; RT, reverse transcriptase; TSD, target site duplication.

disrupt gene expression. Alu integrations caused 33 human hereditary diseases.[67] On average, one new Alu insertion occurs approximately every 20 human births.[68] Fortunately, most integration events have no effect on gene activity (e.g., the integration site is far from functional sequences). However, the frequency of Alu integrations in primates was 100 times more 40–50 million years ago than now, and the rate of Alu-induced mutations was equally higher.[69]

A huge number of SINE copies in the genome underlie another type of genetic damage, unequal crossover. Such homologous recombination between nonallelic SINE sequences gives rise to chromosomal deletions, duplications, translocations, and inversions. To date, many human hereditary diseases caused by nonallelic Alu recombination have been described.[67] Similar events in somatic cells can cause malignant transformation.[70]

Recent data suggest that Alu RNA can have direct toxic effect on cells and cause retinal degeneration.[71] One can expect other examples of SINE RNA cytotoxicity in human and animal tissues.

Stated differently, SINEs can be highly aggressive endogenous factors of mutagenesis, and we can only guess how many biological species have ceased to exist as a result of SINE and LINE expansion. The dangerous nature of SINEs is confirmed by the powerful cell systems repressing their activity. For instance, pol III transcription of Alu and B1 is nearly completely repressed by the methylation of their DNA, which takes place during early embryogenesis. Dnmt3L, an activator of DNA methyl transferases, plays a critical role in the *de novo* methylation of repetitive elements including SINEs, and Dnmt3L-deficient males demonstrated uncontrolled transposon expression and spermatogenic defects[72] prompting that transposon-mediated mutagenesis can cause male sterility. Alu amplification is also suppressed by cytidine deaminase APOBEC3G by a yet unknown mechanism.[73,74] In addition, because SINE amplification depends on LINE RT, LINE repression also represses SINEs. The cell can repress LINE activity by RNA interference (RNAi), APOBEC3 system, and, finally, DNA methylation. (Interestingly, no data support that RNAi suppress pol III-transcribed genes, although DICER1, an enzyme involved in RNAi, can degrade toxic Alu pol III transcripts in the cell.[71])

The interaction between SINEs and LINEs on the one hand and the cell on the other hand resembles an arms race. In particular, this is indicated by the waves of SINE and LINE amplification during evolution,[75,76] reflecting more or less successful cell's control of SINEs and LINEs.

SINEs as Genomic Symbionts

Despite the parasitic nature of SINEs, the coexistence of SINEs and cells over millions of years mutually adapted them to make SINEs an integral or even beneficial part of many genomes. While harmful SINE integrations are eliminated by selection, neutral ones can become beneficial with time and diversify genome activity.

Cis-regulation of Gene Transcription

SINEs can function as promoters, enhancers, silencers, or insulators of protein-coding genes.[77] It is possible due to the presence of sequences in SINEs resembling functional signals of genes. For instance, Alu contains a region quite similar to the binding site of retinoic acid receptor, an important hormone-regulated transcription factor.[78] The ability of this region in many Alu copies to function as the receptor binding site has been experimentally confirmed.[79] Thus, the genomes of primates contain an increased number of binding sites of this receptor, which could play a substantial role in the evolution of this mammalian order.

AmnSINE1 is an ancient SINE family.[17] It amplified in the ancestors of vertebrates and is still active in some fish species (known as SINE3[16]), but became inactive in reptiles, birds, and mammals long ago. Nevertheless, hundreds of its degraded copies were conserved in their genomes, supposedly, because they became functional. Indeed, two such AmnSINE1 loci proved to be enhancers for development-associating mammalian genes *Fgf8* and *Satb2*,[80] and there are more AmnSINE1s that likely act as gene regulatory elements.[81] Apparently, AmnSINE1 and other ancient SINEs were important factors of vertebrate macroevolution.

The insulator function of SINEs can be exemplified by the mouse growth hormone (GH)/B2 system. The GH gene is activated during pituitary gland development, which is accompanied by the locus decondensation. The transcription of a B2 copy approximately 6 kb upstream of the GH gene by pol III and pol II in opposite orientations proved to be necessary and sufficient for the chromatin restructuring in the locus so that B2 was the boundary between heterochromatic and euchromatic regions.[82] Other copies of B2 and other SINEs can act as similar boundary elements controlling gene activity in development.

mRNA Splicing and Polyadenylation

The first publications on SINE sequencing (B1 and B2) pointed out to the regions resembling splice sites.[1,2] Alu consensus sequence contains 10 potential

5′ splice sites and 13 potential 3′ splice sites.[83] Usually these sites are not involved in pre-mRNA splicing, because potential splice sites in SINEs are not optimal and the sites of constitutive exons are much more potent. Nevertheless, alternative splicing of certain genes mediated by Alu, B2, and CHRS sequences have been reported.[84,85] Large-scale sequencing of transcriptomes from different human and mouse cell types provided for a substantial progress in these studies. For instance, at least 5% of all human alternatively spliced mRNA proved to depend on Alu splice sites. Note that 85% of Alu-derived exons involve the minus Alu strand, which is not surprising considering that only 4 of the 23 potential splice sites are in the plus strand.[83] SINE exonization requires changes in their potential splice sites to make them functional. Many human hereditary diseases resulted from Alu- and Ther-1-derived cryptic exons.[86]

Apart from DNA mutations, such changes can be due to A-to-I editing of pre-mRNA. One of introns of the nuclear prelamin A recognition factor gene in some higher primates contains two head-to-head Alu copies, which form a double-stranded region in the pre-mRNA. This hairpin is a target for A-to-I editing, and one of conversions generates a new 3′ splice site and, hence, a new exon.[36,87] Since the rate of A-to-I editing varies between tissues, the frequency of the alternative exon inclusion varies accordingly, thus providing a mechanism of tissue-specific expression of protein isoforms.

Polyadenylation is an important stage of mRNA maturation tightly linked to the formation of the mRNA 3′ end. This process usually includes endonucleolytic cleavage of pre-mRNA at the polyadenylation site, after which nontemplated poly(A) synthesis starts at the 3′ end of the cleaved RNA. Two signals are required for these reactions: an AATAAA (or ATTAAA) hexanucleotide 1–40 nt upstream of the polyadenylation site and a T/G-rich region 1–40 nt downstream of it. Integrated SINEs can introduce the polyadenylation signal into genes. This specifically applies to T^+ SINEs but also occurs with T^- ones having A-rich tails (which can have A to T substitutions). Genes utilizing polyadenylation signals stemming from SINEs were reported long ago.[88] Note that not all polyadenylation signals in the 3′-UTR are involved in the mRNA 3′ end formation, which can be attributed to missing T/G-rich region or other similar signals. Computer analysis of human genomic and RNA sequences has identified hundreds of polyadenylation signals stemming from Alu and Ther-1 SINEs.[89]

Origin of Nonprotein-Coding Genes

Certain nonprotein-coding genes originated from SINEs. All such genes are transcribed by pol III into small RNAs. These genes emerged relatively recently (dozens of MYA) and their taxonomic distribution is accordingly narrow (thus, they are called *steno*RNAs from Greek στενός, narrow[90]).

BC1 RNA is synthesized in the nervous tissue of (presumably) all rodents.[91] Usually, the genome has a single BC1 gene, which is essentially a copy of ID SINE, and the reverse transcription of BC1 RNA leads to ID amplification in the genome.[92] BC1 RNA was shown to repress translation of certain mRNAs in dendrites.[93] BC200 has a similar function in humans and, likely, other higher primates,[93] although it originated from the ancient monomeric Alu FLAM-C.

4.5SI RNA (~100 nt) can be found in all tissues of three rodent families, Muridae, Cricetidae, and Spalacidae.[90] In mouse and rat, it is transcribed from three genes within the same locus.[94] The 5′-terminal sequence of 4.5SI RNA is similar to that of B2 SINE, which points to their common origin. Rodents of the same families plus Dipodidae express 4.5SH RNA of a similar length.[95] The sequence similarity between 4.5SH RNA and pB1 SINE indicates that the 4.5SH gene originated from a pB1 copy. The genomes of these rodents have hundreds of 4–5 kb tandem repeats each containing a 4.5SH gene.[95,96] In contrast to 4.5SI RNA, 4.5SH RNA is short-lived ($T_{1/2} = 30$ min). The sequences of both RNAs are highly conserved, which indicates that they have a function that remains to be determined.

Recently, a small (117 nt) RNA snaR has been identified in some higher primates.[97] SnaR genes originated from 7SL RNA via low-copy number SINEs (FLAM-C, ARS, and CAS) through deletions of different lengths, small tandem duplications, nucleotide substitutions, and the emergence of a transcription terminator. Human genome contains at least 20 snaR genes within large tandem repeats. The half-life of snaR is only 15 min. These properties of primate snaR bring it together with rodent 4.5SH RNA. At the same time, snaR is actively expressed only in the testis, while other tissues demonstrate moderate or low expression levels. SnaR is associated with ribosomes or polysomes, which can point to its involvement in translation control.

MicroRNAs are very short (17–25 nt) noncoding RNAs that guide the repression machinery to specific mRNAs. Most miRNA are generated by processing pol II miRNA transcripts. However, a fraction of primate miRNAs is transcribed by pol III together with the upstream Alus and, possibly, Ther-1.[98,99] Thus, SINE integration into a proper genomic locus

can give birth to miRNA genes. SINEs can also be involved in other aspects of miRNA biology.[100]

Trans-regulation

All functions of SINEs considered above affect DNA or RNA molecules where they reside, i.e., in *cis*. However, SINE RNA can act as *trans* factors as well. Cells exposed to stress (heat shock, viral infection, or toxic agents) demonstrate elevated pol III transcription of SINEs. High levels of Alu RNA in such cells were proposed to favor their survival through translation control[101–103]; however, the underlying mechanism remains to be elucidated.

On the other hand, pol III transcripts of B2 and Alu were shown to bind and inhibit pol II.[104–107] This is used to explain the transcriptional inhibition of most genes during heat shock, while the transcription of heat shock genes is resistant to B2 and Alu RNA. Eventually, the synthesis of heat shock chaperones contributes to cell recovery from stress. This attractive model also has some gaps. Is the inhibitory effect of B2 and Alu RNA strong enough to provide the transcriptional repression observed during heat shock? What underlies the resistance of transcription of heat shock protein genes to high levels of B2 or Alu RNA? Why their inhibitory effect on pol II is not observed in other cells with high levels of B2/Alu RNA (e.g., tumor and testicular cells or early embryonic cells)? As an alternative, one can propose that the capacity of SINE RNA to inhibit pol II can favor SINE expansion in the genome by a yet unknown mechanism.

Recently, Alu-containing RNAs were shown to mediate mRNA degradation. Certain mRNAs with Alu in the 3′-UTR can interact with noncoding Alu-containing RNAs, and the Alu-Alu duplex is recognized by STAU1 protein, the trigger of STAU1-mediated mRNA decay.[108]

CONCLUSION

Thirty years of studies on SINEs largely elucidated the nature of these repetitive sequences of eukaryotic genomes. SINEs are short nonautonomous mobile genetic elements, whose amplification relies on the RT encoded by another type of mobile genetic elements, LINEs. All SINE families originated from one of the three types of short pol III transcripts: tRNAs, 5S rRNA, or 7SL RNA. Apart from the cellular RNA-derived head, most SINEs have a body (whose origin is not always clear) and a tail composed of simple repeats. SINEs are transcribed by pol III, and the resulting small heterogeneous RNAs serve as the template for the LINE RT. The integration of new SINE copies into genes can affect their expression. At the same time, long-term evolution has converted certain SINE copies into promoters, enhancers, silencers, or insulators, and their sequences occasionally provide signals for alternative splicing and polyadenylation in protein-coding genes. The involvement of some SINE RNAs in stress recovery was proposed.

Despite the progress in our understanding of SINE biology, many aspects remain unclear. We consider the following promising trends in the field: (1) elucidation of the mechanisms underlying SINE transcription control as well as SINE RNA processing and transport; (2) investigation of SINE structures critical for their amplification; (3) identification of cell systems suppressing retrotransposition of SINEs and LINEs; (4) evaluation of the role of SINEs in hereditary diseases and carcinogenesis; (5) understanding the absence of SINEs in some eukaryotes and their abundance in other ones; and (6) wide application of SINEs in phylogenetic studies.

ACKNOWLEDGMENTS

The authors are supported by the Molecular and Cellular Biology Program of the Russian Academy of Sciences and the Russian Foundation for Basic Research (project nos 10-04-01259-a and 11-04-00439-a).

REFERENCES

1. Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res* 1980, 8:1201–1215.
2. Krayev AS, Markusheva TV, Kramerov DA, Ryskov AP, Skryabin KG, Bayev AA, Georgiev GP. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucleic Acids Res* 1982, 10:7461–7475.

3. Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J Mol Biol* 1981, 151:17–33.
4. Babushok DV, Kazazian HH Jr. Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat* 2007, 28:527–539.
5. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 1993, 72:595–605.
6. Bailey JA, Liu G, Eichler EE. An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 2003, 73:823–834.
7. Eickbush T, Malik H. Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AL, eds. *Mobile DNA II*. Washington, DC: ASM Press; 2002, 1111–1144.
8. Lovsin N, Gubensek F, Kordis D. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol Biol Evol* 2001, 18:2213–2224.
9. Eickbush TH, Jamburuthugoda VK. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* 2008, 134:221–234.
10. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001, 35:501–538.
11. Deragon JM, Zhang X. Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. *Syst Biol* 2006, 55:949–956.
12. Kramerov DA, Vassetzky NS. Short retroposons in eukaryotic genomes. *Int Rev Cytol* 2005, 247:165–221.
13. Nishihara H, Terai Y, Okada N. Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol Biol Evol* 2002, 19:1964–1972.
14. Vassetzky NS, Ten OA, Kramerov DA. B1 and related SINEs in mammalian genomes. *Gene* 2003, 319:149–160.
15. Zietkiewicz E, Richer C, Sinnett D, Labuda D. Monophyletic origin of Alu elements in primates. *J Mol Evol* 1998, 47:172–182.
16. Kapitonov VV, Jurka J. A novel class of SINE elements derived from 5S rRNA. *Mol Biol Evol* 2003, 20:694–702.
17. Nishihara H, Smit AF, Okada N. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res* 2006, 16:864–874.
18. Gogolevsky KP, Vassetzky NS, Kramerov DA. 5S rRNA-derived and tRNA-derived SINEs in fruit bats. *Genomics* 2009, 93:494–500.
19. Gogolevsky KP, Vassetzky NS, Kramerov DA. Bov-B-mobilized SINEs in vertebrate genomes. *Gene* 2008, 407:75–85.
20. Gilbert N, Labuda D. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci U S A* 1999, 96:2869–2874.
21. Ogiwara I, Miya M, Ohshima K, Okada N. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res* 2002, 12:316–324.
22. Akasaki T, Nikaido M, Nishihara H, Tsuchiya K, Segawa S, Okada N. Characterization of a novel SINE superfamily from invertebrates: "Ceph-SINEs" from the genomes of squids and cuttlefish. *Gene* 2010, 454:8–19.
23. Ohshima K, Hamada M, Terai Y, Okada N. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol Cell Biol* 1996, 16:3756–3764.
24. Okada N, Hamada M. The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. *J Mol Evol* 1997, 44:52–56.
25. Kajikawa M, Okada N. LINEs mobilize SINEs in the eel through a shared 3′ sequence. *Cell* 2002, 111:433–444.
26. Kramerov D, Vassetzky N, Serdobova I. The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. *Mol Biol Evol* 1999, 16:715–717.
27. Veniaminova NA, Vassetzky NS, Kramerov DA. B1 SINEs in different rodent families. *Genomics* 2007, 89:678–686.
28. Nishihara H, Satta Y, Nikaido M, Thewissen JG, Stanhope MJ, Okada N. A retroposon analysis of Afrotherian phylogeny. *Mol Biol Evol* 2005, 22:1823–1833.
29. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. *Bioessays* 2000, 22:148–160.
30. Nishihara H, Okada N. Retroposons: genetic footprints on the evolutionary paths of life. *Methods Mol Biol* 2008, 422:201–225.
31. Perna NT, Batzer MA, Deininger PL, Stoneking M. Alu insertion polymorphism: a new type of marker for human population studies. *Hum Biol* 1992, 64:641–648.
32. Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J. Rodent evolution: back to the root. *Mol Biol Evol* 2010, 27:1315–1326.
33. Murata S, Takasaki N, Saitoh M, Okada N. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc Natl Acad Sci U S A* 1993, 90:6995–6999.
34. Nikaido M, Rooney AP, Okada N. Phylogenetic relationships among cetartiodactyls based on insertions of

short and long interpersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci U S A* 1999, 96:10261–10266.

35. Kramerov DA, Grigoryan AA, Ryskov AP, Georgiev GP. Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. *Nucleic Acids Res* 1979, 6:697–713.
36. Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G. RNA-editing-mediated exon evolution. *Genome Biol* 2007, 8:R29.
37. Chesnokov I, Schmid CW. Flanking sequences of an Alu source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors. *J Mol Evol* 1996, 42:30–36.
38. Arnaud P, Yukawa Y, Lavie L, Pelissier T, Sugiura M, Deragon JM. Analysis of the SINE S1 Pol III promoter from Brassica; impact of methylation and influence of external sequences. *Plant J* 2001, 26:295–305.
39. Schramm L, Hernandez N. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* 2002, 16:2593–2620.
40. Endoh H, Okada N. Total DNA transcription *in vitro*: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. *Proc Natl Acad Sci U S A* 1986, 83:251–255.
41. Matsumoto K, Takii T, Okada N. Characterization of a new termination signal for RNA polymerase III responsible for generation of a discrete-sized RNA transcribed from salmon total genomic DNA in a HeLa cell extract. *J Biol Chem* 1989, 264:1124–1131.
42. Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E. Characterization of a plant SINE, p-SINE1, in rice genomes. *Jpn J Genet* 1992, 67:155–166.
43. Liu WM, Chu WM, Choudary PV, Schmid CW. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* 1995, 23:1758–1765.
44. Bachvarova R. Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. *Dev Biol* 1988, 130:513–523.
45. Kim J, Kass DH, Deininger PL. Transcription and processing of the rodent ID repeat family in germline and somatic cells. *Nucleic Acids Res* 1995, 23:2245–2251.
46. Grigoryan MS, Kramerov DA, Tulchinsky EM, Revasova ES, Lukanidin EM. Activation of putative transposition intermediate formation in tumor cells. *EMBO J* 1985, 4:2209–2215.
47. White RJ, Stott D, Rigby PW. Regulation of RNA polymerase III transcription in response to Simian virus 40 transformation. *EMBO J* 1990, 9:3713–3721.
48. Borodulina OR, Kramerov DA. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mamm Genome* 2001, 12:779–786.
49. Kramerov DA, Tillib SV, Shumyatsky GP, Georgiev GP. The most abundant nascent poly(A) + RNAs are transcribed by RNA polymerase III in murine tumor cells. *Nucleic Acids Res* 1990, 18:4499–4506.
50. Borodulina OR, Kramerov DA. Transcripts synthesized by RNApolymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA* 2008, 14:1865–1873.
51. Roy-Engel AM, El-Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H, Oyeniran OO, Deininger PL. Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet Genome Res* 2005, 110:365–371.
52. Muddashetty R, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskothen J, Duning K, Barnekow A, Huttenhofer A, Tiedge H, et al. Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J Mol Biol* 2002, 321:433–445.
53. West N, Roy-Engel AM, Imataka H, Sonenberg N, Deininger P. Shared protein components of SINE RNPs. *J Mol Biol* 2002, 321:423–432.
54. Dewannieux M, Heidmann T. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J Mol Biol* 2005, 349:241–247.
55. Shumyatsky GP, Tillib SV, Kramerov DA. B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5′ end. *Nucleic Acids Res* 1990, 18:6347–6351.
56. Shumyatsky G, Shimba S, Reddy R. Capping signals correspond to the 5′ end in four eukaryotic small RNAs containing gamma-monomethylphosphate cap structure. *Gene Expr* 1994, 4:29–41.
57. Shimba S, Reddy R. Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. *J Biol Chem* 1994, 269:12419–12423.
58. Shumyatsky G, Wright D, Reddy R. Methylphosphate cap structure increases the stability of 7SK, B2 and U6 small RNAs in Xenopus oocytes. *Nucleic Acids Res* 1993, 21:4756–4761.
59. Matsumoto K, Murakami K, Okada N. Pseudouridylic modification of a 6S RNA transcribed *in vitro* from highly repetitive and transcribable (Hirt) sequences of salmon total DNA. *Biochem Biophys Res Commun* 1984, 124:514–522.
60. Sakamoto K, Okada N. 5-Methylcytidylic modification of *in vitro* transcript from the rat identifier sequence; evidence that the transcript forms a tRNA-like structure. *Nucleic Acids Res* 1985, 13:7195–7206.
61. Rozhdestvensky TS, Crain PF, Brosius J. Isolation and posttranscriptional modification analysis of native BC1 RNA from mouse brain. *RNA Biol* 2007, 4:11–15.

62. Sun FJ, Fleurdepine S, Bousquet-Antonelli C, Caetano-Anolles G, Deragon JM. Common evolutionary trends for SINE RNA structures. *Trends Genet* 2007, 23:26–33.
63. Neeman Y, Levanon EY, Jantsch MF, Eisenberg E. RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA* 2006, 12: 1802–1809.
64. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 2003, 35:41–48.
65. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 1980, 284:601–603.
66. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature* 1980, 284:604–607.
67. Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res* 2008, 18:343–358.
68. Cordaux R, Hedges DJ, Herke SW, Batzer MA. Estimating the retrotransposition rate of human Alu elements. *Gene* 2006, 373:134–137.
69. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet* 2002, 3:370–379.
70. Konkel MK, Batzer MA. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* 2010, 20:211–221.
71. Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, et al. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* 2011, 471:325–330.
72. Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H. Role of the Dnmt3 family in *de novo* methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* 2007, 16:2272–2280.
73. Hulme AE, Bogerd HP, Cullen BR, Moran JV. Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* 2007, 390:199–205.
74. Khatua AK, Taylor HE, Hildreth JE, Popik W. Inhibition of LINE-1 and Alu retrotransposition by exosomes encapsidating APOBEC3G and APOBEC3F. *Virology* 2010, 400:68–75.
75. Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol* 2003, 4:R74.
76. Quentin Y. Successive waves of fixation of B1 variants in rodent lineage history. *J Mol Evol* 1989, 28:299–305.
77. Makalowski W. Genomic scrap yard: how genomes utilize all that junk. *Gene* 2000, 259:61–67.
78. Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N. Association of some potential hormone response elements in human genes with the Alu family repeats. *Gene* 1999, 239:341–349.
79. Laperriere D, Wang TT, White JH, Mader S. Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genomics* 2007, 8:23.
80. Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, Kokubo N, Kimura-Yoshida C, Matsuo I, Sumiyama K, Saitou N, et al. Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci U S A* 2008, 105:4220–4225.
81. Hirakawa M, Nishihara H, Kanehisa M, Okada N. Characterization and evolutionary landscape of AmnSINE1 in Amniota genomes. *Gene* 2009, 441:100–110.
82. Lunyak VV, Prefontaine GG, Nunez E, Cramer T, Ju BG, Ohgi KA, Hutt K, Roy R, Garcia-Diaz A, Zhu X, et al. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 2007, 317:248–251.
83. Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. *Genome Res* 2002, 12: 1060–1067.
84. Pattanakitsakul S, Zheng JH, Natsuume-Sakai S, Takahashi M, Nonaka M. Aberrant splicing caused by the insertion of the B2 sequence into an intron of the complement C4 gene is the basis for low C4 production in H-2k mice. *J Biol Chem* 1992, 267:7814–7820.
85. Shimamura M, Nikaido M, Ohshima K, Okada N. A SINE that acquired a role in signal transduction during evolution. *Mol Biol Evol* 1998, 15:923–925.
86. Vorechovsky I. Transposable elements in disease-associated cryptic exons. *Hum Genet* 2010, 127:135–154.
87. Moller-Krull M, Zemann A, Roos C, Brosius J, Schmitz J. Beyond DNA: RNA editing and steps toward Alu exonization in primates. *J Mol Biol* 2008, 382:601–609.
88. Ryskov AP, Ivanov PL, Kramerov DA, Georgiev GP. Mouse ubiquitous B2 repeat in polysomal and cytoplasmic poly(A)+RNAs: uniderectional orientation and 3′-end localization. *Nucleic Acids Res* 1983, 11:6541–6558.
89. Lee JY, Ji Z, Tian B. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. *Nucleic Acids Res* 2008, 36:5581–5590.
90. Gogolevskaya IK, Kramerov DA. Evolutionary history of 4.5SI RNA and indication that it is functional. *J Mol Evol* 2002, 54:354–364.
91. Martignetti JA, Brosius J. Neural BC1 RNA as an evolutionary marker: guinea pig remains a rodent. *Proc Natl Acad Sci U S A* 1993, 90:9698–9702.

92. Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P. Rodent BC1 RNA gene as a master gene for ID element amplification. *Proc Natl Acad Sci U S A* 1994, 91:3607–3611.

93. Lin D, Pestova TV, Hellen CU, Tiedge H. Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. *Mol Cell Biol* 2008, 28:3008–3019.

94. Gogolevskaya IK, Kramerov DA. 4.5SI RNA genes and the role of their 5′-flanking sequences in the gene transcription. *Gene* 2010, 451:32–37.

95. Gogolevskaya IK, Koval AP, Kramerov DA. Evolutionary history of 4.5SH RNA. *Mol Biol Evol* 2005, 22:1546–1554.

96. Schoeniger LO, Jelinek WR. 4.5S RNA is encoded by hundreds of tandemly linked genes, has a short half-life, and is hydrogen bonded in vivo to poly(A)-terminated RNAs in the cytoplasm of cultured mouse cells. *Mol Cell Biol* 1986, 6:1508–1519.

97. Parrott AM, Tsai M, Batchu P, Ryan K, Ozer HL, Tian B, Mathews MB. The evolution and expression of the snaR family of small non-coding RNAs. *Nucleic Acids Res* 2011, 39:1485–1500.

98. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* 2006, 13:1097–1101.

99. Gu TJ, Yi X, Zhao XW, Zhao Y, Yin JQ. Alu-directed transcriptional regulation of some novel miRNAs. *BMC Genomics* 2009, 10:563.

100. Pouch-Pelissier MN, Pelissier T, Elmayan T, Vaucheret H, Boko D, Jantsch MF, Deragon JM. SINE RNA induces severe developmental defects in Arabidopsis thaliana and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet* 2008, 4:e1000096.

101. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW. Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol Cell Biol* 1998, 18:58–68.

102. Hasler J, Strub K. Alu RNP and Alu RNA regulate translation initiation in vitro. *Nucleic Acids Res* 2006, 34:2374–2385.

103. Rubin CM, Kimura RH, Schmid CW. Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res* 2002, 30:3253–3261.

104. Allen TA, Von Kaenel S, Goodrich JA, Kugel JF. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 2004, 11:816–821.

105. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008, 29:499–509.

106. Ponicsan SL, Kugel JF, Goodrich JA. Genomic gems: SINE RNAs regulate mRNA production. *Curr Opin Genet Dev* 2010, 20:149–155.

107. Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* 2004, 11:822–829.

108. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 2011, 470:284–288.

FURTHER READING

Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know 'bout retroelements in cancer. *Semin Cancer Biol* 2010, 20:200–210.

Kramerov DA, Vassetzky NS. Short retroposons in eukaryotic genomes. *Int Rev Cytol* 2005, 247:165–221.

Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 2011, doi: 10.1038/hdy.2011.43.

Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. *Bioessays* 2000, 22:148–160.

