Mobile Genetic Elements 3:4, e25674; July/August 2013; © 2013 Landes Bioscience

Non-LTR retrotransposons and microsatellites

Partners in genomic variation

Fiorella C. Grandi and Wenfeng An*

School of Molecular Biosciences and Center for Reproductive Biology; Washington State University; Pullman, WA USA

Keywords: Alu, genomic instability, genomic variation, LINE-1, microsatellite, mononucleotide repeat, mutation rate, poly(A), retrotransposition, retrotransposon

The human genome is laden with both non-LTR (long-terminal repeat) retrotransposons and microsatellite repeats. Both types of sequences are able to, either actively or passively, mutagenize the genomes of human individuals and are therefore poised to dynamically alter the human genomic landscape across generations. Non-LTR retrotransposons, such as L1 and Alu, are a major source of new microsatellites, which are born both concurrently and subsequently to L1 and Alu integration into the genome. Likewise, the mutation dynamics of microsatellite repeats have a direct impact on the fitness of their non-LTR retrotransposon parent owing to microsatellite expansion and contraction. This review explores the interactions and dynamics between non-LTR retrotransposons and microsatellites in the context of genomic variation and evolution.

Introduction: Non-LTR Retrotransposons and Microsatellites

The human genome is laden with repetitive sequences in the form of transposable elements (TEs) and tandem repeats, which comprise 45% and 3% of the human genome, respectively.[1] Using new methods capable of annotating repetitive sequences that have substantially diverged from known TE sequences, recent analyses indicate that up to 66–69% of the human genome may be composed of repetitive sequences, predominately in the form of TEs.[2]

TEs, as their name implies, are able to move about in the genome and create interspersed repeats. TEs include DNA transposons, long-terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. In this review, we will focus on non-LTR retrotransposons, which are the only active TEs in the human genome. Non-LTR retrotransposons include long interspersed elements (LINEs) and short interspersed elements (SINEs).[3-5] Both mobilize via a "copy and paste" mechanism, which requires the transcription of a donor element into an RNA intermediate and the subsequent reverse transcription and incorporation of the RNA intermediate into the genome as a new DNA copy (i.e., the retrotransposition process). Among them, LINE-1s (L1s) are autonomous; full-length L1s encode two proteins, ORF1 and ORF2, which are essential for their mobilization (**Fig. 1A**). In contrast, SINEs are non-autonomous and their mobilization relies on L1 proteins. There are two main classes of SINEs.[6,7] The first class includes human Alu (**Fig. 1B**) and mouse B1, both of which are derived from 7SL RNA; the second class is typified by mouse B2 elements, which are derived from tRNAs. The quantity and mobility of non-LTR retrotransposons have made them important drivers of genomic diversity and evolution.[8,9]

In contrast to interspersed repeats, tandem repeats consist of sequences that are sequentially repeated. Tandem repeats are classified by the size of their repeating unit: those with a repeating unit length of 1–6 base pairs (bp) are commonly referred to as microsatellites (**Fig. 2A**).[10-12] However, the boundary between microsatellites and minisatellites is often disputed, where the latter typically refer to tandem repeats of greater unit sizes.[13] As they are highly polymorphic, both microsatellites and minisatellites have been instrumental in genetic mapping, evolutionary and phylogenetic studies as well as DNA forensics.[14] This review focuses on microsatellites, which encompass mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeats. Notably, mononucleotide A repeats (i.e., poly(A) DNA tracts) represent the most abundant microsatellite class with an overall length of 12 bp in the human genome.[15] Due to the repeating nature of these DNA tracts, they are prone to rapid and variable contraction and expansion of repeat length (**Fig. 2B**).

Collectively, the dynamics of these two entities are a major source of genomic variation and serve as the architectural framework for protein-coding genes. Recent work has revealed an intimate relationship between non-LTR retrotransposons and microsatellites. On one hand, non-LTR retrotransposons are a major source of new microsatellites as they give birth to microsatellites both concurrent and subsequent to their integration into the genome. On the other hand, the inherent instability of microsatellite repeats has a direct impact on the fitness of their non-LTR retrotransposon parent. This review will focus on the interactions and mutation dynamics between non-LTR retrotransposons and microsatellites as well as the variability and rates of mutation of microsatellite repeats that make them instigators of genomic evolution.

*Correspondence to: Wenfeng An; Email: wenfengan@vetmed.wsu.edu
Submitted: 04/22/13; Revised: 07/07/13; Accepted: 07/09/13
Citation: Grandi FC, An W. Non-LTR retrotransposons and microsatellites: Partners in genomic variation. Mobile Genetic Elements 2013; 3:e25674; http://dx.doi.org/10.4161/mge.25674

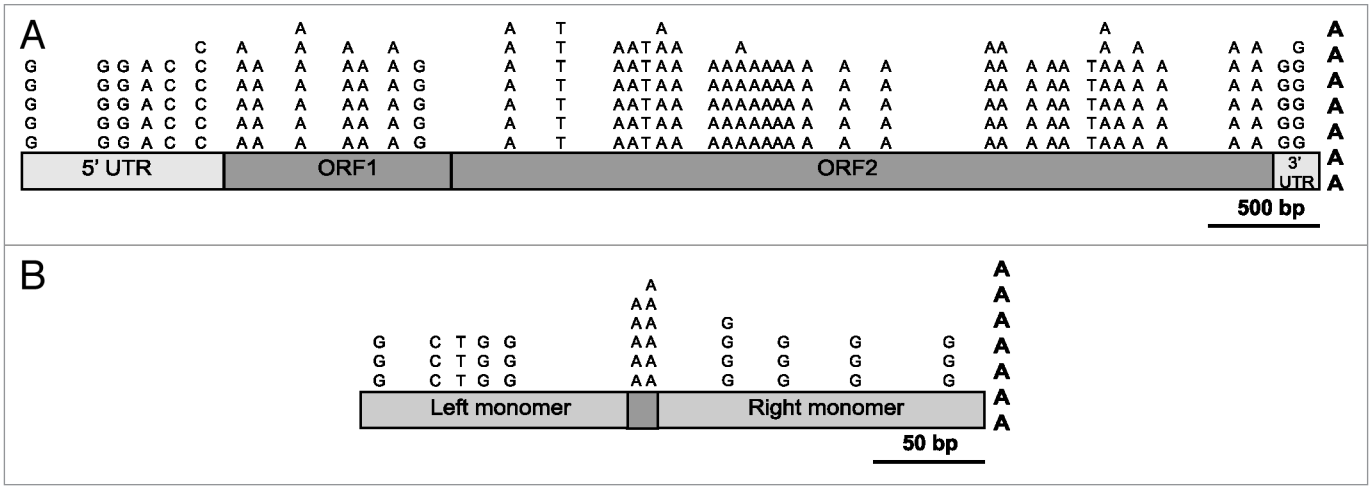

Figure 1. Non-LTR retrotransposons are hotbeds for new microsatellites. (**A**) The locations of mononucleotide microsatellite seed sequences (i.e., 5–7 bp proto-microsatellites) are diagramed along the length of the consensus human L1 sequence. Five proto-microsatellites within the sequence are 7 units long, only one nucleotide below the threshold for mononucleotide microsatellites. A long poly(A) tail is highlighted at the 3' end of the L1 element (in bold font; only 7 A's shown due to space limitation). (**B**) The locations of 3–6 bp mononucleotide proto-microsatellites are diagramed along the length of the consensus Alu sequence. Unlike L1, Alu is G/C rich and only two poly(A) proto-microsatellites are found in the middle linker region between left and right monomers. Over time, the short proto-microsatellites may expand beyond the microsatellite threshold. A long poly(A) tail is highlighted at the 3' end of the Alu element (in bold font; only 7 A's shown due to space limitation). Note the 10-fold difference in scale between L1 and Alu elements.

Non-LTR Retrotransposons Give Birth to Microsatellites

Relatively little attention has been paid to the evolutionary origins of microsatellites. It is proposed that each microsatellite locus possesses a lifecycle that begins with its "birth" into the genome and ends with its "death" (**Fig. 2B**).[16] Most studies focus on the intermediate "growth" phase, a period that is punctuated by contractions and expansions, which dynamically shorten and lengthen the overall microsatellite length. The most commonly proposed mechanism for microsatellite contractions and expansions is replication slippage by DNA polymerases.[10] The main factors involved in replication slippage will be discussed in the next section. Two other mutation mechanisms, base substitution and indel slippage, are also thought to play an important role in microsatellite birth and death.[12,17] Overall, two distinct pathways have been postulated for microsatellite genesis (**Fig. 2B**). One is through de novo retrotransposition by non-LTR retrotransposons, while the other involves the birth of proto-microsatellites from random sequences.[10,16] It is the goal of this review to integrate these two pathways and highlight the role of non-LTR retrotransposons in microsatellite genesis.

Non-LTR retrotransposons can give birth to microsatellites concurrently to their integration into the genome (i.e., the first pathway; **Fig. 2B**). The association between A-rich microsatellites and non-LTR retrotransposons has been repeatedly observed by many independent studies.[18-24] A recent analysis of microsatellites at orthologous loci in three primate genomes estimated that upon retrotransposition, Alus and L1s contribute to 36% of mono-, di, tri- and tetranucleotide microsatellites in human, chimpanzee and orangutan genomes.[17] This pathway is best illustrated by the birth of poly(A) microsatellites. To mobilize, non-LTR retrotransposons require a poly(A) tail at the 3' end of their sequence (discussed later).[3,25] Thus, a common feature of Alu and L1 insertions that have integrated into the genome is the presence of a 3' poly(A) DNA tail (**Fig. 1**). These A-rich sequences are born into the genome as mature microsatellites, far above the threshold size (**Fig. 3**). In humans, 90% of these poly(A) repeats are derived from poly(A) tails of L1 and Alu sequences.[26] Even this may be an underestimate because L1 can integrate a pure poly(A) DNA tract into the genome if the insertion is severely 5' truncated;[27] such events would typically be undetected. The abundance of L1 and Alu sequences in the primate genomes likely accounts for the overrepresentation of poly(A)/poly(T) repeats compared with poly(C)/poly(G) repeats.[28] In contrast, dinucleotide microsatellites are most abundant in rodents, and AC repeats are the most frequent dinucleotide repeat in all vertebrates.[28] Not surprisingly, 15% of AC repeats are derived from the AC-rich tails of B2 SINE elements found in the mouse genome.[26]

The second pathway posits that microsatellites initially arise from random sequences via mutation as proto-microsatellites, small tandem repeated sequences that are below the length threshold for mature microsatellites (**Fig. 2B**).[29,30] Theoretically, these initiating loci can be found everywhere in the genome. However, the overwhelming abundance of TEs in the human genome made it the perfect hotbed for microsatellite births.[1,2] It has been suggested that the conversion of AAA to NAA is a major mechanism for the generation of all A-rich trinucleotide repeats in the human genome as over 60% of GAA, CAA and TAA repeats with repeat number of 8 or higher are located within Alu poly(A) tails.[31] Furthermore, poly(A) tails have also been implicated in the birth of dinucleotide and tetranucleotide microsatellites.[21,22] Aside from their 3' tails, both Alu and L1 have internal proto-microsatellite sequences that can birth new microsatellites subsequent to their integration (**Fig. 1**; **Fig. 3**). An analysis of microsatellites at orthologous loci in three primate

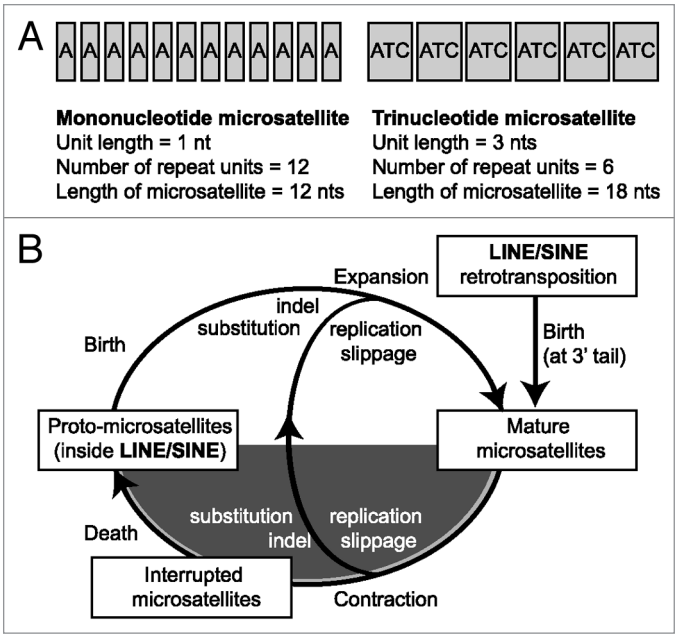

Figure 2. The anatomy and lifecycle of a microsatellite. (**A**) Microsatellite anatomy and terminology. A mononucleotide microsatellite and a trinucleotide microsatellite are depicted. Each repeating unit is depicted as a box. The repeat unit length, the number of repeating units, and the overall length are noted for each microsatellite. Both microsatellites have the same number of repeating units, but the unit length and the overall length vary among them. (**B**) The birth and death lifecycle for microsatellites. Two pathways lead to microsatellite birth. The first involves LINE/SINE retrotransposition; a microsatellite, typically a mature poly(A) repeat, is born into the genome concurrently upon LINE/SINE integration. The second pathway involves birth from random sequences (not shown) and/or proto-microsatellites via base substitution, indel and replication slippage. When a poly(A) tract reaches 8–9 A's, it is considered a mature microsatellite. As the length increases, length contractions outweigh expansions, or vice versa, forming an indefinite loop (thus, bypassing the path to death). Microsatellite sequences can also be interrupted by mutations. The interrupted microsatellites may recover and continue expanding (not shown) or may be further mutated and eventually die.

genomes indicates that 26% of microsatellite births and 24% of microsatellite deaths occur within Alu and L1 sequences subsequent to retrotransposition.[17] It is noteworthy that microsatellites are found in distinct distributions over the length of L1 and Alu elements, likely reflecting differences in their sequence compositions.[17] L1 sequences are AT-rich and give rise to microsatellites during and after their insertion along their entire length (**Fig. 1A**).[17] In contrast, Alus are GC-rich and microsatellite births and deaths are enriched at the 3' poly(A) tail and the middle A stretch that connects the left and the right monomers (**Fig. 1B**).[17,22,23,32] The middle A-rich region of Alu elements tends to expand and mature into poly(A) microsatellites[32] and can also give rise to GAA repeats.[31] Approximately 2% of GAA repeats with a repeat number of 8 or longer are located in the Alu middle A tract; one of such GAA microsatellite is positioned in intron 1 of the *FRDA* gene and its expansion is responsible for the neuromuscular disorder Friedreich's ataxia.[31]

Microsatellite Mutation Dynamics

An important characteristic of microsatellites is their high mutation rates. Classic studies on microsatellite mutation rates utilize pedigree analysis, usually in cancer patients with microsatellite instability. Several model systems, including *E. coli*, yeast, human and mouse cells have also been described and used to obtain in vivo mutation rates. In addition, computational and mathematical models have been utilized to describe the behavior of microsatellites by comparing genomic sequences from large scale sequencing efforts and from evolutionarily related species. The majority of computational studies and in vivo studies

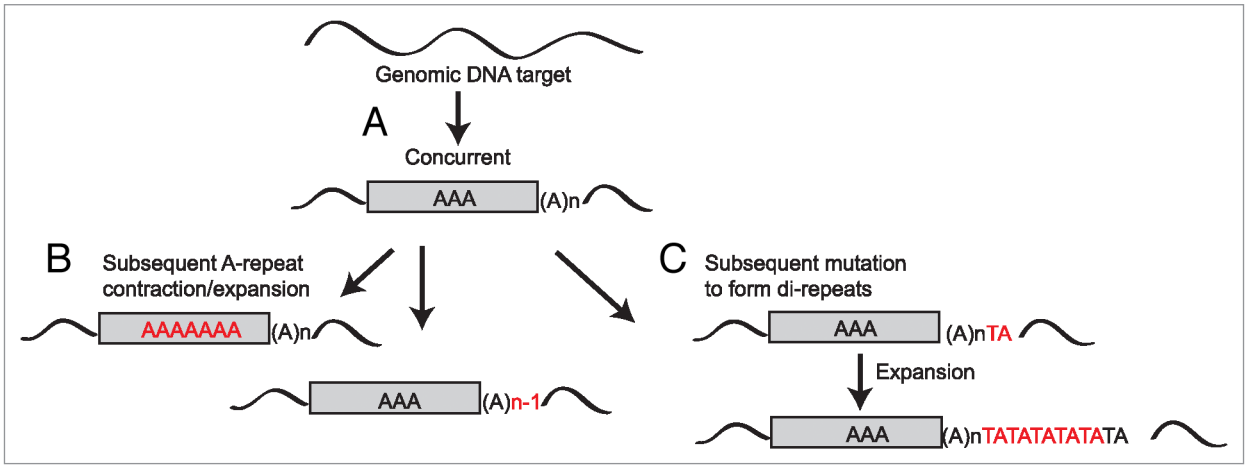

Figure 3. Non-LTR retrotransposons can give birth to microsatellites both concurrently and subsequently to their integration in the genome. (**A**) Concurrent birth occurs via the integration of a long poly(A) tract, which is part of the Alu or L1 element. Additionally, the integrated element carries proto-microsatellites, depicted by the middle A stretch. Here, the waved line represents genomic DNA target, and the gray box represents the body of an Alu or L1 element. (**B**) Subsequent to integration, both the poly(A) tail microsatellite and the internal proto-microsatellites may expand or contract due to DNA polymerase slippage. This process creates a variety of microsatellite length polymorphisms in somatic and germ cells. (**C**) Subsequent to integration, the poly(A) tail microsatellite may also be mutated via base substitution or indel to form a new repeating unit, here depicted as TA at the end of the poly(A) tract. The combination of mutational forces may eventually lead to the formation a new dinucleotide microsatellite at the 3' end of the retrotransposon.

Table 1. Summary of major studies characterizing the DNA polymerase derived slippage rate of microsatellite repeats in mouse and human model systems

Biological system	Repeat size	Repeat sequence	Rate	Units	Ref.
Mononucleotide repeats					
Human lymphoblastoid cell line	10	C	2.9E-05	per generation	12
Human pedigree	~40	A	6.8E-02	per locus per generation	38
Mouse pedigree with L1 transgene	> > 50	A	1.33 to 1.48E-01[a]	per locus per generation	39
Dinucleotide repeats					
Human lymphoblastoid cell line	11	GA	3.2E-06	per generation	12
Mouse fibroblast cell line (CAK-Stu3)	17	GA	3.0E-05	per cell per generation	40
H6 cells (colon tumor origin), mismatch repair deficient	17	GA	1.6E-04	per locus per generation	41
Human lymphoblastoid cell line	17	GA	9.8E-06	per generation	12
Human lymphoblastoid cell line	20	GA	2.1E-04	per generation	12
Mouse fibroblast cell line (CAK-Stu3)	8	CA	3.7E-06 (7.9E-08)[b]	per cell per generation	42
Human lymphoblastoid cell line	10	CA	2.1E-07	per generation	43
Human lymphoblastoid cell line	13	CA	6.9E-07	per generation	43
Human lymphoblastoid cell line	16	CA	3.4E-06	per generation	12
Mouse fibroblast cell line (CAK-Stu3)	17	CA	2.4E-05 (1.7E-05)[b]	per cell per generation	42
Mouse fibroblast cell line (CAK-Stu3)	30	CA	1.6E-04 (3.8E-05)[b]	per cell per generation	42
Human pedigree analysis; 15 autosomal locations	variable	CA	5.6E-04	per generation	44
Trinucleotide repeats					
Human embryonic kidney cells (293T)	33	CAG	5.7E-03	per generation	45
Human embryonic kidney cells (293T)	33	CAG	5.0E-04	per generation	45
Tetranucleotide repeats					
CAK mouse cells	17	GAAA	1.2E-06	per cell per generation	46
H6 cells (colon tumor origin), mismatch repair deficient	17	GAAA	3.3E-05	per cell per generation	46
Pedigree analysis; Y chromosome sperm	8 to 15	GATA	4.0E-03	per generation	47
Pedigree analysis; autosomal locations	variable	GATA	2.1E-03	per generation	44
human lymphoblastoid cell line	9	TTCC	5.6E-06	per generation	12
human lymphoblastoid cell line	9	TTTC	3.5E-05	per generation	12
human lymphoblastoid cell line	9	TCTA	4.8E-05	per generation	12

[a]Contraction rate; [b]Number in parenthesis is the contraction rate for the same repeat; [c]Trinucleotide studies focus primarily on mismatch repair deficient cells. Trinucleotide repeats that cause disease tend to be more unstable in the germline than in somatic cell lines. This review does not focus on the mismatch repair component of microsatellite slippage.

focus exclusively on pure or perfect repeats, which contain only nucleotides pertaining to the repeating unit. Imperfect repeats, in contrast, have additional nucleotides that do not belong to the repeating units. Overall, the rate of microsatellite mutation is highly variable, and it is influenced by several factors, including the number of repeating units, unit repeat length, sequence composition, genomic loation and the secondary structures of the repeat.[12,33,34]

The number of repeating units is a major factor in microsatellite mutation rates. Specifically, microsatellites become highly mutable when the number of repeating units exceeds a threshold value.[33,35,36] Although there is disagreement on the exact threshold value, the consensus is that the threshold represents the size that allows the sequence to acquire slippage mutations at a higher rate than the genomic average. The threshold values for mono-, di-, tri- and tetranucleotide repeats are estimated to be 8–9, 4–5, 3.4–4 and 4 respectively.[35-37] In fact, the distinct mutational behaviors manifested by microsatellite before and after they reach the threshold have prompted some researchers to consider microsatellites only as those that exceed their respective threshold.[33,36] Once a tandem repeat has reached its threshold, experimental studies with human and mouse models have shown that the mutation rate increases exponentially as the number of repeating units increases (**Table 1**).[12,33,39-47] The same trend is also observed in computational studies.[33,35,36] However, due to the differences in the experimental systems used, it is difficult to compare across

these studies. The exponential increase in mutation rates is not seen in mismatch repair deficient cells, indicating that mismatch repair proteins are implicated in the expansion process (**Table 1**).[42]

Sequence composition also affects mutability. For similar sized repeats, GA repeats are more mutable than CA repeats and sequences with higher GC content have lower rates of mutation (**Table 1**).[12,43] DNA structure and thermodynamic stability have been implicated in explaining this phenomenon.[12,43] Stability increases with the complexity of the repeating pattern, as demonstrated in tetranucleotide repeat studies. Despite a doubling of repeat unit length, di- and tetranucleotide repeats have similar stabilities (**Table 1**; TC vs. TTCC).[12] In contrast, mononucleotide microsatellites, such as those associated with retrotransposons, are highly unstable and hypermutable in the germline.[38,39] For example, the human BAT-40 locus is an intronic poly(A) tract of 40 A's in the 3-β-hydroxysteroid dehydrogenase gene (HSD3B1).[48] Its instability is well documented in tumor cells deficient in mismatch repair. This locus is also highly polymorphic among healthy individuals. In fact, pedigree analysis indicates BAT-40 is inherently unstable in the germline. Five-fold more mutant alleles are found in sperm than in leukocytes of the same individual.[38] The BAT-40 poly(A) microsatellite has not been identified to be associated with non-LTR retrotransposons; here, we report that it is in fact the poly(A) tail of an AluJb element.

Despite experimental focus on the expansion dynamics of microsatellites, expansion is not indefinite; instead, longer alleles show a bias toward contraction.[10] The mutation dynamics for very long (> 50 bp) mononucleotide repeats had been frequently overlooked because of their rarity in the genome. Using a transgenic L1 mouse model, we characterized the contraction dynamics of long poly(A) tracts associated with nascent L1 insertions.[39] These poly(A) microsatellites show rapid and variable contraction to lengths below 50 bp in both somatic and germ cells at frequencies of 14.8% and 13.3%, respectively.[39] Of note, these values represent the highest mutation frequencies hitherto reported for all microsatellite repeats, highlighting the extraordinary mutability of long poly(A) tracts (**Table 1**). Thus, long poly(A) microsatellites birthed into the genome via retrotransposition are predicted to shorten in length rapidly within the first few generations.[39] This prediction is consistent with observations from surveys of genomic resident Alu and L1 sequences, which demonstrated an inverse correlation of poly(A) tract length with the evolutionary age of the elements.[27,49,50]

One prominent form of microsatellite death is interruption by a non-repeat base, which can arise from base substitutions or short indel mutations.[12,17,51,52] In particular, a comparison of human, chimpanzee and orangutan genomes suggests that substitutions are the primary cause of deaths for microsatellites of all lengths.[17] Although not widely recognized, non-LTR retrotransposons may also act as a mechanism for microsatellite interruption. Such interruption can occur when new copies of non-LTR retrotransposons insert into pre-existing microsatellites. Mammalian L1s are members of a diverse and widely distributed group of autonomous non-LTR retrotransposons that encode an endonuclease with homology to apurinic/apyrimidinic endonucleases (APE).[53] A small subset of these APE-type elements insert in a sequence-specific manner into microsatellites: (1) ACAY microsatellites are targeted by three Waldo elements (WaldoAg1, WaldoAg2 in the African malaria mosquito *A. gambiae* and WalsoFs1 in the earwig *F. scudderi*);[54] (2) AC microsatellites are targeted by a Mino element in *A. gambiae* (MinoAg1);[54] (3) TC microsatellites are targeted by four Kibi elements (KibiDr1 and KibiDr2 in zebrafish *D. rerio*, KibiFr1 in the torafugu *F. rubripes*, and KibiTn1 in the green spotted pufferfish *T. nigroviridis*);[55] (4) TTC microsatellites are targeted by two Koshi elements (KoshiFr1 in *F. rubripes* and KoshiTn1 in *T. nigroviridis*);[55] (5) TAA microsatellites are targeted by three Dong elements (Dong in the silkworm *B. mori*, DongAg in *A. gambiae* and DongBg in the freshwater snail *B. glabrata*).[55,56] The molecular mechanisms underlying microsatellite targeting by these APE-type non-LTR retrotransposons are not well understood. It is believed that the sequence-specific cleavage by the endonuclease plays a major role in active targeting.[53] In addition, two APE-type elements (Waldo-A and Waldo-B in the fruitfly *D. melanogaster*) are frequently found within or close to CA microsatellites, most likely as a result of passive accumulation in these low recombination regions.[57,58] Microsatellites may also be targeted by DNA transposons; examples include MINE-1 helitrons and Micron MITEs. The former insert at GAAA microsatellites in the genomes of the European corn borer *O. nubilalis*, the silkworm *B. mori* and the pink bollworm *P. gossypiella*,[59] while the latter insert into TA microsatellites in the rice genome.[60,61] In contrast, mammalian L1s have a weak target site preference with a consensus sequence 5'-TTAAAA-3'.[53,62-64] Although they may fortuitously land in or near certain microsatellite sites, they are unlikely to serve as a major factor in disrupting microsatellites.

Microsatellite Instability Affects Non-LTR Retrotransposon Mobility

The relationship between microsatellites and non-LTR retrotransposons is not unidirectional. While both L1s and Alus give birth to microsatellites, especially poly(A) mononucleotide microsatellites, these microsatellite sequences can also affect the fitness of their "parent" due to their unusually high mutation rates. The impact of microsatellite instability on non-LTR retrotransposons depends on the location of the microsatellite, i.e., whether it is internal or at the 3' terminal of the element. The effect of variation in microsatellites internal to non-LTR retrotransposons is less understood. Newly inserted L1 and Alu copies carry many mononucleotide proto-microsatellites (**Fig. 1**). As microsatellites are predicted to mutate faster than the genomic average, expansion and contraction of microsatellite loci internal to L1s may introduce frameshift mutations, abolishing L1 coding capacity (**Fig. 4A**).

The 3' poly(A) tail of an L1 or Alu element is a crucial component of the retrotransposition process and therefore, its length directly affects their retrotransposition potential. Distinct cellular processes are responsible for the poly(A) tail formation in L1 and Alu elements. L1 elements are transcribed by RNA polymerase II and poly(A) polymerases produce a poly(A) RNA tail as the 3' end of an L1 mRNA. In contrast, Alu elements are

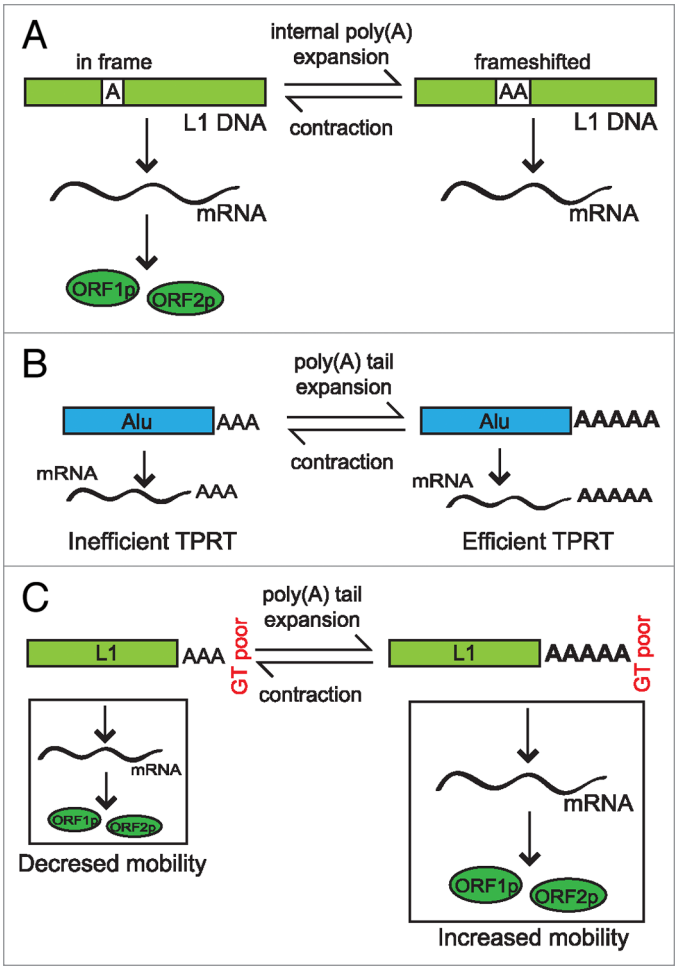

Figure 4. Microsatellite instability alters the retrotransposition potential of non-LTR retrotransposons. (**A**) Effect of internal microsatellite loci. Expansions or contractions of proto-microsatellite loci inside an L1 element can cause frameshift mutations, leading to loss of ORF1 and/or OFR2 function. The mutagenized L1 element would be unable to mobilize itself and to support Alu retrotransposition. (**B**) Effect of the 3' poly(A) microsatellite on Alu elements. The contraction or expansion of an Alu's poly(A) tail diminishes or stimulates its retrotransposition, respectively, presumably because the length of its poly(A) RNA tail is positively correlated with the efficiency of target-primed reverse transcription (TPRT). (**C**) Effect of the 3' poly(A) microsatellite on L1 elements. Due to the random nature of integration, an L1 sequence may not have a GT rich region downstream to its polyadenylation signal. However, there is evidence that the L1 poly(A) tail exerts a crucial role in ensuring proper polyadenylation of L1 mRNA in the absence of a strong GT-rich downstream sequence. The relative efficiency in mobilization is depicted by a difference in size of L1 mRNA and protein products.

transcribed by RNA polymerase III and the resulting transcripts are not polyadenylated.[3] However, active Alu elements have a poly(A) DNA tract, which is transcribed as part of the Alu RNA. The 3' poly(A) RNA tail is predicted to serve two important roles during retrotransposition. First, the initial base pairing of these A's with the T-rich DNA sequence at the target site may be required for efficient first-strand cDNA synthesis during target-primed reverse transcription (TPRT).[64-66] The L1 ORF1 protein is an RNA-binding protein with nucleic acid chaperone activity and it may facilitate this strand transfer and annealing process.[65] Second, increasing evidence indicates that these poly(A) RNA tails are bound by poly(A) binding proteins (PABPs) and that this interaction is critical for the formation of the ribonucleoprotein complex between L1 proteins and L1/Alu RNAs.[67,68] PABPC1 may also facilitate the nuclear import of L1 RNP.[67,69] Indeed, retrotransposition assays demonstrate that the poly(A) tail is strictly required for Alu mobilization and that its retrotransposition activity is positively correlated with the length of poly(A) tails.[70,71] In parallel, poly(A) tail shortening in expressed L1 mRNAs impairs RNP formation and retrotransposition.[67]

Therefore, the high mutability of poly(A) tails has direct impact on L1 and Alu retrotransposition. Once captured in genomic DNA, the initial long 3' poly(A) tract undergoes rapid shortening in the first few generations.[39] In fact, genome-wide, the length of the poly(A) tract is inversely correlated with the evolutionary age of Alu and L1 subfamilies.[49,50] As the length of the poly(A) tail is a key determinant of Alu activity, A-tail expansion due to microsatellite instability is predicted to promote its retrotransposition and, conversely, its contraction would predict the extinction of an Alu's ability to mobilize (**Fig. 4B**). Indeed, newly generated Alu copies typically have longer poly(A) tails than the original donor elements.[72] Additionally, a nascent retrotransposition-competent Alu element is able to propagate its newfound mobility to progeny copies through poly(A) tail expansion during reverse transcription.[72]

What impact does poly(A) tail instability have on L1 retrotransposition? It is generally accepted that the L1 poly(A) tail is regenerated during each round of retrotransposition via reverse transcription of the polyadenylated L1 mRNA. Thus, the length of the poly(A) tail in the donor L1 element may affect L1 retrotransposition only if it alters transcription and polyadenylation. A canonical polyadenylation signal contains a conserved AATAAA hexamer and a downstream GT-rich region. The L1 polyadenylation signal is atypical in that a poly(A) tail separates the AATAAA hexamer from any downstream sequence; the sequence composition of the latter is determined by chance depending on where the donor element is inserted during the previous round of retrotransposition. Such a configuration may predispose an L1 element to bypass its own polyadenylation signal and instead to use a stronger polyadenylation signal fortuitously situated further downstream.[73] Nevertheless, the poly(A) tail appears to exert a critical role in proper polyadenylation of L1 mRNA in the absence of a strong GT-rich downstream sequence (**Fig. 4C**).[74] Although full length transcription is possible in the context of a weak polyadenylation context, the presence of a string of 17 A's brings about significant increases in the transcriptional efficiency.[74] Although its presence and length is not as critical for L1 mobility as for Alu mobility, the poly(A) tail's dynamics still subtly influence the efficiency of its parent in a disadvantageous genomic neighborhood.

Genomic Distribution and Polymorphisms

The availability of complete genome sequences for many organisms has enabled genome-wide analyses of repeat

distribution. Microsatellites with overall repeat length of 12 bp or longer make up 1.3% of the reference human genome (i.e., 13.3 kb per Mb genome, excluding hexameric repeats at telomeres).[15] Although the overall density of microsatellites is relatively uniform in all human chromosomes, the distribution of each repeat class (mono-, di-, tri- etc.) varies: (1) Mononucleotide microsatellites show a uniform distribution among exonic, intronic and intergenic regions (note "exonic regions" herein include not only protein-coding sequences but also 5' and 3' untranslated regions); (2) Di-, tetra- and pentanucleotide microsatellites are relatively depleted in exonic regions (74%, 75% and 87% of the genome average, respectively); (3) Tri- and hexanucleotide microsatellites are enriched in exonic regions (191% and 146% of the genome average, respectively).[15] The vast majority of human microsatellites that are conserved during vertebrate evolution are found in intergenic and intronic regions, consistent with the prediction that microsatellites in exonic regions are more disruptive to gene function.[75]

In comparison, the two major types of non-LTR retrotransposons, L1s and Alus, display contrasting distribution patterns relative to genes. Alus typically are overrepresented inside genes (i.e., residing anywhere in a transcription unit) as well as within 30 kb from annotated genes.[76] In addition, Alus display no orientation bias relative to genes.[76,77] In contrast, L1s show prominent orientation bias within genes: while antisense L1s are close to the predicted density, sense L1s are strongly depleted within genes.[76,77] In addition, both sense and antisense L1s are significantly underrepresented within 5 kb from annotated genes.[76] The differential distribution of Alus and L1s relative to genes is not a result of target preference because both are retrotransposed by L1 proteins. Rather it is consistent with the predicted impact of sense-oriented L1s on premature polyadenylation, cryptic splicing and/or inhibition of transcriptional elongation.[77-80] This idea is further supported by the random chromosomal distribution of recent endogenous L1 and Alu copies in the human genome[24,81-83] and de novo insertions from L1 transgenes in mice.[84,85] Furthermore, the large catalog of polymorphic Alu and L1 insertions clearly shows that exons can be targeted in proportion to their abundance in the genome.[83,86]

Due to their dynamic nature, microsatellite loci are highly polymorphic among individuals. By comparing the reference human genome to a second genome, a survey of variation in microsatellites with at least three perfect repeats was conducted. Overall, 2.7% of the microsatellite loci are polymorphic, a frequency that is 45-fold higher than single nucleotide polymorphism (0.06%).[14] Exonic, especially the coding exonic sequences, show the least amount of microsatellite variation as compared with intergenic and intronic sequences, reflecting increased selective constraints on exonic sequences.[14] Polymorphism incidence increases with repeat number, an observation that is reproduced by surveying polymorphic microsatellite loci among individual genomes from the 1000-genome project.[14,36] There is remarkable heterogeneity in the levels of variation at specific microsatellite loci. For example, the trinucleotide repeat in the 5'UTR of the fragile X mental retardation 1 (FMR1) gene is one of the most studied microsatellite loci with respect to polymorphism in ethnically diverse, healthy individuals.[87] The average number of repeat variants is approximately 9.4 per 100 individuals surveyed. The range of repeat number varies from population to population; the overall range is 6 to 87 in surveyed populations.[87] A repeat number of > 200 is considered a full mutation, which is typically associated with intellectual and developmental disabilities.

Whole-genome as well as targeted sequencing has also started to reveal the extent of genetic variations caused by non-LTR retrotransposons. Polymorphisms in non-LTR retrotransposons are manifested as indels (i.e., insertions in one genome and deletions in another genome), a common form of structural variations (other forms include tandem duplications, inversions and translocations). Recent whole-genome sequencing efforts have cataloged a large number of structural variations among diverse human populations. Detailed analyses of mutational mechanisms indicate 20–30% of SVs are caused by non-LTR retrotransposons.[88-91] On average, each human genome contains 1000–2000 polymorphic non-LTR retrotransposons, with 79–85% Alu, 12–17% L1s and 3% SVAs.[82,88,92,93] Non-LTR retrotransposons are actively modifying individual genomes. Based on the number of polymorphic insertions and the estimated genetic distance between individual genomes, the frequency of retrotransposition has been calculated as one Alu insertion in 21 births and one L1 insertion in 108 to 212 births.[88,94,95] Most intriguingly, an analysis of the 1000 Genomes Project pilot data set suggests a recent increase in retrotransposition rates in humans.[82] These insertions continuously generate new germline polymorphisms that can be passed across generations (**Fig. 5**). As compared with humans, TEs play an even bigger role in generating structural variations in the mouse genome.[96-98] This discrepancy can be largely explained by the presence of a large number of active LTR retrotransposons in the mouse genome besides non-LTR retrotransposons. Together, retrotransposon insertions contribute to ¾ of structural variations between two inbred mouse lines, with 52%, 43% and 5% contribution from LTR retrotransposons, L1s and SINEs, respectively.[96] In addition to retrotransposition in the germline, non-LTR retrotransposons are active in somatic tissues, including normal brain tissues and most epithelial cancers (further discussed in the next section). Despite the significant progress made thus far, it should be noted that the discovery of structural variations mediated by TEs remains a challenge because of the technical difficulty in de novo assembling repetitive sequences from short sequence reads as well as the high sequence depth that is required for detecting low-frequency alleles.

Functional Impacts on Genomes and Epigenomes

The impact of a microsatellite or a transposable element and its associated variability is dependent on its genomic location. The most direct phenotypic impact of microsatellites and retrotransposons is manifested as disease causing mutations. Microsatellite instability has been implicated in a variety of human diseases with over 40 neurological, neurodegenerative and neuromuscular disorders associated with trinucleotide repeat instability.[99,100] Well known examples include Huntington's disease, Fragile X

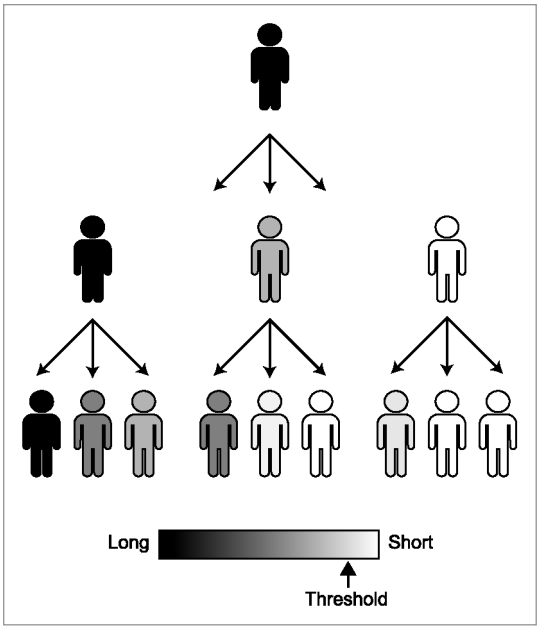

Figure 5. Polymorphism and population dynamics of poly(A) microsatellites. The schematic illustrates the mutation dynamics of poly(A) microsatellites across generations, beginning with an individual who has a newly acquired long poly(A) microsatellite from L1 or Alu retrotransposition. Color is used to distinguish the size and mutability of the microsatellite, where darker color indicates longer poly(A) tails that are more mutable. A long poly(A) microsatellite can expand or contract between generations and give rise to variations in the offspring. However, when a poly(A) microsatellite mutates below the threshold repeat number, the rate of mutation decreases to the average genomic mutation rate.

syndrome and Friedreich's Ataxia. Unlike static mutations, which can be stably passed from one generation to the other, alleles with expanded microsatellites may continue to expand over the course of the individual's life or through germline transmission.[100] Depending on the genomic neighborhood, variation in microsatellite length can induce phenotypic changes through multiple pathways. These mechanisms have been best characterized in trinucleotide repeat expansion disorders and typically consist of one (or the combination) of three following three pathways: the loss of function at the protein level, the gain of function at the protein level, or the gain of function at the RNA level.[99] The latter is gaining increased attention due to the recognized role of non-coding RNA in epigenetic pathways.[101] A noticeable absence among these molecular mechanisms is a loss of function at the RNA level, a potential mechanism that should be investigated in future studies.

Likewise, de novo retrotransposition has been implicated in a variety of sporadic human diseases, including hemophilia, Apert syndrome, neurofibromatosis type 1 and congenital muscular dystrophy.[102,103] A multitude of molecular mechanisms have been proposed to account for the mutagenesis effect for intronic or exonic insertions.[103] A less emphasized impact is post-insertional genome instability mediated by non-LTR retrotransposons, which likely poses even greater hazard to the human genome.[104,105] For example, nonallelic homologous recombination (NAHR) between two separate copies of Alu repeats may generate structural variations, such as indel, inversion and translocation mutations.[104,105] In addition, a hallmark of non-LTR retrotransposition is the formation of a pair of target site duplications (TSDs) that flank the new insertion. These TSDs represent pre-insertion L1 endonuclease cleavage sites and may be targeted for subsequent cleavage by other L1 endonucleases. It is proposed that new double strand breaks in TSDs may trigger rearrangement events due to their adjacency to non-LTR retrotransposons with abundant homologous copies.[105] These mechanisms highlight the dynamic nature of L1 and Alu sequences in the human genome. Furthermore, as has been discussed earlier, non-LTR retrotransposons are not static entities after insertion; they are a significant source for microsatellite genesis.

Both microsatellites and non-LTR retrotransposons have been associated with tumorigenesis and cancer. High-frequency microsatellite instability (MSI-H) is a molecular feature of tumors associated with hereditary nonpolyposis colorectal cancers (HNPCC; Lynch syndrome) and has been seen in other sporadic cancers, usually of the gastrointestinal system.[106,107] These tumors generally arise from defects in the mismatch repair system, leaving DNA vulnerable to a high degree of mutation within microsatellites.[106-108] Detection of MSI-H tumors is often achieved with a reference panel of five microsatellites, including two mononucleotide and three dinucleotide markers.[109] The two mononucleotide markers, BAT26 and BAT25, contain $(A)_{26}$ and $(A)_{25}$ repeats, respectively; their length changes can be easily discerned even in the absence of normal tissue.[109] It is thought that MSI-H contributes to tumorigenesis mainly through mutating genes with coding microsatellites.[107] The role of microsatellite instability in non-coding regions remains to be characterized. A recent genome-wide survey yields the initial glimpse of microsatellite mutational landscapes in MSI-H gastric cancers.[110] There are four to 6-fold higher microsatellite mutations in genes in MSI-H gastric cancers than microsatellite stable (MSS) cancers.[110] Only a very small fraction of microsatellite mutations (0.1%) are located in protein-coding exonic regions; the remaining are found in 5'UTR, 3'UTR, intronic and intergenic regions of the genome.[110] Importantly, genes with significant changes in mRNA expression between MSI-H and MSS samples are enriched for microsatellite mutations in their UTRs, indicating a role of non-coding microsatellite instability in regulating gene expression.[110]

A role of non-LTR retrotransposons in tumorigenesis has gained increasing attention in recent years.[111-113] Several studies have used NextGen sequencing approaches to detect genome-wide structural variations or specifically new retrotransposon insertions in tumor samples.[83,93,114-116] The first study identified 9 tumor-specific (i.e., somatic) L1 insertions in 6 of the 20 lung tumors but none in 10 brain tumors by targeted sequencing.[83] A subsequent study characterized retrotransposon insertions in 43 human tumors of different origins by whole-genome sequencing.[93] Tumor-specific somatic insertions were found in all tumors of epithelial origin (colorectal, prostate and ovarian), but not in blood or brain cancers.[93] Among epithelial cancers, colorectal cancers had the highest incidence of L1 insertions at an average of 9 insertions per tumor among 4 colorectal samples; the fifth colorectal tumor sample alone had an astonishingly high

number of 102 L1 insertions.[93] Interestingly, this tumor was the only colorectal sample that manifested MSI-H with altered DNA repair pathways, suggesting a common origin to both sources of genomic instability. Microsatellite instability may be accompanied by increased retrotransposon activity. An independent study mapped tumor-specific L1 insertions in 16 colorectal tumors by targeted L1 sequencing.[114] An average of 4.4 somatic L1 insertions was found. Intriguingly, although two samples with the highest number of somatic insertions showed microsatellite instability, no correlation was found between the number of insertions and the microsatellite instability status.[114] Future studies with higher coverage and/or larger sample size may be needed to establish a connection between microsatellite instability and retrotransposon mutagenesis. It should be emphasized that, in addition to somatic insertions, germline insertions may also be an important etiological factor in predisposing the carriers to cancers, as has been recently shown for hepatocellular carcinoma patients.[116] The precise role of retrotransposon insertions in cancer remains to be determined but they can have a significant impact on gene function. Consistent with sense-oriented L1 insertions being more disruptive to gene expression, genes targeted by sense insertions, but not those targeted by antisense insertions, showed statistically significant decrease in transcription.[93]

Although usually highlighted for their negative roles, microsatellites and retrotransposons have evolutionary importance due to their ability to provide genetic novelties in human populations.[8,9,117] A salient example is sequence shuffling through 3' transduction.[73,118,119] Due to the weak nature of the L1 poly(A) signal, it is possible for L1 to bypass the signal during transcription and use an alternative poly(A) signal downstream.[73] This results in the mobilization of the 3' non-L1 sequence. Surveys of genomic L1 copies indicate up to 23% of them underwent 3' transduction.[118,119] Some 3' transductions may remain undetected due to so-called orphan 3' transduction.[120] The latter involves a 3' transduction event that is severely 5' truncated, which leaves behind no L1 sequence to be associated with the mobilized sequence. Such an event was responsible for Duchene muscular dystrophy in a Japanese patient.[120] Thus, 3' transduction has the potential to generate new genes and bring new regulatory features to the vicinity of other genes. Although no cases have been reported, it is conceivable that L1s could also mobilize microsatellites by the same mechanism. Another way of introducing evolutionary novelties is through trans-mobilization of cellular RNAs, which form processed pseudogenes.[121] There are approximately 10,000 copies of processed pseudogenes in the human genome.[122,123] Some of them have acquired novel functions and are thus classified as retrogenes.[124] Lastly, since the exciting discovery of somatic retrotransposition in the brain, it has been postulated that L1-induced neuronal diversity may be an important source for behavioral phenotypes and confer evolutionary advantage.[125-129]

Beyond genome sequence changes, the importance of other epigenetic mechanism, including DNA methylation, histone modification and nucleosome occupancy, is coming to light. Retrotransposons are known to be associated with epigenetic changes.[130] Therefore, the insertion of an L1 into a genomic region may alter the previously existing epigenetic state ([128,131] and our unpublished data). Due to their unique sequence composition, microsatellites can change the physical forms of DNA where they occur,[12,43] which can have implications for gene expression and genome stability. Though not all microsatellites meet the criteria to do so, many repeats form non-B DNA structures. These non-B forms of DNA are often associated with locations of chromosomal breakage.[132] An inverse correlation exists between a repeat's ability to form non-B form DNA and its abundance in the genome, suggesting that these unstable structures are selected against due to reduced evolutionary fitness. The presence of poly(A) tracts within promoter sequences also has implications for nucleosome organization.[133] Poly(A) repeats form shorter and more rigid helix structures and have a narrower minor groove, altering nucleosome positioning in manners which increase the binding ability of transcription factors.[134,135] Thus, the mobilization of retroelements and their associated poly(A) tails present a direct manner for L1, Alu and SVA elements to affect the genome compaction and expression dynamics of the local regions where they land.[39]

In summary, microsatellites and non-LTR retrotransposons are dynamic DNA sequences that rapidly change from generation to generation. This makes them independent drivers of genomic variation and evolution. Expansion and mutation of microsatellites within L1 and Alu sequences may change the size and sequence identity of these non-LTR retrotransposons, perhaps modulating their mobility. Future studies to compare the sequence of full length L1s may reveal how the dynamics of microsatellites affect the retrotransposons from which they are born.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors wish to thank James Rosser and Simon Newkirk as well as an anonymous reviewer for critical reading and helpful suggestions in the composition of this review.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11237011; http://dx.doi.org/10.1038/35057062
2. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 2011; 7:e1002384; PMID:22144907; http://dx.doi.org/10.1371/journal.pgen.1002384
3. Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 2003; 13:651-8; PMID:14638329; http://dx.doi.org/10.1016/j.gde.2003.10.013
4. Singer MF. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 1982; 28:433-4; PMID:6280868; http://dx.doi.org/10.1016/0092-8674(82)90194-5
5. Weiner AM. SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol 2002; 14:343-50; PMID:12067657; http://dx.doi.org/10.1016/S0955-0674(02)00338-1
6. Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. Heredity (Edinb) 2011; 107:487-95; PMID:21673742; http://dx.doi.org/10.1038/hdy.2011.43
7. Okada N. SINEs. Curr Opin Genet Dev 1991; 1:498-504; PMID:1668310; http://dx.doi.org/10.1016/S0959-437X(05)80198-4

8. Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science 2004; 303:1626-32; PMID:15016989; http://dx.doi.org/10.1126/science.1089670
9. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet 2009; 10:691-703; PMID:19763152; http://dx.doi.org/10.1038/nrg2640
10. Ellegren H. Microsatellites: simple sequences with complex evolution. Nat Rev Genet 2004; 5:435-45; PMID:15153996; http://dx.doi.org/10.1038/nrg1348
11. Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 2002; 11:2453-65; PMID:12453231; http://dx.doi.org/10.1046/j.1365-294X.2002.01643.x
12. Eckert KA, Hile SE. Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. Mol Carcinog 2009; 48:379-88; PMID:19306292; http://dx.doi.org/10.1002/mc.20499
13. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev 2008; 72:686-727; PMID:19052325; http://dx.doi.org/10.1128/MMBR.00011-08
14. Payseur BA, Jing P, Haasl RJ. A genomic portrait of human microsatellite variation. Mol Biol Evol 2011; 28:303-12; PMID:20675409; http://dx.doi.org/10.1093/molbev/msq198
15. Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol 2003; 4:R13; PMID:12620123; http://dx.doi.org/10.1186/gb-2003-4-2-r13
16. Buschiazzo E, Gemmell NJ. The rise, fall and renaissance of microsatellites in eukaryotic genomes. Bioessays 2006; 28:1040-50; PMID:16998838; http://dx.doi.org/10.1002/bies.20470
17. Kelkar YD, Eckert KA, Chiaromonte F, Makova KD. A matter of life or death: how microsatellites emerge in and vanish from the human genome. Genome Res 2011; 21:2038-48; PMID:21994250; http://dx.doi.org/10.1101/gr.122937.111
18. Economou EP, Bergen AW, Warren AC, Antonarakis SE. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. Proc Natl Acad Sci U S A 1990; 87:2951-4; PMID:2326257; http://dx.doi.org/10.1073/pnas.87.8.2951
19. Zuliani G, Hobbs HH. A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. Am J Hum Genet 1990; 46:963-9; PMID:2339694
20. Beckman JS, Weber JL. Survey of human and rat microsatellites. Genomics 1992; 12:627-31; PMID:1572635; http://dx.doi.org/10.1016/0888-7543(92)90285-Z
21. Nadir E, Margalit H, Gallily T, Ben-Sasson SA. Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc Natl Acad Sci U S A 1996; 93:6470-5; PMID:8692839; http://dx.doi.org/10.1073/pnas.93.13.6470
22. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA. Alu repeats: a source for the genesis of primate microsatellites. Genomics 1995; 29:136-44; PMID:8530063; http://dx.doi.org/10.1006/geno.1995.1224
23. Jurka J, Pethiyagoda C. Simple repetitive DNA sequences from primates: compilation and analysis. J Mol Evol 1995; 40:120-6; PMID:7699718; http://dx.doi.org/10.1007/BF00167107
24. Ovchinnikov I, Troxel AB, Swergold GD. Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. Genome Res 2001; 11:2050-8; PMID:11731495; http://dx.doi.org/10.1101/gr.194701
25. Boeke JD. LINEs and Alus--the polyA connection. Nat Genet 1997; 16:6-7; PMID:9140383; http://dx.doi.org/10.1038/ng0597-6
26. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al.; Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420:520-62; PMID:12466850; http://dx.doi.org/10.1038/nature01262
27. Chen JM, Stenson PD, Cooper DN, Férec C. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet 2005; 117:411-27; PMID:15983781; http://dx.doi.org/10.1007/s00439-005-1321-0
28. Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 2000; 10:967-81; PMID:10899146; http://dx.doi.org/10.1101/gr.10.7.967
29. Jarne P, David P, Viard F. Microsatellites, transposable elements and the X chromosome. Mol Biol Evol 1998; 15:28-34; PMID:9491602; http://dx.doi.org/10.1093/oxfordjournals.molbev.a025844
30. Wilder J, Hollocher H. Mobile elements and the genesis of microsatellites in dipterans. Mol Biol Evol 2001; 18:384-92; PMID:11230539; http://dx.doi.org/10.1093/oxfordjournals.molbev.a003814
31. Clark RM, Dalgliesh GL, Endres D, Gomez M, Taylor J, Bidichandani SI. Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. Genomics 2004; 83:373-83; PMID:14962663; http://dx.doi.org/10.1016/j.ygeno.2003.09.001
32. Roy AM, Carroll ML, Nguyen SV, Salem AH, Oldridge M, Wilkie AO, et al. Potential gene conversion and source genes for recently integrated Alu elements. Genome Res 2000; 10:1485-95; PMID:11042148; http://dx.doi.org/10.1101/gr.152300
33. Kelkar YD, Strubczewski N, Hile SE, Chiaromonte F, Eckert KA, Makova KD. What is a microsatellite: a computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. Genome Biol Evol 2010; 2:620-35; PMID:20668018; http://dx.doi.org/10.1093/gbe/evq046
34. Webster MT, Smith NG, Ellegren H. Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. Proc Natl Acad Sci U S A 2002; 99:8748-53; PMID:12070344; http://dx.doi.org/10.1073/pnas.122067599
35. Lai Y, Sun F. The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol Biol Evol 2003; 20:2123-31; PMID:12949124; http://dx.doi.org/10.1093/molbev/msg228
36. Ananda G, Walsh E, Jacob KD, Krasilnikova M, Eckert KA, Chiaromonte F, et al. Distinct mutational behaviors differentiate short tandem repeats from microsatellites in the human genome. Genome Biol Evol 2013; 5:606-20; PMID:23241442; http://dx.doi.org/10.1093/gbe/evs116
37. Dechering KJ, Cuelenaere K, Konings RN, Leunissen JA. Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. Nucleic Acids Res 1998; 26:4056-62; PMID:9705519; http://dx.doi.org/10.1093/nar/26.17.4056
38. Bacon AL, Dunlop MG, Farrington SM. Hypermutability at a poly(A/T) tract in the human germline. Nucleic Acids Res 2001; 29:4405-13; PMID:11691928; http://dx.doi.org/10.1093/nar/29.21.4405
39. Grandi FC, Rosser JM, An W. LINE-1-derived poly(A) microsatellites undergo rapid shortening and create somatic and germline mosaicism in mice. Mol Biol Evol 2013; 30:503-12; PMID:23125228; http://dx.doi.org/10.1093/molbev/mss251
40. Twerdi CD, Boyer JC, Farber RA. Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. Proc Natl Acad Sci U S A 1999; 96:2875-9; PMID:10077604; http://dx.doi.org/10.1073/pnas.96.6.2875
41. Hanford MG, Rushton BC, Gowen LC, Farber RA. Microsatellite mutation rates in cancer cell lines deficient or proficient in mismatch repair. Oncogene 1998; 16:2389-93; PMID:9620556; http://dx.doi.org/10.1038/sj.onc.1201751
42. Yamada NA, Smith GA, Castro A, Roques CN, Boyer JC, Farber RA. Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. Mutat Res 2002; 499:213-25; PMID:11827714; http://dx.doi.org/10.1016/S0027-5107(01)00282-2
43. Baptiste BA, Ananda G, Strubczewski N, Lutzkanin A, Khoo SJ, Srikanth A, et al. Mature microsatellites: mechanisms underlying dinucleotide microsatellite mutational biases in human cells. G3 (Bethesda) 2013; 3:451-63; PMID:23450065; http://dx.doi.org/10.1534/g3.112.005173
44. Weber JL, Wong C. Mutation of human short tandem repeats. Hum Mol Genet 1993; 2:1123-8; PMID:8401493; http://dx.doi.org/10.1093/hmg/2.8.1123
45. Pelletier R, Farrell BT, Miret JJ, Lahue RS. Mechanistic features of CAG*CTG repeat contractions in cultured cells revealed by a novel genetic assay. Nucleic Acids Res 2005; 33:5667-76; PMID:16199754; http://dx.doi.org/10.1093/nar/gki880
46. Lee JS, Hanford MG, Genova JL, Farber RA. Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. Hum Mol Genet 1999; 8:2567-72; PMID:10556306; http://dx.doi.org/10.1093/hmg/8.13.2567
47. Holtkemper U, Rolf B, Hohoff C, Forster P, Brinkmann B. Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. Hum Mol Genet 2001; 10:629-33; PMID:11230182; http://dx.doi.org/10.1093/hmg/10.6.629
48. Parsons R, Myeroff LL, Liu B, Willson JK, Markowitz SD, Kinzler KW, et al. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. Cancer Res 1995; 55:5548-50; PMID:7585632
49. Ovchinnikov I, Rubin A, Swergold GD. Tracing the LINEs of human evolution. Proc Natl Acad Sci U S A 2002; 99:10522-7; PMID:12138175; http://dx.doi.org/10.1073/pnas.152346799
50. Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, et al. Active Alu element "A-tails": size does matter. Genome Res 2002; 12:1333-44; PMID:12213770; http://dx.doi.org/10.1101/gr.384802
51. Varela MA, Sanmiguel R, Gonzalez-Tizon A, Martinez-Lage A. Heterogeneous nature and distribution of interruptions in dinucleotides may indicate the existence of biased substitutions underlying microsatellite evolution. J Mol Evol 2008; 66:575-80; PMID:18496726; http://dx.doi.org/10.1007/s00239-008-9107-3
52. Hile SE, Eckert KA. DNA polymerase kappa produces interrupted mutations and displays polar pausing within mononucleotide microsatellite sequences. Nucleic Acids Res 2008; 36:688-96; PMID:18079151; http://dx.doi.org/10.1093/nar/gkm1089
53. Zingler N, Weichenrieder O, Schumann GG. APE-type non-LTR retrotransposons: determinants involved in target site recognition. Cytogenet Genome Res 2005; 110:250-68; PMID:16093679; http://dx.doi.org/10.1159/000084959
54. Kojima KK, Fujiwara H. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol 2003; 20:351-61; PMID:12644555; http://dx.doi.org/10.1093/molbev/msg031
55. Kojima KK, Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol 2004; 21:207-17; PMID:12949131; http://dx.doi.org/10.1093/molbev/msg235

56. Xiong Y, Eickbush TH. Dong, a non-long terminal repeat (non-LTR) retrotransposable element from Bombyx mori. Nucleic Acids Res 1993; 21:1318; PMID:8385316; http://dx.doi.org/10.1093/nar/21.5.1318
57. Busseau I, Berezikov E, Bucheton A. Identification of Waldo-A and Waldo-B, two closely related non-LTR retrotransposons in Drosophila. Mol Biol Evol 2001; 18:196-205; PMID:11158378; http://dx.doi.org/10.1093/oxfordjournals.molbev.a003793
58. Blumenstiel JP, Hartl DL, Lozovsky ER. Patterns of insertion and deletion in contrasting chromatin domains. Mol Biol Evol 2002; 19:2211-25; PMID:12446812; http://dx.doi.org/10.1093/oxfordjournals.molbev.a004045
59. Coates BS, Sumerford DV, Hellmich RL, Lewis LC. A helitron-like transposon superfamily from lepidoptera disrupts (GAAA)(n) microsatellites and is responsible for flanking sequence similarity within a microsatellite family. J Mol Evol 2010; 70:275-88; PMID:20217059; http://dx.doi.org/10.1007/s00239-010-9330-6
60. Akagi H, Yokozeki Y, Inagaki A, Mori K, Fujimura T. Micron, a microsatellite-targeting transposable element in the rice genome. Mol Genet Genomics 2001; 266:471-80; PMID:11713677; http://dx.doi.org/10.1007/s004380100563
61. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001; 11:1441-52; PMID:11483586; http://dx.doi.org/10.1101/gr.184001
62. Ichiyanagi K, Nishihara H, Duvernell DD, Okada N. Acquisition of endonuclease specificity during evolution of L1 retrotransposon. Mol Biol Evol 2007; 24:2009-15; PMID:17602167; http://dx.doi.org/10.1093/molbev/msm130
63. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A 1997; 94:1872-7; PMID:9050872; http://dx.doi.org/10.1073/pnas.94.5.1872
64. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. EMBO J 2002; 21:5899-910; PMID:12411507; http://dx.doi.org/10.1093/emboj/cdf592
65. Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol 2001; 21:467-75; PMID:11134335; http://dx.doi.org/10.1128/MCB.21.2.467-475.2001
66. Monot C, Kuciak M, Viollet S, Mir AA, Gabus C, Darlix JL, et al. The specificity and flexibility of l1 reverse transcription priming at imperfect T-tracts. PLoS Genet 2013; 9:e1003499; PMID:23675310; http://dx.doi.org/10.1371/journal.pgen.1003499
67. Dai L, Taylor MS, O'Donnell KA, Boeke JD. Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. Mol Cell Biol 2012; 32:4323-36; PMID:22907758; http://dx.doi.org/10.1128/MCB.06785-11
68. West N, Roy-Engel AM, Imataka H, Sonenberg N, Deininger P. Shared protein components of SINE RNPs. J Mol Biol 2002; 321:423-32; PMID:12162956; http://dx.doi.org/10.1016/S0022-2836(02)00542-9
69. Xie Y, Mates L, Ivics Z, Izsvák Z, Martin SL, An W. Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. Mob DNA 2013; 4:10; PMID:23497436; http://dx.doi.org/10.1186/1759-8753-4-10
70. Dewannieux M, Heidmann T. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. J Mol Biol 2005; 349:241-7; PMID:15890192; http://dx.doi.org/10.1016/j.jmb.2005.03.068
71. Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? Genome Res 2009; 19:545-55; PMID:19273617; http://dx.doi.org/10.1101/gr.089789.108
72. Wagstaff BJ, Hedges DJ, Derbes RS, Campos Sanchez R, Chiaromonte F, Makova KD, et al. Rescuing Alu: recovery of new inserts shows LINE-1 preserves Alu activity through A-tail expansion. PLoS Genet 2012; 8:e1002842; PMID:22912586; http://dx.doi.org/10.1371/journal.pgen.1002842
73. Moran JV, DeBerardinis RJ, Kazazian HH Jr. Exon shuffling by L1 retrotransposition. Science 1999; 283:1530-4; PMID:10066175; http://dx.doi.org/10.1126/science.283.5407.1530
74. Belancio VP, Whelton M, Deininger P. Requirements for polyadenylation at the 3' end of LINE-1 elements. Gene 2007; 390:98-107; PMID:17023124; http://dx.doi.org/10.1016/j.gene.2006.07.029
75. Buschiazzo E, Gemmell NJ. Conservation of human microsatellites across 450 million years of evolution. Genome Biol Evol 2010; 2:153-65; PMID:20333231; http://dx.doi.org/10.1093/gbe/evq007
76. Medstrand P, van de Lagemaat LN, Mager DL. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res 2002; 12:1483-95; PMID:12368240; http://dx.doi.org/10.1101/gr.388902
77. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999; 9:657-63; PMID:10607616; http://dx.doi.org/10.1016/S0959-437X(99)00031-3
78. Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet 2003; 35:363-6; PMID:14625551; http://dx.doi.org/10.1038/ng1269
79. Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic Acids Res 2006; 34:1512-21; PMID:16554555; http://dx.doi.org/10.1093/nar/gkl027
80. Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature 2004; 429:268-74; PMID:15152245; http://dx.doi.org/10.1038/nature02536
81. Cordaux R, Lee J, Dinoso L, Batzer MA. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene 2006; 373:138-44; PMID:16527433; http://dx.doi.org/10.1016/j.gene.2006.01.020
82. Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, et al.; 1000 Genomes Project. A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet 2011; 7:e1002236; PMID:21876680; http://dx.doi.org/10.1371/journal.pgen.1002236
83. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, et al. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 2010; 141:1253-61; PMID:20603005; http://dx.doi.org/10.1016/j.cell.2010.05.020
84. An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, et al. Active retrotransposition by a synthetic L1 element in mice. Proc Natl Acad Sci U S A 2006; 103:18662-7; PMID:17124176; http://dx.doi.org/10.1073/pnas.0605300103
85. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr. L1 integration in a transgenic mouse model. Genome Res 2006; 16:240-50; PMID:16365384; http://dx.doi.org/10.1101/gr.4571606
86. Witherspoon DJ, Zhang Y, Xing J, Watkins WS, Ha H, Batzer MA, et al. Mobile element scanning (ME-Scan) identifies thousands of novel Alu insertions in diverse human populations. Genome Res 2013; 23:1170-81; PMID:23599355; http://dx.doi.org/10.1101/gr.148973.112
87. Peprah E. Fragile X syndrome: the FMR1 CGG repeat distribution among world populations. Ann Hum Genet 2012; 76:178-91; PMID:22188182; http://dx.doi.org/10.1111/j.1469-1809.2011.00694.x
88. 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; 19:1516-26; PMID:19439515; http://dx.doi.org/10.1101/gr.091827.109
89. Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, et al. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. Cell 2010; 143:837-47; PMID:21111241; http://dx.doi.org/10.1016/j.cell.2010.10.027
90. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science 2007; 318:420-6; PMID:17901297; http://dx.doi.org/10.1126/science.1149504
91. Lam HY, Mu XJ, Stütz AM, Tanzer A, Cayting PD, Snyder M, et al. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. Nat Biotechnol 2010; 28:47-55; PMID:20037582; http://dx.doi.org/10.1038/nbt.1600
92. Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. Natural genetic variation caused by transposable elements in humans. Genetics 2004; 168:933-51; PMID:15514065; http://dx.doi.org/10.1534/genetics.104.031757
93. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, et al.; Cancer Genome Atlas Research Network. Landscape of somatic retrotransposition in human cancers. Science 2012; 337:967-71; PMID:22745252; http://dx.doi.org/10.1126/science.1222077
94. Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res 2010; 20:1262-70; PMID:20488934; http://dx.doi.org/10.1101/gr.106419.110
95. Huang CR, Schneider AM, Lu Y, Niranjan T, Shen P, Robinson MA, et al. Mobile interspersed repeats are major structural variants in the human genome. Cell 2010; 141:1171-82; PMID:20602999; http://dx.doi.org/10.1016/j.cell.2010.05.026
96. Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, et al. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res 2010; 20:623-35; PMID:20308636; http://dx.doi.org/10.1101/gr.102970.109
97. Akagi K, Li J, Stephens RM, Volfovsky N, Symer DE. Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. Genome Res 2008; 18:869-80; PMID:18381897; http://dx.doi.org/10.1101/gr.075770.107
98. Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL. Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. PLoS Genet 2008; 4:e1000007; PMID:18454193; http://dx.doi.org/10.1371/journal.pgen.1000007
99. Gatchel JR, Zoghbi HY. Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet 2005; 6:743-55; PMID:16205714; http://dx.doi.org/10.1038/nrg1691
100. Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 2005; 6:729-42; PMID:16205713; http://dx.doi.org/10.1038/nrg1689
101. Renoux AJ, Todd PK. Neurodegeneration the RNA way. Prog Neurobiol 2012; 97:173-89; PMID:22079416; http://dx.doi.org/10.1016/j.pneurobio.2011.10.006
102. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. Curr Opin Genet Dev 2012; 22:191-203; PMID:22406018; http://dx.doi.org/10.1016/j.gde.2012.02.006

103. Kaer K, Speek M. Retroelements in human disease. Gene 2013; 518:231-41; PMID:23333607; http://dx.doi.org/10.1016/j.gene.2013.01.008
104. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab 1999; 67:183-93; PMID:10381326; http://dx.doi.org/10.1006/mgme.1999.2864
105. Hedges DJ, Deininger PL. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat Res 2007; 616:46-59; PMID:17157332; http://dx.doi.org/10.1016/j.mrfmmm.2006.11.021
106. Imai K, Yamamoto H. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. Carcinogenesis 2008; 29:673-80; PMID:17942460; http://dx.doi.org/10.1093/carcin/bgm228
107. Woerner SM, Kloor M, von Knebel Doeberitz M, Gebert JF. Microsatellite instability in the development of DNA mismatch repair deficient tumors. Cancer Biomark 2006; 2:69-86; PMID:17192061
108. Preston BD, Albertson TM, Herr AJ. DNA replication fidelity and cancer. Semin Cancer Biol 2010; 20:281-93; PMID:20951805; http://dx.doi.org/10.1016/j.semcancer.2010.10.009
109. Laghi L, Bianchi P, Malesci A. Differences and evolution of the methods for the assessment of microsatellite instability. Oncogene 2008; 27:6313-21; PMID:18679418; http://dx.doi.org/10.1038/onc.2008.217
110. Yoon K, Lee S, Han TS, Moon SY, Yun SM, Kong SH, et al. Comprehensive genome- and transcriptome-wide analyses of mutations associated with microsatellite instability in Korean gastric cancers. Genome Res 2013; 23:1109-17; PMID:23737375; http://dx.doi.org/10.1101/gr.145706.112
111. Rodić N, Burns KH. Long interspersed element-1 (LINE-1): passenger or driver in human neoplasms? PLoS Genet 2013; 9:e1003402; PMID:23555307; http://dx.doi.org/10.1371/journal.pgen.1003402
112. Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know 'bout retroelements in cancer. Semin Cancer Biol 2010; 20:200-10; PMID:20600922; http://dx.doi.org/10.1016/j.semcancer.2010.06.001
113. Wilkins AS. The enemy within: an epigenetic role of retrotransposons in cancer initiation. Bioessays 2010; 32:856-65; PMID:20715060; http://dx.doi.org/10.1002/bies.201000008
114. Solyom S, Ewing AD, Rahrmann EP, Doucet T, Nelson HH, Burns MB, et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res 2012; 22:2328-38; PMID:22968929; http://dx.doi.org/10.1101/gr.145235.112
115. Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. Cell 2013; 153:919-29; PMID:23663786; http://dx.doi.org/10.1016/j.cell.2013.04.010
116. Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 2013; 153:101-11; PMID:23540693; http://dx.doi.org/10.1016/j.cell.2013.02.032
117. Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 2004; 21:991-1007; PMID:14963101; http://dx.doi.org/10.1093/molbev/msh073
118. Goodier JL, Ostertag EM, Kazazian HH Jr. Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 2000; 9:653-7; PMID:10699189; http://dx.doi.org/10.1093/hmg/9.4.653
119. Pickeral OK, Makałowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 2000; 10:411-5; PMID:10779482; http://dx.doi.org/10.1101/gr.10.4.411
120. Solyom S, Ewing AD, Hancks DC, Takeshima Y, Awano H, Matsuo M, et al. Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. Hum Mutat 2012; 33:369-71; PMID:22095564; http://dx.doi.org/10.1002/humu.21663
121. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 2000; 24:363-7; PMID:10742098; http://dx.doi.org/10.1038/74184
122. Torrents D, Suyama M, Zdobnov E, Bork P. A genome-wide survey of human pseudogenes. Genome Res 2003; 13:2559-67; PMID:14656963; http://dx.doi.org/10.1101/gr.1455503
123. Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res 2003; 13:2541-58; PMID:14656962; http://dx.doi.org/10.1101/gr.1429003
124. Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. Nat Rev Genet 2003; 4:865-75; PMID:14634634; http://dx.doi.org/10.1038/nrg1204
125. Singer T, McConnell MJ, Marchetto MC, Coufal NG, Gage FH. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci 2010; 33:345-54; PMID:20471112; http://dx.doi.org/10.1016/j.tins.2010.04.001
126. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature 2011; 479:534-7; PMID:22037309; http://dx.doi.org/10.1038/nature10531
127. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell 2012; 151:483-96; PMID:23101622; http://dx.doi.org/10.1016/j.cell.2012.09.035
128. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 2005; 435:903-10; PMID:15959507; http://dx.doi.org/10.1038/nature03663
129. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. Nature 2009; 460:1127-31; PMID:19657334; http://dx.doi.org/10.1038/nature08248
130. Rosser JM, An W. L1 expression and regulation in humans and rodents. [Elite Ed]. Front Biosci (Elite Ed) 2012; 4:2203-25; PMID:22202032
131. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, et al. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature 2010; 466:769-73; PMID:20686575; http://dx.doi.org/10.1038/nature09209
132. Zhao J, Bacolla A, Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci 2010; 67:43-62; PMID:19727556; http://dx.doi.org/10.1007/s00018-009-0131-2
133. Segal E, Widom J. Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol 2009; 19:65-71; PMID:19208466; http://dx.doi.org/10.1016/j.sbi.2009.01.004
134. Suter B, Schnappauf G, Thoma F. Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. Nucleic Acids Res 2000; 28:4083-9; PMID:11058103; http://dx.doi.org/10.1093/nar/28.21.4083
135. Shimizu M, Mori T, Sakurai T, Shindo H. Destabilization of nucleosomes by an unusual DNA conformation adopted by poly(dA) small middle dotpoly(dT) tracts in vivo. EMBO J 2000; 19:3358-65; PMID:10880448; http://dx.doi.org/10.1093/emboj/19.13.3358