

# Transposable Elements: No More 'Junk DNA'

Yun-Ji Kim, Jungnam Lee, Kyudong Han\*

Department of Nanobiomedical Science, WCU Research Center, Dankook University, Cheonan 330-714, Korea

Since the advent of whole-genome sequencing, transposable elements (TEs), just thought to be 'junk' DNA, have been noticed because of their numerous copies in various eukaryotic genomes. Many studies about TEs have been conducted to discover their functions in their host genomes. Based on the results of those studies, it has been generally accepted that they have a function to cause genomic and genetic variations. However, their infinite functions are not fully elucidated. Through various mechanisms, including *de novo* TE insertions, TE insertion-mediated deletions, and recombination events, they manipulate their host genomes. In this review, we focus on *Alu*, L1, human endogenous retrovirus, and short interspersed element/variable number of tandem repeats/*Alu* (SVA) elements and discuss how they have affected primate genomes, especially the human and chimpanzee genomes, since their divergence.

**Keywords:** *Alu* elements, DNA transposable elements, endogenous retroviruses, long interspersed nucleotide elements, SVA

## Introduction

Transposable elements (TEs), mobile segments of genetic material, were first discovered by McClintock [1]. Since then, they have been identified in a variety of eukaryotes [2]. Recent genome sequencing projects have consistently shown that TEs make up ~50% of primate genomes, while coding DNA occupies only ~2% of the genomes [3-5]. TEs are generally divided into two categories, DNA transposons and retrotransposons (Fig. 1), based on their manner of mobilization. DNA transposons move using a cut-and-paste mechanism [6]. In contrast, retrotransposons move in a copy-and-paste fashion by duplicating the element into a new genomic location via an RNA intermediate [7]. Thus, retrotransposons increase their copy number more rapidly than DNA transposons.

Retrotransposons include short interspersed element (SINE), long interspersed element (LINE), and human endogenous retrovirus (HERV). *Alu* and short interspersed element/variable number of tandem repeats/*Alu* (SVA) elements are primate-specific retrotransposons, and their full-length is 300 bp and 2 kb, respectively. The *Alu* element is the most successful SINE in terms of its copy number; ~1.2 million *Alu* copies exist in the human genome. LINE is ~6 kb in length, and thus, it is much longer than the SINEs. This

element has two open reading frames encoding enzymatic machineries essential for the propagation of the three elements; the *Alu* element depends on reverse transcriptase of LINE for making their dispersed copies in the host genome [8]. In contrast to SINE and LINE, which do not have long terminal repeats (LTRs), the full-length HERV (~10 kb) has two LTRs, and three genes-*gag*, *pol*, and *env*-are located between them [8, 9].

Studies on active TEs have suggested that the elements could alter gene expression by providing *cis*-regulatory elements, such as promoters, enhancers, and transcription factor binding sites [10]. Through these mechanisms, altered transcriptional activity could lead to dysfunctional and abnormal proteins. Through *de novo* TE insertion within a gene, TEs could alter a gene product, which could be either harmful or beneficial to its host genome [11, 12]. In cases where an inserted TE causes a harmful effect on its host genome, the TE is likely to go to inactivation and fossilization by evolutionary accumulation of mutations and silencing effects [13].

Since the divergence of the human and chimpanzee, ~6 million years ago, many TEs have propagated in each genome. Among them are the *Alu*, L1, SVA, and HERV-K (HML-2) elements. During the past 6 million years, 5,530 *Alu*, 1,835 L1, 864 SVA, and 113 HERV-K (HML-2) elements

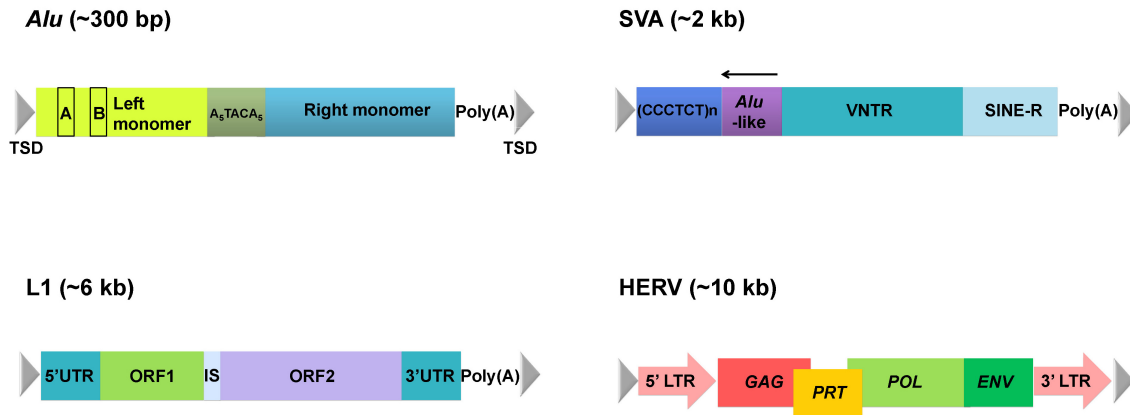
Received November 1, 2012; Revised November 16, 2012; Accepted November 17, 2012

\*Corresponding author: Tel: +82-41-550-3567, Fax: +82-41-550-3450, E-mail: jim97@dankook.ac.kr

Copyright © 2012 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).

## (A) Retrotransposon



## (B) DNA transposon



**Fig. 1.** Structures of transposable elements. These elements could be categorized into retrotransposons (*Alu*, long interspersed element [LINE], and human endogenous retrovirus [HERV]) (A) and DNA transposons (e.g., MARINER) (B) based on their manner of mobilization. In addition, autonomous elements (e.g., HERV and LINE) have coding genes responsible for their own mobilization but also other nonautonomous elements (e.g., *Alu* and short interspersed element/variable number of tandem repeats/*Alu* [SVA]). *Alu* consists of two monomers separated by an A-rich connector, one of which, the left monomer, includes internal RNA polymerase III promoter (A and B boxes). Full-length of LINE is ~6 kb and has open reading frames (ORFs) encoding RNA-binding protein, endonuclease, and reverse-transcriptase, which are flanked by untranslated regions (UTRs). ORF1 and ORF2 are separated by an ~60-bp-long intergenic spacer (IS). SVA contains a (CCCTCT)<sub>n</sub> hexamer, *Alu*-like sequences, variable number of tandem repeat (VNTR), and short interspersed element-R (SINE-R). An arrow on *Alu*-like sequences indicates the direction of *Alu*. HERV has *gag*, *prt*, *pol*, and *env* genes flanked by a long terminal repeat (LTR), which encodes capsid protein, protease, polymerase, and envelop protein, respectively, used in viral infection. As an example of DNA transposon, *mariner* has a gene encoding transposase with a DNA-binding domain and catalytic domain flanked by an inverted repeat (IR). All elements are flanked by target site duplication (TSD) through integration. DDE, the conserved DDE sequence of the mariner transposase; NLS, nuclear localization signal.

are estimated to have been newly inserted in the human genome (Table 1) [8, 14, 15]. These elements could act as an agent causing human-specific genomic rearrangements via *de novo* TE insertions, TE insertion-mediated deletions, and homologous recombination events [12, 16-19]. Furthermore, some of the recent integrated TEs are capable of producing new copies in the human genome. These *de novo* TE insertions have the potential to cause a genomic difference among human populations and even human individuals, which could be related to human phenotypes and diseases [20].

In this review, we describe species-specific TEs and discuss

how they affect their host genomes, focusing on illustrating the mechanisms that they utilize with examples. Taken together, we suggest that TEs, often called 'junk' DNA, in fact have many functions and play a significant and dynamic role in primate genomic evolution.

### TEs Recently Inserted into the Human Genome

Comparative genomics allows us to investigate species-specific TEs. Through a combinational method of computational data mining and experimental verification, species-

**Table 1.** Genomic rearrangement associated with active transposable elements (TEs) in human genome

Mechanism	Type of TE	No. of events (deletion size, bp)	References
<i>de novo</i> insertion	<i>Alu</i>	5,530	[22]
	L1	1,835	[15]
	SVA	864	[22]
	HERV-K	113	[30]
Recombination-mediated deletion	<i>Alu-Alu</i>	492 (396,420)	[16]
	L1-L1	73 (447,567)	[24]
	SVA-SVA	1 (589)	[33]
Insertion-mediated deletion	<i>Alu</i>	23 (11,206)	[17, 31]
	L1	31 (22,873)	[18, 32]
	SVA	13 (30,785)	[33]

specific TEs have been well studied in various primate genomes, including human, chimpanzee, and rhesus macaque. During the past 6 million years, more than 10,000 human-specific TEs were newly inserted in the human [21]. The majority of the identified elements belong to *Alu*, L1, and SVA retrotransposons. These elements have contributed to the genomic difference between human and non-human primates through insertion and post-insertion recombination between the elements [8, 22].

*Alu* elements are most ubiquitous in the human genome. A master gene model has been generally accepted to explain the amplification of *Alu* elements. In the model, only limited numbers of hyperactive “source” or “master” genes are able to produce a high number of *Alu* copies over evolutionary time. One of the previous studies about the *AluYb* subfamily, one of the most active *Alu* subfamilies in the human genome, found that the copy number of *AluYb* elements is very different between human and non-human primates; a high copy number of *AluYb* elements exists in the human, but a very small number of the elements exists in non-human primates [23]. The oldest *AluYb* element resides at its orthologous position in all hominid primate genomes, demonstrating that the *AluYb* subfamily emerged 18 to 25 million years ago. Then, after approximately 20 million years of retrotranspositional quiescence, a major expansion of the subfamily occurred only in the human genome within the past several million years. To explain their successful proliferation in the human genome, a new model, “stealth driver,” was introduced. In this model, a high copy of the *Alu* elements could be driven at least in part by “stealth driver” elements, which maintain low retrotransposition activity over extended periods of time. Although this element has low retrotransposition activity, they have the potential to produce short-lived hyperactive copies responsible for the remarkable expansion of *AluYb* elements within the human genome [24].

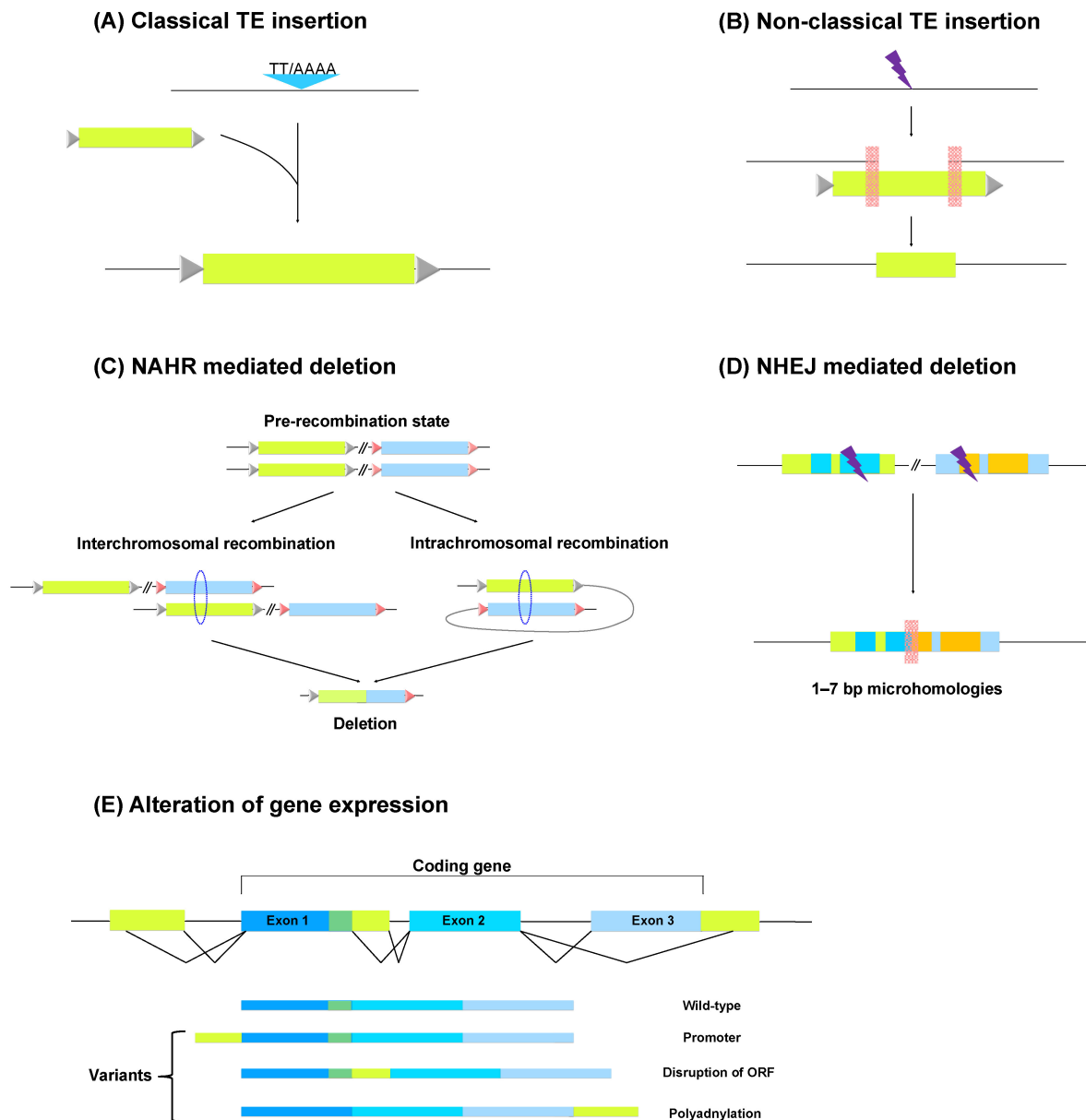
L1 is another successful element, occupying ~17% of the

human genome. The copy number of *Alu* elements is much higher than that of L1 (*Alu*, ~1.2 million copies; L1, ~520,000 copies). Nonetheless, *Alu* elements are responsible for ~11% of the human genome, because the *Alu* element is much shorter than L1 (*Alu* average length, 300 bp; full-length of LINE, 6 kb) [25]. Through characterization of sequence diversity of chimpanzee-specific L1 subfamilies as compared to their human-specific counterparts, it was concluded that L1s experienced different evolutionary fates between humans and chimpanzees within the past ~6 million years. Although the species-specific L1 copy numbers were on the same order in the two species (1,800 human-specific L1s vs. 1,200 chimpanzee-specific L1s), the number of retrotransposition-competent elements was much higher in the human genome than in the chimpanzee genome. The species-specific L1s were grouped into several L1 subfamilies. All human L1 subfamilies belonged to a single lineage, but two distinct L1 lineages were identified in the chimpanzee genome [15].

SVA is shared only in human and apes. Fewer than 1,000 copy numbers were detected in orangutan (~15 million years ago), and no SVA detected in Old World monkeys indicating SVA is a hominid-specific element. Like *Alu* elements, SVA elements retrotranspose to another locus *in trans* by using reverse-transcriptase encoded by L1s [8]. Due to the limited mobilization and short evolutionary time, the copy number of SVA is very small compared to *Alu* and L1 elements [15, 26, 27]. There are six SVA subfamilies (SVA\_A to SVA\_F) in the human genome. The older, SVA\_A to SVA\_D, evolved in a single lineage, whereas human-specific SVA\_E and SVA\_F were derived independently from their ancestral sequences. The copy number of SVA was estimated throughout the primate genomes of human, chimpanzee, and gorilla, and there was no significant difference among them. The two elements, *Alu* and L1, showed a huge expansion at a specific evolutionary time along the primate lineage, but SVA still did not show any burst in its copy number [26, 27].

## Genomic Rearrangements by TEs

The comparison of human and chimpanzee genomic sequences showed that the two genomes have a much higher sequence identity than we expected [3, 4]. In spite of the sequence similarity, TEs have remarkably generated genomic differences between the two species since their divergence [22]. Many studies have suggested that a number of TEs are still active to retrotranspose and have the potential to cause genomic variations as a major driver [12, 16-19, 28]. In reality, TEs have rearranged human and non-human primate genomes through various mechanisms, such as *de*



**Fig. 2.** Schematic representation of genomic rearrangement and gene expression alteration by transposable elements (TEs) in host genome. (A) Classical TE insertion by recognizing 5'-TTAAA-3', (B) non-classical TE insertion, (C) nonallelic homologous recombination (NAHR)-mediated deletion, (D) nonhomologous end-joining (NHEJ)-mediated deletion, (E) mechanism of gene expression alteration by TEs integrated into the host gene. Depending on location of insertion in the host gene, TEs could generate alternative transcripts or disrupt the expression. ORF, open reading frame; Grey and pink arrow boxes, target site duplication; black line, flanking region; grey line, intervening region; dotted circles, homologous recombination regions; pink boxes, microhomology region.

*novo* TE insertion, TE insertion-mediated deletion, and homologous recombination between them (Fig. 2) [29]. These genomic changes caused by TEs have increased the genomic difference between human and non-human primates, and some of the human-specific genomic rearrangements caused human diseases [28, 34].

Advanced sequencing technology, including next-generation sequencing, and combined computational analyses have

accelerated the studies on the dynamics of TE mobilization [35]. In reality, human-specific TEs have been continuously investigated, and the majority of them are *Alu*, L1, and SVA elements [8, 14]. The relationship between human brain evolution and *Alu* elements was studied. Since the divergence of the human and chimpanzee lineages, the human brain has rapidly changed in terms of mass [36]. It is not an exaggeration to say that *Alu* elements are in part responsible

for the human brain mass. Interestingly, *de novo* *Alu* insertions have been identified in many human brain genes that are related to neuronal functions and neurological disorders [37]. The inserted *Alu* elements belong to *AluYa5*, *AluYb8*, and *AluYc1*, which are human-specific *Alu* subfamilies [37].

Approximately 1,800 human-specific L1s were identified in the human genome [15]. They belonged to two different subfamilies, pre-Ta and Ta; Ta is subdivided into Ta-0 and Ta-1 by diagnostic nucleotides [38]. Among hominid-specific SVA subfamilies, SVA\_E and SVA\_F are only detected in the human genome, but the other four subfamilies, SVA\_A, SVA\_B, SVA\_C, and SVA\_D, are shared in human and other apes, including chimpanzee and gorilla [26]. HERV appeared in the primate genome through germ-line infection [30]. There are approximately 98,000 HERVs in the human genome. Full-length HERVs are ~10 kb in length, but most of the HERVs existing in the human are defective due to truncation and accumulation of mutations during primate evolution [39]. Among various HERV subfamilies, HERV-K (HML2) is the youngest element in the human genome [8, 40]; 113 human-specific HERV-Ks were identified in the human genome, and among them, there were 15 and 98 full-length HERV-Ks and solitary LTRs, respectively [39]. These *de novo* TE insertions showed polymorphisms among human populations and even human individuals [31, 39, 41, 42]. Therefore, they have the potential to be used as a genetic marker for racial identification [31, 41, 42].

*De novo* TE insertions contribute to the genome expansion. Actually, some of them somewhat decreased their host genomes involving the insertion-mediated deletion of host genome sequences [32, 33]. Through comparative genomic analyses, 50 L1 insertion-mediated deletion events were found in the human and chimpanzee genomes [18]. The sizes of the deleted sequences were variable, and in sum, ~18 kb and ~15 kb of sequences were removed from the human and chimpanzee genomes, respectively. Based on the result, it was estimated that L1 insertions may have deleted up to 7.5 Mb of target genomic sequences during the primate radiation. *Alu* insertions were also involved in the genomic deletions at its insertion target regions through *Alu* retrotransposition-mediated deletion. A total of 33 deletion events responsible for a ~9,000-bp deletion in human and chimpanzee genomes were identified. It was suggested that *Alu* retrotransposition may have contributed to over 3,000 deletion events, leading to a ~900-kb deletion during primate evolution [17]. Additionally, 13 SVA insertion-mediated deletions (SIMDs) were also identified in the human genome, and they deleted 30,785 bp of the human genome compared with the chimpanzee genome (Table 1). Among the 13 SIMDs, 9 were associated with the SVA\_D subfamily, occupying the largest portion of SVAs, which suggests that

SIMD frequency is directly correlated to the copy number of SVA elements. Furthermore, one of the deletion events occurred in the *tMDC II* gene associated with sperm-egg binding prior to fertilization [43, 44].

After TE insertions into the host genome, they could generate genomic variations through unequal homologous recombination events between them [17, 19, 43]. The copy number of TEs is closely related to the frequency of the recombination between them. Thus, compared to other TEs, *Alu* and L1 elements have a high probability of generating genomic structural variations due to their ubiquity. In reality, 492 *Alu* recombination-mediated deletions (ARMDs) were identified in the human genome, and they deleted ~400 kb of human genomic sequences (Table 1). About 60% of the deletion events were related to known or predicted genes, including three that deleted functional exons. Thus, the ARMD process has produced a considerable portion of the genomic and phenotypic variations between humans and chimpanzees since the divergence of the two species [16]. The recombination between L1 elements has also deleted human genome sequences. Seventy-three L1 recombination-associated deletions (L1RADs) were identified in the human genome [45]. The sizes of the deletion events range from 56 to 64,113 bp, and ~450 kb of human genomic sequence was deleted through this L1RAD process (Table 1). Thus, the L1RAD event has deleted 25 times as much human genomic sequence as the L1 insertion-mediated deletion event [18, 45].

## Genomic Instability Generated by TEs

The TEs inserted in intra- and inter-genic regions could alter cellular gene expression, increasing genomic instability [9]. About a decade ago, gene regulation by TEs was studied only in specific genes through experimental validation. However, genome-wide analyses of gene regulation by TEs were recently conducted due to the developed high-throughput technologies. The findings showed that TEs have many regulatory sequences, such as promoters, enhancers, polyadenylation signals, and cryptic splicing donor (5') and acceptor (3') sites, by which the transcript architecture of nearby genes can be altered [10, 46, 47].

When TEs insert into the intronic region of genes, they could create a new exon by offering splicing sites, and this process is called "exonization" [48]. This mechanism is related to exon variations, such as cassette exons and intron retention in exons, increasing mRNA instability [49]. The TEs residing upstream of any gene could act as an alternative promoter, leading to new alternative transcripts with a new transcriptional start site [10]. Some TEs carry bidirectional promoters with transcription factor binding sites. For

example, LTR and L1 have sense promoters initiating their transcription and antisense promoters having the potential to initiate the transcription of other genes in the opposite direction [50, 51]. There is a microRNA gene cluster in human chromosome 19 (C19MC), over 100 kb in length. This cluster consists of the duplication of a core cassette, including a minus-strand *Alu* element. The cluster grew successfully during primate evolution, and the *Alu* element promotes microRNA expression by RNA polymerase III [52]. TEs not only initiate transcription but also terminate it by offering a polyadenylation signal. A full-length L1 contains 19 polyadenylation signals that could cause premature mRNA truncation [53]. The genes that contain TEs in their genic region have a tendency to produce various transcript forms, causing transcriptome diversity [10].

The orientation of TEs could be a factor affecting gene expression, which is well described by a “head-on collision” hypothesis. During DNA replication, DNA polymerase collides with RNA polymerase transcription complexes moving in the opposite direction to the movement of the DNA polymerase. It was observed that the collision slows down the DNA replication [54]. In cases where active TE exists in the opposite direction to its nearby genes, an RNA polymerase transcription complex transcribing the TE could encounter any of the RNA polymerase transcription complexes transcribing nearby genes, which could reduce the expression of the gene [54, 55]. In reality, transcription of the E-globin gene is repressed by an *Alu* element that has been inserted in the opposite direction to the gene [56]. On the other hand, TEs inserting in the same direction to its nearby genes show no effect on the expression of the genes [55].

Histone modification plays an important role in gene transcriptional regulation, and through this process, the host genome could regulate the activation of TEs [57, 58]. In reality, most TEs are accompanied by repressive histone modifications (e.g., H3K9me2 and H3K27me3), which cause the formation of heterochromatin. Conversely, TEs could affect the expression of host genes through histone modifications [59, 60]. The level of histone modifications was calculated in all families of human TEs, and older TE families carried more histone modifications than younger families. Interestingly, TEs proximal to genes carry more histone modifications than the ones that are distal to genes, which suggests that some epigenetic modifications of TEs may serve to regulate the expression of host genes [61].

DNA methylation is a strict silencing mechanism, and the host genome could use this process to repress the activation of TEs [61]. In general, DNA methylation occurs in a CpG dinucleotide. Because *Alu* and SVA elements have a high degree of CpG dinucleotides, they are vulnerable to methy-

lation [26, 62]. It was observed that TEs regain their activity to mobilize and regulate the expression of host genes when the silencing effect becomes slackened with increasing genomic instability. In addition, the demethylation of TEs is associated with human diseases, commonly in cancer [63-65].

miRNAs, one of the most active factors regulating gene expression, could be derived from TEs. Fifty-five genes derived from TEs were identified in the human genome, and their characterization showed that TE-derived miRNAs could potentially regulate the complex and dynamics of human genes [66].

## Conclusions

TEs have shown a variety of impacts on their host genomes. In this review, we describe HERV, *Alu*, L1, and SVA elements, which are thought to still be active in the human genome. A number of research studies related to TEs have shed new light on their amplification mechanisms and their function in primate genomes. Furthermore, recent research of TEs in the rhesus macaque genome provides a glimpse into their diversity and strong influence on the overall differences in genomic architecture between the Old World monkey (e.g., rhesus macaque) and hominid (e.g., human and ape) lineages [67]. The occurrence of *de novo* TE insertions, TE insertion-mediated deletions, and post-insertion recombination between TEs within the human and chimpanzee lineages has caused genetic alteration, lineage-specific genomic rearrangements, and phenotypic variations, further contributing to the divergence of humans and chimpanzees. As a whole, this review calls into question whether TEs should be considered “junk” DNA at all. Rather, TEs represent a potent evolutionary force associated with genomic fluidity in their host genomes.

## Acknowledgments

This work was supported by the Basic Science Research Program (no. 2011-0009080) through the National Research Foundation (NRF) of Korea, funded by the Ministry of Education, Science, and Technology.

## References

1. McClintock B. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol* 1956;21:197-216.
2. Alzohairy AM, Gyulai G, Jansen RK, Bahieldin A. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. *Plasmid* 2012 Aug 30 [Epub]. <http://dx.doi.org/10.1016/j.plasmid.2012.08.001>.
3. Chimpanzee Sequencing and Analysis Consortium. Initial se-

- quence of the chimpanzee genome and comparison with the human genome. *Nature* 2005;437:69-87.
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
  5. Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 2007;316:222-234.
  6. Mizuuchi K. Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu Rev Biochem* 1992;61:1011-1051.
  7. 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.
  8. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends Genet* 2007;23:183-191.
  9. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 2009;10:691-703.
  10. Warnefors M, Pereira V, Eyre-Walker A. Transposable elements: insertion pattern and impact on gene expression evolution in hominids. *Mol Biol Evol* 2010;27:1955-1962.
  11. Moran JV, DeBerardinis RJ, Kazazian HH Jr. Exon shuffling by L1 retrotransposition. *Science* 1999;283:1530-1534.
  12. Deininger PL, Batzer MA. *Alu* repeats and human disease. *Mol Genet Metab* 1999;67:183-193.
  13. Venner S, Feschotte C, Biémont C. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet* 2009;25:317-323.
  14. Baskaev KK, Buzdin AA. Evolutionary recent insertions of mobile elements and their contribution to the structure of human genome. *Zh Obshch Biol* 2012;73:3-20.
  15. Lee J, Cordaux R, Han K, Wang J, Hedges DJ, Liang P, et al. Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene* 2007;390:18-27.
  16. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, et al. Human genomic deletions mediated by recombination between *Alu* elements. *Am J Hum Genet* 2006;79:41-53.
  17. Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA. *Alu* retrotransposition-mediated deletion. *J Mol Biol* 2005;348:791-800.
  18. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, et al. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res* 2005;33:4040-4052.
  19. Gilbert N, Lutz-Prigge S, Moran JV. Genomic deletions created upon LINE-1 retrotransposition. *Cell* 2002;110:315-325.
  20. O'Donnell KA, Burns KH. Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mob DNA* 2010;1:21.
  21. Ahmed M, Liang P. Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp Funct Genomics* 2012;2012:947089.
  22. Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, et al. Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* 2006;78:671-679.
  23. Carter AB, Salem AH, Hedges DJ, Keegan CN, Kimball B, Walker JA, et al. Genome-wide analysis of the human *Alu* Yb-lineage. *Hum Genomics* 2004;1:167-178.
  24. Han K, Xing J, Wang H, Hedges DJ, Garber RK, Cordaux R, et al. Under the genomic radar: the stealth model of *Alu* amplification. *Genome Res* 2005;15:655-664.
  25. Cordaux R. The human genome in the LINE of fire. *Proc Natl Acad Sci U S A* 2008;105:19033-19034.
  26. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, et al. SVA elements: a hominid-specific retroposon family. *J Mol Biol* 2005;354:994-1007.
  27. Hancks DC, Kazazian HH Jr. SVA retrotransposons: evolution and genetic instability. *Semin Cancer Biol* 2010;20:234-245.
  28. Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. Natural genetic variation caused by transposable elements in humans. *Genetics* 2004;168:933-951.
  29. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, et al. Mobile elements create structural variation: analysis of a complete human genome. *Genome Res* 2009;19:1516-1526.
  30. Kim HS. Genomic impact, chromosomal distribution and transcriptional regulation of HERV elements. *Mol Cells* 2012;33:539-544.
  31. Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, et al. Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Res* 2003;13:1607-1618.
  32. Srikanta D, Sen SK, Huang CT, Conlin EM, Rhodes RM, Batzer MA. An alternative pathway for *Alu* retrotransposition suggests a role in DNA double-strand break repair. *Genomics* 2009;93:205-212.
  33. Sen SK, Huang CT, Han K, Batzer MA. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res* 2007;35:3741-3751.
  34. Callinan PA, Batzer MA. Retrotransposable elements and human disease. *Genome Dyn* 2006;1:104-115.
  35. Bergman CM. A proposal for the reference-based annotation of *de novo* transposable element insertions. *Mob Genet Elements* 2012;2:51-54.
  36. Rogers J, Kochunov P, Zilles K, Shelledy W, Lancaster J, Thompson P, et al. On the genetic architecture of cortical folding and brain volume in primates. *Neuroimage* 2010;53:1103-1108.
  37. Britten RJ. Transposable element insertions have strongly affected human evolution. *Proc Natl Acad Sci U S A* 2010;107:19945-19948.
  38. Boissinot S, Chevret P, Furano AV. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol* 2000;17:915-928.
  39. Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J Virol* 2005;79:12507-12514.
  40. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribu-

- tion of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* 2011;8:90.
41. Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, *et al.* A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet* 2011;7:e1002236.
  42. Konkel MK, Wang J, Liang P, Batzer MA. Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. *Gene* 2007;390:28-38.
  43. Lee J, Ha J, Son SY, Han K. Human genomic deletions generated by SVA-associated events. *Comp Funct Genomics* 2012; 2012:807270.
  44. Frayne J, Hall L. The gene for the human tMDC I sperm surface protein is non-functional: implications for its proposed role in mammalian sperm-egg recognition. *Biochem J* 1998; 334(Pt 1):171-176.
  45. Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA. L1 recombination-associated deletions generate human genomic variation. *Proc Natl Acad Sci U S A* 2008;105:19366-19371.
  46. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 2008;9:397-405.
  47. Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 2011;12: 615-627.
  48. Schmitz J, Brosius J. Exonization of transposed elements: a challenge and opportunity for evolution. *Biochimie* 2011;93: 1928-1934.
  49. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev* 2003;17:419-437.
  50. Mätlik K, Redik K, Speek M. L1 antisense promoter drives tissue-specific transcription of human genes. *J Biomed Biotechnol* 2006;2006:71753.
  51. Domansky AN, Kopantzev EP, Snezhkov EV, Lebedev YB, Leib-Mosch C, Sverdlov ED. Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett* 2000;472:191-195.
  52. Lehnert S, Van Loo P, Thilakarathne PJ, Marynen P, Verbeke G, Schuit FC. Evidence for co-evolution between human microRNAs and *Alu*-repeats. *PLoS One* 2009;4:e4456.
  53. Kaer K, Branovets J, Hallikma A, Nigumann P, Speek M. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. *PLoS One* 2011;6:e26099.
  54. Liu B, Alberts BM. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* 1995;267:1131-1137.
  55. Cui F, Sirotnin MV, Zhurkin VB. Impact of *Alu* repeats on the evolution of human p53 binding sites. *Biol Direct* 2011;6:2.
  56. Wu J, Grindlay GJ, Bushel P, Mendelsohn L, Allan M. Negative regulation of the human epsilon-globin gene by transcriptional interference: role of an *Alu* repetitive element. *Mol Cell Biol* 1990;10:1209-1216.
  57. Kondo Y, Issa JP. Enrichment for histone H3 lysine 9 methylation at *Alu* repeats in human cells. *J Biol Chem* 2003;278: 27658-27662.
  58. Pauler FM, Sloane MA, Huang R, Regha K, Koerner MV, Tamir I, *et al.* H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res* 2009;19:221-233.
  59. Huda A, Bowen NJ, Conley AB, Jordan IK. Epigenetic regulation of transposable element derived human gene promoters. *Gene* 2011;475:39-48.
  60. Huda A, Mariño-Ramírez L, Jordan IK. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. *Mob DNA* 2010;1:2.
  61. Huda A, Tyagi E, Mariño-Ramírez L, Bowen NJ, Jjingo D, Jordan IK. Prediction of transposable element derived enhancers using chromatin modification profiles. *PLoS One* 2011;6:e27513.
  62. Xing J, Hedges DJ, Han K, Wang H, Cordaux R, Batzer MA. *Alu* element mutation spectra: molecular clocks and the effect of DNA methylation. *J Mol Biol* 2004;344:675-682.
  63. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, *et al.* Hypomethylation of a *LINE-1* promoter activates an alternate transcript of the *MET* oncogene in bladder with cancer. *PLoS Genet* 2010;6:e1000917.
  64. Matsuda Y, Yamashita S, Lee YC, Niwa T, Yoshida T, Gyobu K, *et al.* Hypomethylation of *Alu* repetitive elements in esophageal mucosa, and its potential contribution to the epigenetic field for cancerization. *Cancer Causes Control* 2012;23:865-873.
  65. Bae JM, Shin SH, Kwon HJ, Park SY, Kook MC, Kim YW, *et al.* *ALU* and *LINE-1* hypomethylations in multistep gastric carcinogenesis and their prognostic implications. *Int J Cancer* 2012;131:1323-1331.
  66. Piriyaopongsa J, Mariño-Ramírez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. *Genetics* 2007;176:1323-1337.
  67. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, *et al.* Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. *Science* 2007;316:238-240.