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

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. 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 have substantially diverged from known TE sequences, recent methods capable of annotating repetitive sequences that are sequentially repeated. 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). However, the boundary between microsatellites and minisatellites is often disputed, where the latter typically refer to tandem repeats of greater unit sizes. As they are highly polymorphic, both microsatellites and minisatellites have been instrumental in genetic mapping, evolutionary and phylogenetic studies as well as DNA forensics. This review focuses on microsatellites, which encompass mono-, di-, tri-, tetra-, penta- and hexanucleotide 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. 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.

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.

*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
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.29 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...
genomes indicates that 26% of microsatellite births and 24% of microsatellite deaths occur within Alu and L1 sequences subsequent to retrotransposition. 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. L1 sequences are AT-rich and give rise to microsatellites during and after their insertion along their entire length (Fig. 1A). 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). The middle A-rich region of Alu elements tends to expand and mature into poly(A) microsatellites and can also give rise to GAA repeats. 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.

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.

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.
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

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 | 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) | 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) | per cell per generation | 42 |
| Mouse fibroblast cell line (CAK-Stu3) | 30 | CA | 1.6E-04 (3.8E-05) | 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 |

*Contraction rate; Number in parenthesis is the contraction rate for the same repeat; 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 location 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
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 As 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.12,17,47,52 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).39 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 WaldoSt1 in the earwig F. scudder i);44 (2) AC microsatellites are targeted by a Mino element in A. gambiae (MinoAg);44 (3) TC microsatellites are targeted by four Kibi elements (KibiDr1 and KibiDr2 in zebrafish D. rerio, KibiFrl in the torafugu F. rubripes, and KibiTnl in the green spotted pufferfish T. nigroviridis);50 (4) TTC microsatellites are targeted by two Koshi elements (KoshiFrl in F. rubripes and KoshiTnl in T. nigroviridis);59 (5) TAA microsatellites are targeted by three Dong elements (Dong in the silkworm B. mori, DongAg in A. gambiae and DongBsg in the freshwater snail B. glabrata).50,56 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.35 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,69 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 induce 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...
transcribed by RNA polymerase III and the resulting transcripts are not polyadenylated. 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 associated with the efficiency of integration, an L1 sequence may not have a GT rich region (Fig. 4B). Indeed, newly generated Alu copies typically have longer poly(A) tails than the original donor elements. 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.

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 in 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. 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). 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. 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). 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). 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.

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. In addition, Alus display no orientation bias relative to genes. 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. In addition, both sense and antisense L1s are significantly underrepresented within 5 kb from annotated genes. 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. This idea is further supported by the random chromosomal distribution of recent endogenous L1 and Alu copies in the human genome and de novo insertions from L1 transgenes in mice. 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.

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%). 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. Polymorphism incidence increases with repeat number, an observation that is reproduced by surveying polymorphic microsatellite loci among individual genomes from the 1000-genome project. 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. 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. 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. On average, each human genome contains 1000–2000 polymorphic non-LTR retrotransposons, with 79–85% Alu, 12–17% L1s and 3% SINEs. 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. Most intriguingly, an analysis of the 1000 Genomes Project pilot data set suggests a recent increase in retrotransposition rates in humans. 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. 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. 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. Well known examples include Huntington’s disease, Fragile X syndrome, and 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.
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. 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. The latter is gaining increased attention due to the recognized role of non-coding RNA in epigenetic pathways. 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. A multitude of molecular mechanisms have been proposed to account for the mutagenesis effect for intronic or exonic insertions. A less emphasized impact is post-insertional genome instability mediated by non-LTR retrotransposons, which likely poses even greater hazard to the human genome. For example, nonallelic homologous recombination (NAHR) between two separate copies of Alu repeats may generate structural variations, such as indel, inversion and translocation mutations. 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. 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. These tumors generally arise from defects in the mismatch repair system, leaving DNA vulnerable to a high degree of mutation within microsatellites. Detection of MSI-H tumors is often achieved with a reference panel of five microsatellites, including two mononucleotide and three dinucleotide markers. 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. It is thought that MSI-H contributes to tumorigenesis mainly through mutating genes with coding microsatellites. 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. There are four to 6-fold higher microsatellite mutations in genes in MSI-H gastric cancers than microsatellite stable (MSS) cancers. 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. 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.

A role of non-LTR retrotransposons in tumorigenesis has gained increasing attention in recent years. Studies have used NextGen sequencing approaches to detect genomewide structural variations or specifically new retrotransposon insertions in tumor samples. 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. A subsequent study characterized retrotransposon insertions in 43 human tumors of different origins by whole-genome sequencing. Tumor-specific somatic insertions were found in all tumors of epithelial origin (colorectal, prostate and ovarian), but not in blood or brain cancers. 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. 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. 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. 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, germ line insertions may also be an important etiological factor in predisposing the carriers to cancers, as has been recently shown for hepatocellular carcinoma patients.

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, showed statistically significant decrease in transcription.

Although usually highlighted for their negative roles, microsatellites and retrotransposons have evolutionary importance due to their ability to provide genetic novelties in human populations. A salient example is sequence shuffling through 3’ transduction. 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. 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. Some 3’ transductions may remain undetected due to so-called orphan 3’ truncation. 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. 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. There are approximately 10,000 copies of processed pseudogenes in the human genome.

Some of them have acquired novel functions and are thus classified as retrogenes. 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.

Beyond genome sequence changes, the importance of other epigenetic mechanisms, including DNA methylation, histone modification and nucleosome occupancy, is coming to light. Retrotransposons are known to be associated with epigenetic changes. Therefore, the insertion of an L1 into a genomic region may alter the previously existing epigenetic state and our unpublished data). Due to their unique sequence composition, microsatellites can change the physical forms of DNA where they occur, 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. 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. 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. 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.

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, Barter 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, Barter MA, Kazazian HH Jr. Mobile elements and mammalian genome evolution.Curr Opin Genet Dev 2003; 13:655-8; PMID:14638239; 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)90198-2
5. Weiner AM. SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol 2002; 14:343-50; PMID:12007657; http://dx.doi.org/10.1016/S0955-0674(02)00338-1
6. Kamenov DA, Vasserdy NS. Origin and evolution of SINEs in eukaryotic genomes. Heredity (Edinb) 2001; 1011:487-95; PMID:12173742; http://dx.doi.org/10.1038/hdy.2011.43
7. Okada N. SINEs. Curr Opin Genet Dev 1999; 1:498-504; PMID:1068310; http://dx.doi.org/10.1016/S0959-437X(98)00198-4
8. Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science 2004; 303:1626-32; PMID:15106898; http://dx.doi.org/10.1126/science.1089670

9. Cordaux R, Batzer MA. The impact of retropositioning on human genome evolution. Nat Rev Genet 2009; 10:691-703; PMID:19763512; http://dx.doi.org/10.1038/nrg2560

10. Ellegren H. Microsatellites: simple sequences with complex evolution. Nat Rev Genet 2004; 5:455-65; PMID:15285906; http://dx.doi.org/10.1038/nrg1348

11. Li YC, Korot AL, Fatima T, Beles A, Newo 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.1365-2942.2002.01643.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:19366292; http://dx.doi.org/10.1002/mc.20499

13. Richard GJ, Kerrest A, Dupuy B. Comparative genomic and molecular dynamics of DNA repeats in eukaryotes. Microb Mol Biol Rev 2008; 72:686-727; PMID:19053235; http://dx.doi.org/10.1128/MMBR.00111-08

14. Payseur BA, Jing P, Haal RJ. A genomic portrait of slipped strand mispairing 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. Buschhau E, Gennell NJ, The rise, fall and renaissance of microsatellites in eukaryotic genomes. Biochem Biophys Res Commun 2006; 343:181-5; PMID:16999838; http://dx.doi.org/10.1016/j.bbrc.2006.04.099

17. Kelkar YD, Eckert KA, Chioromonte 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.123971.111

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:23623657; http://dx.doi.org/10.1073/pnas.87.8.2951

19. Zuliani G, Hobbs HH. The high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. Am J Hum Genet 1990; 46:963-9; PMID:41802171

20. Beckman JS, Weber JL. Survey of human and rat microsatellites. Genomics 1992; 12:627-37; PMID:1572653; http://dx.doi.org/10.1016/0888-7543(92)90275-Z

21. Nadir E, Margalit H, Gallyl T, Ben-Sasson SA. Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc Natl Acad Sci USA 1996; 93:6470-5; PMID:7548508; http://dx.doi.org/10.1073/pnas.96.6.2875

22. Jarne P, David P, Viard F. Microsatellites, transposable elements and the X chromosome. Mol Biol Evol 1998; 15:258-34; PMID:9491602; http://dx.doi.org/10.1093/oxfordjournals.molbev.a025844

23. Wilder J, Hollocher H. Mobile elements and the genesis of microsatellites in diploids. Mol Biol Evol 2001; 18:984-92; PMID:11205359; http://dx.doi.org/10.1093/molbev/msq198

24. Nijholt J, Bruckner J, Ovchinnikov I, Troel AB, Swergold GD. 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.119701

25. Bocke JD. LINEs and Alus—polynucleotide connections. Nat Genet 1997; 16:6-7; PMID:9140838; http://dx.doi.org/10.1038/385097a0
Xiong Y, Eickbush TH, Dong, a non-long terminal repeat (non-LTR) retrotransposable element from Bombyx mori. Nucleic Acids Res 1993; 21:1318; PMID:8358116; http://dx.doi.org/10.1093/nar/21.5.1318

Benett GA, Benett JS, Yamasaki Y, Ohrn WA, Rice JS. Identification of a helitron-like transposable superfamily from lepidoptera. Science 1993; 259:701-3; PMID:8437716; http://dx.doi.org/10.1126/science.8437716

Xie Y, Mates L, Ivics Z, Izsvák Z, Martin SL, An W. RNA structures indicate an enzymatic activity of the ORF1 protein from the mouse LINE-1 retrotransposon. EMBO J 2002; 21:5899-910; PMID:12411507; http://dx.doi.org/10.1093/emboj/cdf592

Monot C, Kuciak M, Viollet S, Mir AA, Gabus C, Cost GJ, Feng Q, Jacquier A, Boeke JD. Human retrotransposition of murine B1 and B2 SINEs requires a helicase. J Mol Biol 2001; 21:1318-26; PMID:11158378; http://dx.doi.org/10.1006/jmbi.2000.40045

Coates BS, Suescun DV, Hellmich RL, Lewis LC. A helitron-like transposable superfamily from lepidoptera. Science 1993; 259:701-3; PMID:8437716; http://dx.doi.org/10.1126/science.8437716
Hedges DJ, Deininger PL. Inviting instability: T ransposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat Res 2007; 616:66-59; PMID:17157332; http://dx.doi.org/10.1016/j.mrme.2006.11.021

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

Wiermer SM, Kloor M, von Knebel Doeberitz M, Imai K, Yamamoto H. Carcinogenesis and microsatellite instability in colorectal tumors. Genome Res 2000; 10:411-5; PMID:10779482; http://dx.doi.org/10.1101/gr.10.4.411

Solyom S, Ewing AD, Rahrmann EP, Doucet T, Rodi N, Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley A. DNA replication fidelity and cancer. Semin Cancer Biol 2010; 20:281-93; PMID:19657334; http://dx.doi.org/10.1016/j.semcancer.2010.10.009

Laghi I, 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

Torröntes 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.145593.103

Zhao J, Bacolla A, Wang G, Vasquez KM. Non-B DNA conformation adopted by poly(dA) small middle sequences exist as rigid DNA structures in nucleosomes. EMBO J 2000; 19:3358-65; PMID:11058103; http://dx.doi.org/10.1093/embj.2000.19.21.4083

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

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

Moustri AR, Cha 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

Koulaf NG, Garcia-Perez JL, Peng GE, Yoo 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/nature08828

Russer JM, An W. L1 expression and regulation in humans and rodents. [Elite Ed]. Front Biosci (Elite Ed) 2012; 4:2203-25; PMID:22202032

García-Perez JL, Morell M, Scheps JO, Kalpa 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:20686757; http://dx.doi.org/10.1038/nature09209

Ding L, Xu X, Dong Z, Wang G, Cao Y. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Hum Mol Genet 2000; 9:653-7; PMID:10699189; http://dx.doi.org/10.1093/hmg/9.4.653

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

Solyom S, Ewing AD, Hancks DC, Takeshima Y, Awano H, Munso 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

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

Moustri AR, Cha 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

Koulaf NG, Garcia-Perez JL, Peng GE, Yoo 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/nature08828

Russer JM, An W. L1 expression and regulation in humans and rodents. [Elite Ed]. Front Biosci (Elite Ed) 2012; 4:2203-25; PMID:22202032

García-Perez JL, Morell M, Scheps JO, Kalpa 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:20686757; http://dx.doi.org/10.1038/nature09209

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

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

Suter B, Schnappauf G, Thoma F. Poly(dA:dT) tracts sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. Nucl Acids Res 2000; 28:4083-9; PMID:11058103; http://dx.doi.org/10.1093/nar/28.21.4083

Shimizu M, Mori T, Sakurai T, Shindo H. Destabilization of nucleosomes by an unusual DNA conformation adopted by poly(dA) small middle dtopoly(dT) tracts in vivo. EMBO J 2000; 19:3558-65; PMID:10880448; http://dx.doi.org/10.1093/emboj/19.13.3558

Radman TC, Hsu J, Bigg C. Endogenous 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

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