Published in final edited form as:

Adv Genet. 2011 ; 73: 219–262. doi:10.1016/B978-0-12-380860-8.00006-9.

Restless Genomes: Humans as a Model Organism for Understanding Host-Retrotransposable Element Dynamics

Dale J. Hedges* and Victoria P. Belancio†

*Hussman Institute for Human Genomics, Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, Florida, USA

†Department of Structural and Cellular Biology, Tulane School of Medicine, Tulane Cancer Center, Tulane Center for Aging, New Orleans, Louisiana, USA

Abstract

Since their initial discovery in maize, there have been various attempts to categorize the relationship between transposable elements (TEs) and their host organisms. These have ranged from TEs being selfish parasites to their role as essential, functional components of organismal biology. Research over the past several decades has, in many respects, only served to complicate the issue even further. On the one hand, investigators have amassed substantial evidence concerning the negative effects that TE-mutagenic activity can have on host genomes and organismal fitness. On the other hand, we find an increasing number of examples, across several taxa, of TEs being incorporated into functional biological roles for their host organism. Some 45% of our own genomes are comprised of TE copies. While many of these copies are dormant, having lost their ability to mobilize, several lineages continue to actively proliferate in modern human populations. With its complement of ancestral and active TEs, the human genome exhibits key aspects of the host–TE dynamic that has played out since early on in organismal evolution. In this review, we examine what insights the particularly well-characterized human system can provide regarding the nature of the host–TE interaction.

I. INTRODUCTION

Transposable elements (TEs) are, at their essence, stretches of DNA that have the capacity to mobilize themselves to different locations throughout a genome. The wealth of new genomic sequencing data made available over the past decade has allowed for significant advances in our understanding of the distribution and diversity of these elements. In conjunction with important experimental work, the analysis of sequencing data has served to highlight the importance of TE mobilization and proliferation on organismal evolution. From the introduction of alternative splice variants to the reshuffling of exons and entire genes, mobile element activity has constituted a potent force shaping genome architecture (Babushok *et al.*, 2007; Belancio *et al.*, 2008b; Lev-Maor *et al.*, 2003; Moran *et al.*, 1999; Sorek *et al.*, 2002; Xing *et al.*, 2006). TE activity is not relegated to the distant evolutionary past, however. Active mobile lineages continue to propagate within the majority of

eukaryotic species surveyed thus far, including humans. Whereas the activity of these elements was once believed to be focused exclusively within the germline, there is now emerging evidence for substantial TE activity within somatic tissues. Much remains to be understood concerning the impact of this ongoing somatic activity on organismal fitness and, ultimately, upon human health. Here, we examine recent progress in human TE research, focusing on advances in our understanding of the relationship between element and host. Concentrating on the human genome and its current set of TE inhabitants, we examine both the positive and negative consequences of historical and ongoing TE activity.

A. Mobilization and classification

TE mobilization mechanisms can be divided into two broad categories, based on whether an RNA or DNA intermediate is used during the transposition process. These two varieties of mobilization have been conveniently described as “copy and paste” and “cut and paste,” and they form the basis of a common categorization scheme of TEs introduced by Finnegan (1989), which divides TEs into Class I and Class 2 elements. According to the original criteria, Class 2 elements are those that mobilize directly from DNA to DNA, without the use of an RNA intermediate (Van Duyne, 2002). With some exceptions, noted below, DNA transposition occurs through a “cut and paste,” strategy, where the original double-stranded DNA source element is excised from its existing location and reintroduced to a novel location in the genome. This excision is accomplished by transposase proteins that are encoded directly by the transposons themselves, or, in some cases, hijacked from other transposons in the host genome that possess protein-coding capacity. These same transposase proteins also facilitate the reintegration of the excised element elsewhere in the genome. With some important exceptions that allow for copy number increase, this type of mobilization is conservative in nature, resulting in no overall increase in the total number of TE elements within a host genome. It was representatives of the Class 2 type of elements that McClintock first observed in maize (McCLINTOCK, 1950). Discoveries of additional DNA transposon varieties that, despite their use of a DNA intermediate, did not include a double-stranded DNA removal step during the transposition process, necessitated the updating of Class 2 DNA transposon class to include three distinct subcategories. These include the original “cut and paste transposons,” Helitrons, and Mavericks (reviewed in Feschotte and Pritham, 2007). The second major class of TEs, termed Class 1 elements (also commonly referred to as retrotransposons), mobilize by generating RNA transcripts that are subsequently converted into DNA prior to reintegration at a novel location in the host genome. Again, the process is mediated by proteins that are either encoded directly by the retrotransposon itself or hijacked from other protein-coding TEs. This mobilization strategy, termed retrotransposition, is an inherently expansive force within genomes. The total number of elements increases each time a successful retrotransposition event occurs, and, as a direct consequence, so does the total size of the host genome. Retrotranspositional activity has played a tremendously important role in genome expansion and diversification during evolutionary history, providing, in the process, abundant material for natural selection to carve out new functions. It is frequently noted that TEs from various lineages comprise roughly 45–50% of the human genome (Lander *et al.*, 2001). A considerable fraction of the remaining half of our genome likely has an origin in mobile element activity as well, but

these elements have since diverged so far from their initial sequences that they can no longer be readily identified through nucleotide or protein homology to known element classes.

B. LINEs and Tigger (the diversity and complexity of TE repertoire)

Beyond the simple classification scheme of TEs into Class 1 and Class 2 elements, there exists an ever-growing bestiary of mobile element lineages, with colorful names ranging from “SPACE INVADERS” to “Tigger.” Identifying and classifying TEs from diverse species has fueled the development of an array of important software tools as well as the establishment of curated databases. Examples include Censor (Jurka *et al.*, 1996) and RepeatMasker (www.repeatmasker.org) for identifying elements based on known libraries of elements, and several methods, including RepeatScout, Recon, RepeatFinder, PILER, and ReAs for the *ab initio* discovery of elements (Edgar and Myers, 2005; Li *et al.*, 2005; Price *et al.*, 2005; Quesneville *et al.*, 2005; Volfovsky *et al.*, 2001). Repbase and dbRIP provide information about categorized TEs and the polymorphic status of particular inserts, respectively (Jurka *et al.*, 2005; Wang *et al.*, 2006). As the genome of any given organism may hold from a few to hundreds of TE varieties, each of which is diverging independently from its relatives inhabiting the genomes of other organismal species, TE nomenclature has the potential to be orders of magnitude more extensive and cumbersome than analogous systems for organisms. The discovery of TEs that propagate in a “copy and paste” manner without the need for an RNA intermediate served to undermine the basis of the Type1/Type2 classification scheme (Morgante *et al.*, 2005; Wicker *et al.*, 2007). It has, however, proven challenging to devise a taxonomic system for mobile elements that manages to reflect the ancestral relationships among elements, while, at the same time, avoids being too unwieldy for researchers to practically employ on a regular basis. As a consequence of this challenge, multiple categorization schemes have been proposed and are simultaneously in use (Wicker *et al.*, 2007). The past decade’s precipitous increase in sequencing capacity, and the “tsunami” of data that continues to ensue from it, has led to the identification of a growing list of TEs from an increasingly diverse set of sampled organisms. The rising abundance of mobile element literature from this wealth of data poses a formidable challenge when attempting a reasonably comprehensive overview of progress in the area of TE research. We cannot hope, within the scope of this review, to address the full diversity of elements and their associated biology, or even to do adequate justice to describing what is known about the major TE branches. As our own research largely focuses on human TEs, we will primarily concentrate on progress in our understanding of the human host–element relationship (Fig. 6.1). Unlike many other higher eukaryotes, humans are currently bereft of active Class 2 type TEs, although elements of this variety existed relatively recently (evolutionarily speaking) within our primate lineage (Pace and Feschotte, 2007). Despite the human and primate-centric focus of this review, it is of impossible to avoid noting that critical insights into both the molecular biology and evolution of TEs have been obtained from the study of TEs from a wide assortment of nonhuman organisms. We attempt to draw upon these diverse systems throughout our review to illustrate important concepts in the host–element relationship.

Advances in the field of TE research over the past two decades have effectively put to rest lingering questions about their overall relevance within biology. Although once relegated as

mere evolutionary curiosities, clear and abundant evidence of their far-reaching impact now extends across diverse disciplines, ranging from evolutionary biology to medical genetics. This new perspective on mobile elements is perhaps best summed up by Lynch (2007), "... they are so ubiquitous, so diverse, and have such a profound effect on eukaryotic chromosomal architecture that one can reasonably argue that an overview of genomic evolution ought to start with them, before moving on to the host genes themselves" (Lynch, 2007). Among other things, the activity of TEs has been an expansionary force in the genome that provided abundant fodder with which natural selection could experiment. TEs have been exapted to perform functional biological roles numerous times during the history of life (reviewed in Bowen and Jordan, 2007; Dooner and Weil, 2007). The fact that such an important role was played by TEs is no longer matter of debate. That this role was cultivated and maintained by natural selection to promote "evolvability" is a stronger claim, and one for which both theoretical and empirical support remains lacking. This last claim is closely related to what has been an enduring question in TE research. Namely, how do we best describe the unusual relationship between host genomes and their TE residents? Is it mutualistic, symbiotic, parasitic, or something that defies our typical classification schemes? Views range from, on the one extreme, TEs as vital components of organismal biology to, at the other extreme, completely parasitic selfish genes undermining organismal fitness. The true nature of the host-TE dynamic no doubt rests somewhere between these polar extremes. A more nuanced view is offered by Kidwell and Lisch (1997), when they write, "the idea that TEs are primarily parasitic is not at all inconsistent with a role for these elements in the evolution of their hosts." Throughout this review, we indicate empirical evidence supporting what is ultimately an ambivalent role for TEs in organismal biology.

II. THE HUMAN ELEMENTS

The TE families remaining active in modern humans include the LINE-1 (L1) element, Alu, SVA, and human endogenous retroviruses (HERVs). All of these lineages belong to the Type 1, retrotransposon class (Fig. 6.1). Their basic structure is outlined below; a more detailed description of their molecular biology is provided in Belancio *et al.* (2010a).

A. L1

L1 elements are members of the long interspersed element. The L1 sequence encodes a bicistronic polymerase II (Pol-II) transcript that produces two proteins labeled ORF1 and ORF2, both of which are necessary for successful retrotransposition. ORF1's function is poorly understood, although it has a strong affinity for the LI RNA sequence and is reported to exhibit nucleic acid chaperone activity (Martin and Bushman, 2001; Martin *et al.*, 2005). The larger ORF2 protein encodes a protein harboring both endonuclease (EN) and reverse transcriptase (RT) activities. The protein itself has multiple functions that are conferred by the presence of several distinct domains; these include an N-terminal EN domain (Feng *et al.*, 1996), a central RT domain (Mathias *et al.*, 1991), and a C-terminal cysteine-rich domain of unknown function (Fanning and Singer, 1987). LI and the nonautonomous retroelements (such as Alu and SVA, described below) replicate via a process known as target-primed reverse transcription (TPRT; Cost *et al.*, 2002; Luan *et al.*, 1993), wherein reverse transcription of the RNA intermediate is primed by an exposed 3' hydroxyl. The 3'

hydroxyl strand of the host genome is exposed as the result endonucleatic cleavage by ORF2. Based on protein sequence analysis, the L1 lineage descends from a group of non-LTR (long terminal repeat) elements (elements lacking long terminal repeats) that have existed, in some form or another, since early eukaryotic evolution (Xiong and Eickbush, 1990). It remains a matter of debate whether L1 and similar elements were the predecessors or ancestors of modern viruses (Xiong and Eickbush, 1988). There are approximately 500,000 L1 copies in the human genome (haploid), comprising roughly 17% of total genome content (Lander *et al.*, 2001). Only a fraction of these elements are intact, full-length elements that retain the ability to retrotranspose. Most elements are either 5' truncated during the insertion process, deactivated through postinsertional mutation events, or otherwise removed from the host population due to genetic drift and/or natural selection. There is evidence that the majority of retrotranspositional activity actually results from a relatively small number of particularly active, "hot," L1 elements (Beck *et al.*, 2010; Brouha *et al.*, 2003; Seleme *et al.*, 2006). As is the case with the Alu, SVA, and—to a lesser extent—HERV elements described below, modern humans are polymorphic with regard to the insertion status of numerous L1 inserts, making L1 an integral component of human genetic diversity (Beck *et al.*, 2010; Ewing and Kazazian, 2010; Huang *et al.*, 2010; Iskow *et al.*, 2010).

B. Alu

Alu are classified as Short INterspersed Elements (SINEs). These 300 bp Pol-III transcribed sequences are comprised of two separate domains tethered by an A-rich linker region. Originally derived from 7SLRNA, which is a member of the signal recognition complex, the current dimeric Alu found in humans formed from the merger of two monomeric units during early primate evolution. There are approximately 1 million copies of Alu in the human genome (haploid), constituting roughly 10% of the total genomic content (Lander *et al.*, 2001). Alu elements do not encode the protein machinery required for their retrotransposition. Instead, it has been experimentally demonstrated that they can effectively commandeer L1 protein products to facilitate their proliferation (Dewannieux *et al.*, 2003). While the ORF2 protein is sufficient for their retrotransposition (Dewannieux *et al.*, 2003), the efficiency of Alu transposition can be improved by the ORF1 expression (Wallace *et al.*, 2008a).

C. SVA

SVA elements arose more recently in the evolutionary timeline than the other active elements, appearing later in primate evolution than Alu (Ostertag *et al.*, 2003; Wang *et al.*, 2005; classified as nonautonomous, they defy easy categorization into existing schemas due to their intermediate size and chimeric nature). These odd elements appear to be cobbled together from several distinct genomic sources, including Alu-like sequence as well as sequence from viral envelope genes and LTR regions (Ostertag *et al.*, 2003). Lacking protein-coding capacity, they are nonautonomous elements that parasitize L1 to obtain the necessary protein machinery for retrotransposition (Ostertag *et al.*, 2003). The presence of several potential Pol-III termination sequences within their consensus sequences argues for a Pol-II transcription and the involvement of the serendipitous adjacent host promoters in driving SVA transcription has been reported (Damert *et al.*, 2009).

D. Human endogenous retroviruses

HERVs are thought to be the ancestors of ancient retroviruses that infected the germline and, subsequently, lost their ability to encode active envelope proteins and move from host to host. Collectively, ERV copies make up ~8% of the human genome with human-specific ERVs (i.e., HERVs) constituting a smaller fraction (Lander *et al.*, 2001). Due to their ability to be brought in through the infection of the germline by exogenous retroviruses, HERVs have been introduced and reintroduced a multitude of times during human evolution. While their rate of proliferation in modern human germlines appears to be limited, with only a modest number of polymorphic locations being identified (Turner *et al.*, 2001), the activity of their remnant promoters has significant biological consequences, having been often adopted in several instances for host genome regulation [reviewed in Maksakova *et al.*, 2006; Sverdlov, 2000]. Also, see Cohen *et al.* (2009) for a more critical examination of the use of LTRs from endogenous retroviruses as functional human promoters (Cohen *et al.*, 2009).

E. Brief evolutionary history

Despite the fact that a spectacularly diverse set of TE lineages proliferated during the course of metazoan evolution, only a handful of families remain active within modern humans (Furano *et al.*, 2004; Pritham and Feschotte, 2007). It has been noted, for example, that mammals in general retain far fewer ancient L1 lineages compared to the fish species that have been examined thus far, and, in primates, this number has been whittled down to a single L1 family during the course of human evolution (Furano *et al.*, 2004). The bulk of ancient transposon diversity appears to have been lost during the evolution of tetrapods, possibly during synapsid evolution (Kordis *et al.*, 2006). The forces leading to this shift in TE genomic ecology remain unclear. Both demographic processes (e.g., population bottlenecks) and molecular changes are likely to have played their part in the transition. Evidence that only a handful of members of the L1 elements—so called “hot” elements—exhibit the highest levels of activity (Brouha *et al.*, 2003; Seleme *et al.*, 2006), suggest that random demographic fluctuations could have profound impacts on the persistence of TE diversity. As ecological sustainable population sizes generally decrease within higher trophic levels, loss of TE diversity could result from the fact that the loss of active lineages due to genetic drift occurs more rapidly than new lineages are emerging. If drift were the predominant force driving the loss of TE diversity, however, we would arguably find more examples of lineages that have lost their entire L1 complement (Cantrell *et al.*, 2008; Casavant *et al.*, 2000). The adaptive evolution of TE nucleotide and protein sequences, possibly to evade host repression mechanisms, likely play a substantial role in determining levels of TE diversity. Boissinot and Furano (2001) provided convincing evidence, based on the ratio of nonsynonymous to synonymous amino acid changes, that adaptive evolution occurred in the L1 lineage during the last 25 myrs of human evolution (Boissinot and Furano, 2001), although the evolutionary pressures driving these changes have not been definitively established.

Alu elements arrived considerably later on the evolutionary scene than L1 and have enjoyed spectacular success, reaching approximately 1 million copies in the haploid human genome. Thought to have evolved from the human *7SLRNA* gene during early primate radiation, Alus

have far outperformed their L1, SVA, and HERV counterparts. Examination of sequence data does suggest, however, that the overall amplification rate of both Alu and L1 elements has attenuated over time from a peak rate approximately 40–50 mya (Ohshima *et al.*, 2003). The cause of the amplification burst and subsequent attenuation has not been established, although changes in host molecular biology and/or demographic effects, such as population bottlenecks, have been suggested. Both the evidence of changing levels of TE diversity, as well as for large fluctuation in insertion rates over evolutionary time, demonstrate that the relationship between TE and host genome is not a static one. Both host and TE are concurrently evolving and external perturbations from the environment can cause the landscape of insertion rates and element diversity to shift, sometimes dramatically. Across taxa, this can be observed in the existence of a diverse genomic ecology of TEs and host [discussed in Brookfield, 2005]. These range from systems containing a large diversity of elements, each of which have but a few genomic instances, to those systems, like the human, where a select few families achieve very high copy numbers. While many important questions remain to be answered regarding the impact of host demographic history on TE expression, it has become clear that factoring heavily into the resulting TE ecology is the extent to which host molecular biology augments or intervenes with TE proliferation. To understand how the molecular biology of the host cell can help or hinder the mobilization process, we will now examine in more detail the life cycle of the human retrotransposon.

III. EXPRESSION AND REGULATION

In order to survive and impose lasting genetic alterations on future generations, TE-associated modification of the host genome must ultimately occur in the germline—at least within organisms that have a sequestered germline. From the selfish gene perspective, activity in somatic cells would appear to only reduce genetic fitness of the host without resulting in any increase in TE copy number. This relatively simple—yet ultimately incomplete—evolutionary logic resulted in decades of near-exclusive focus on the germline as the principal site of mammalian TE expression. While TE expression and proliferation within the germline remains the only means of ensuring long-term TE survival, as we describe below, there is now emerging evidence that somatic TE activity is considerably higher than once suspected. A more complete characterization of this somatic activity is a vital component of understanding the full impact of TEs on organismal fitness and, ultimately, human health, and well being. On account of the ongoing tension in the TE–host coexistence, a network of defense mechanisms has been erected by host cells to shield the genome from unchecked TE activity. An understanding of the TE life cycle is essential for appreciating the breadth and complexity of this regulatory network.

A. Life cycle of human retroelements

Among the human TEs that have retained their activity throughout the course of evolution are three main groups of non-LTR retrotransposons (L1, Alu, and SVA) that stand apart from the LTR group of retroelements such as HERVs. Although all TEs are frequently considered to have a parasitic aspect, the nonautonomous group of non-LTR transposons, such as SINEs (represented by Alu elements in the primate lineage) and SVA (hominoid-specific retroelements) take this parasitism to another level. Because SINEs and SVA

elements do not encode any proteins for orchestrating their own mobilization (hence the “nonautonomous”) they have evolved to parasitize the retrotransposition machinery of the currently active human autonomous non-LTR retroelement, LINE-1. The retrotransposition process of non-LTR retrotransposons begins with transcription of a retrotranspositionally competent locus within the host genome. L1, Alu, and SVA elements accumulated in the human genome to 500,000, 1,000,000, and 1700 copies, respectively, the majority of which are inactive (Lander *et al.*, 2001). Out of the structurally intact TE copies only a fraction of loci retained functional promoters. Finally, a large proportion of the potentially expressed loci have lost the ability to mobilize due to mutations inactivating functional protein domains or sequences within their RNAs critical for amplification. The former applies only to L1 elements (because only they encode proteins), while the latter is true for all human retrotransposons. As a result there are about 100 predicted active L1 loci per haploid human genome (Penzkofer *et al.*, 2005). L1 elements are transcribed by an atypical bidirectional internal RNA Pol-II promoter located within the L1 5' untranslated region, 5'UTR (Speek, 2001; Swergold, 1990). At least some of the produced L1 mRNAs are capped and RNA capping is reported to stimulate translation of L1 proteins (Dmitriev *et al.*, 2007; Kulpa and Moran, 2005; McMillan and Singer, 1993). As a result of Pol-II transcription, L1 mRNAs are polyadenylated. The presence of this polyA tail at the 3' end is critical for the L1 integration process (Symer *et al.*, 2002).

Very little is known about transcription of SVA elements. They are most likely transcribed by RNA Pol-II because of the presence of multiple RNA Pol-III terminators within their sequence and evidence for posttranscriptional processing of SVA RNA characteristic of the Pol-II generated RNAs (Damert *et al.*, 2009; Hancks *et al.*, 2009; Ostertag *et al.*, 2003; Wang *et al.*, 2005). Bioinformatic analysis of SVA elements identified in the human genome, followed by empirical testing, suggested that SVA retrotransposons do not contain an internal promoter that can be carried over into the new genomic location to ensure efficient transcription of the *de novo* insertions (Damert *et al.*, 2009). SVA loci appear to rely on the existence of the functional promoters at the sites of their integration (Damert *et al.*, 2009). Along with their relatively recent appearance on the evolutionary scene, the lack of a mobile promoter may explain their relatively low copy numbers.

In contrast to L1 and SVA elements, Alu transcription is carried out by the RNA Pol-III complex as is the case with its 7SLRNA ancestor. Alu RNA is not capped and the artificial generation of the capped Alu transcripts significantly changes requirements for Alu mobilization (Kroutter *et al.*, 2009). Similarly to the L1 and SVA elements, Alu RNAs contain a polyA tail at their 3' end, but in contrast to the polyA tails of L1 and SVA, it is encoded within genomic Alu sequences rather than being generated through polyadenylation by cellular proteins.

Mammalian L1 elements encode two proteins, open reading frame 1 (ORF1) and open reading frame 2 (ORF2), that are absolutely essential for L1 retrotransposition (Fig. 6.1). L1 ORF1 and ORF2 proteins are presumed to be made in the cytoplasm of the cells that permit transcription of a full length, bicistronic L1 mRNA. Very little is known about L1 mRNA translation and even less about the mechanisms that may regulate L1 protein production (Alisch *et al.*, 2006; Dmitriev *et al.*, 2007; Leibold *et al.*, 1990; Li *et al.*, 2006; McMillan

and Singer, 1993). Extensive empirical evidence points out that L1 ORF1 plays a crucial role in the formation of ribonucleoprotein (RNP) complexes with L1 mRNA that are believed to represent important retrotransposition intermediates (Kolosha and Martin, 1997, 2003; Kulpa and Moran, 2005; Martin, 1991; Martin *et al.*, 2000). L1 ORF2 is also associated with the L1 RNPs (Kulpa and Moran, 2005; Wei *et al.*, 2001) and these L1 RNA/protein complexes gain access by an undefined passive or active mechanism to genomic DNA. L1 ORF2 contains apurinic-like EN and RT domains that are responsible for a series of sequential enzymatic steps in the LINE, SINE, and SVA integration process (Clements and Singer, 1998; Cost and Boeke, 1998; Cost *et al.*, 2002; Feng *et al.*, 1996; Martin *et al.*, 1998). While the requirement for the functional L1 ORF2 is shared by all three non-LTR retroelements (LINEs, SINEs, and SVA), L1 ORF1 protein has been reported to be indispensable only in the case of L1 mobilization (Dewannieux *et al.*, 2003; Moran *et al.*, 1996). Currently there is very limited understanding of how, when, and where Alu and SVA RNAs hijack the L1 ORF2 protein. The means of gaining access to L1 ORF2 by SINEs and SVAs are complicated by the existence of a very strong cis-preference of L1 proteins for the specific L1 mRNA instance that was the source of their translation (Wei *et al.*, 2001). This cis-preference may be one of the mechanisms that serve to prevent retroelements from wreaking havoc in the host genome through mobilization of otherwise retrotranspositionally incompetent TE RNAs and normal cellular transcripts resulting in the bombardment of the human genome with retro-pseudogene copies. Nevertheless, despite the apparent barriers to commandeering the L1 mobilization apparatus, both SINEs and SVA elements have been extremely successful at doing so (Lander *et al.*, 2001; Wang *et al.*, 2005), suggesting a potential active mechanism of gaining access to the L1 ORF2. Alu elements have been particularly adept at this process, which is likely one of the reasons that allowed them to flourish to over 1,000,000 copies in the human genome surpassing L1 copy number by more than a half (Lander *et al.*, 2001).

The initiation of retrotransposition in the nucleus involves recognition of an EN target site within the host genomic DNA by the L1 ORF2 followed by the first-strand DNA cleavage by the L1 ORF2 EN that produces a DNA nick at the EN site (Cost and Boeke, 1998; Feng *et al.*, 1996; Jurka, 1997). Even though the consensus L1 EN site is deduced to be 5'-TTTTAA-3', some of the sites with a single nucleotide substitution within this sequence are known to be efficiently used by the L1 EN (Jurka, 1997; Szak *et al.*, 2002). The cleavage of genomic DNA at the T-A junction is proposed to generate a single strand (ss), 5'-TTTT-3', DNA available for a noncovalent interaction with the polyA tail at the 3' end of the L1 mRNA (or SINE and SVA RNAs; Feng *et al.*, 1996; Symer *et al.*, 2002). Free 3' end of the ss genomic DNA serves as a primer for the L1 RT to synthesize first-strand L1 cDNA (Martin *et al.*, 1998; Piskareva and Schmatchenko, 2006). This process is known as TPRT and is common among retroelements from species as diverse as the silk worm, *Bombyx mori*, and human, *Homo sapiens* (Christensen and Eickbush, 2005; Cost *et al.*, 2002). The remaining steps of the retrotransposition process that include second strand L1 DNA synthesis (Piskareva and Schmatchenko, 2006), removal of the RNA template, and covalent connection of the *de novo* L1 (or Alu and SVA) insert with genomic DNA at the site of integration continue to be very poorly characterized. It is speculated that cellular DNA repair proteins or other cellular factors are likely to assist in the completion of some or all of

these steps. This assumption is based on the apparent lack of certain enzymatic functions within L1 proteins, such as ligase activity, that would be required to carry out these steps. It is further based on the recent discoveries of the opposing effects of cellular proteins from various DNA repair pathways on the efficiency of L1 retrotransposition (Gasior *et al.*, 2007, 2008; Suzuki *et al.*, 2009). The final product of retrotransposition contains distinct structural characteristics such as L1, SINE, SVA, or cellular transcript sequences flanked by the EN recognition site known as the target site duplication (TSD). These signature features unequivocally identify L1 machinery involvement in the generation of integration events.

Even though L1, SINEs, and SVA elements rely completely on the L1 retrotransposition apparatus for their mobilization, different elements exhibit distinct differences between the retrotransposition requirements. One of these prerequisites is the dependence or the lack of it on the L1 ORF1 protein (Dewannieux *et al.*, 2003; Moran *et al.*, 1996). As mentioned above, in contrast to L1, which absolutely requires participation of the functional ORF1 protein in addition to the ORF2 protein for successful retrotransposition, Alu elements can be efficiently mobilized by the L1 ORF2 alone (Dewannieux *et al.*, 2003; Moran *et al.*, 1996). While the presence of the L1 ORF1 protein slightly stimulates Alu retrotransposition in tissue culture (Wallace *et al.*, 2008a), this modest effect supports the observation that Alu retrotransposition is largely independent of the L1 ORF1. The proposed explanations for this discrepancy include the difference in the origin of transcripts through RNA Pol-II in the case of L1 mRNA and Pol-III in the case of Alu, as well as the association with the cellular translation machinery (neither Alu nor SVA transcripts are believed to contain any ORFs; Kroutter *et al.*, 2009). Interestingly, an artificial switch from Pol-III to Pol-II generated Alu transcripts results in the requirement of ORF1 for Alu mobilization (Kroutter *et al.*, 2009).

The multistep life cycle of human TEs provides ample opportunities for the host to erect barriers that would allow prevention of efficient amplification of TEs. Understanding of the TE life cycle makes it very clear that suppression of TE expression can be a very effective way of minimizing their negative impact on the genome stability.

B. TE expression

TEs are families of elements interspersed throughout the human genome, as a result, each of human TEs is represented by hundreds or thousands of nonidentical loci that can be expressed in any given cell (Lander *et al.*, 2001). Thus, endogenous expression detected by conventional techniques is usually a combination of transcripts produced by a subpopulation of loci expressed in the assessed cell type. Due to the sequence variation among the majority of TE loci found in the human genome, each method used to analyze TE expression is often accompanied by its own set of specific detection biases, discussion of which is beyond the scope of this review. Technical difficulties associated with detection of human TEs combined with the presumed relatively low expression levels and incomplete knowledge of rules governing TE expression have accounted for the impediment of progress in this area of TE research. The expression of endogenous L1 elements (mRNA and ORF1 protein) was reported to be restricted to the mouse germline (Branciforte and Martin, 1994; Ostertag *et al.*, 2002). The expression of the L1 proteins was also detected in specific somatic cell types (such as Leydig, Sertoli, and vascular endothelial cells; Ergun *et al.*, 2004) and in normal

breast tissues (Asch *et al.*, 1996). Analysis of endogenous L1 expression in various types of human tumors and cancer cell lines established that L1 expression, as a rule, is significantly upregulated in human malignancies (Bratthauer and Fanning, 1992, 1993; Bratthauer *et al.*, 1994). As is obvious from the TE life cycle (described above), ongoing L1 expression is necessary for the activity of Alu and SVA elements. Even though it is deduced from the Alu and SVA retrotransposition events occurred in the germline that both of these elements are expressed at minimum in the germline, the endogenous expression of Alu and SVA elements is very poorly characterized (Fuhrman *et al.*, 1981; Shaikh *et al.*, 1997). Few studies have attempted to assess Alu transcripts in a manner that distinguishes Pol-II- and Pol-III- promoted transcripts (Shaikh *et al.*, 1997).

Several lessons have been derived concerning L1 expression from transgenic mouse models. Transgenic animals containing L1 expressed from its native promoter supported L1 expression in testis and ovaries (Ostertag *et al.*, 2002). Pol-II driven L1 was also expressed in kidney, lung, intestine, liver, and brain (Ostertag *et al.*, 2002). RT-PCR analysis of germ cell fractions such as pachytene spermatocytes, round spermatids, and condensing spermatids demonstrated L1 expressed with varying efficiency within these three cell types (Ostertag *et al.*, 2002). Another transgenic mouse model of human L1 driven by the mouse pHsp70-2 promoter demonstrated very strong positional effect of the transgene integration site on retrotransposition (Babushok *et al.*, 2006). This observation suggested that L1 transgene expression in the transgenic mouse model can be significantly affected by its specific location, consistent with the reports of positional effects on the expression of endogenous L1 loci (Lavie *et al.*, 2004). Transgenic mouse and rat models of human and mouse L1 retrotransposition driven by their respective endogenous promoters demonstrated L1 mRNA presence in both germ cells and embryo. In these models retrotransposition events occurred in embryogenesis creating nonheritable somatic mosaicism (Kano *et al.*, 2009). In contrast to the previous models, a transgenic animal expressing a single copy of the synthetic mouse L1 element demonstrated significantly higher retro-transposition compared to retrotransposition frequencies detected in animals containing multiple integrated copies of the same L1 transgene (An *et al.*, 2006, 2008). Retrotransposition in a transgenic mouse model was also reported to take place in the mouse brain strongly supporting L1 expression in normal brain cells (Muotri *et al.*, 2005). Analysis of the methylation status of endogenous L1 promoters coupled with detection of *de novo* L1 retrotransposition events (Coufal *et al.*, 2009) in normal human brain produced results consistent with ongoing expression of endogenous L1 elements in human brain cells. More comprehensive analysis of endogenous L1 mRNA expression in normal human tissues and adult stem cells demonstrated that the majority of the examined human tissues support endogenous L1 mRNA expression (Belancio *et al.*, 2010b). However, the abundance of the total L1-related transcripts and the amount of the full-length L1 mRNA varies significantly among tissues. This variation appears to be at least in part influenced by the tissue-specific differences in the efficiency and pattern of L1 mRNA processing (Belancio *et al.*, 2010b). The same is true for human cancer cell lines where the amount of the full-length L1 mRNA is dictated by the efficiency of L1 mRNA splicing and premature polyadenylation (Belancio *et al.*, 2010b). Somatic endogenous L1 expression combined with the lack of obvious restrictions for somatic retrotransposition (An *et al.*, 2006, 2008; Kano *et al.*, 2009; Kubo *et*

al., 2006) and L1-induced double-strand DNA break (DSB) formation in normal cells (Belancio *et al.*) suggests that ongoing L1-induced DNA damage is likely to take place in somatic tissues. Because of the reported variation in the L1 promoter strength in different cell types (Swergold, 1990; Yang *et al.*, 2003) and the reported variation in the extent of L1 mRNA processing (Belancio *et al.*, 2010b) the exposure to the L1-associated DNA damage is expected to vary in a tissue-specific manner. Along with the significant broadening of the potential spectrum of L1-induced DNA damage comes the importance of understanding of the time, location, and levels of TE expression in human tissues making unraveling of the somatic L1 expression an important aspect of L1 biology.

C. Mechanisms controlling TE expression and activity

Even though many critical elements controlling TE expression and activity remain unknown, a plethora of cellular mechanisms that control human TE expression and activity has been unveiled to date. Based on the fact that TEs can significantly perturb the normal function of the genome it is not surprising that host organisms employed a multifaceted network of road blocks designed to suppress TE expression and mobilization. Of the three currently active human TEs, L1 expression and activity is the best studied to date, mostly due to the fact that L1s are the driving force of any TE-related unrest in the human genome. Among the well-characterized mechanisms influencing human TE expression and activity are transcriptional regulation, promoter methylation, RNA processing, and cellular pathways involved in controlling TE-induced damage. The number and diversity of the barriers erected by mammalian cells to suppress TE expression speaks to the importance of keeping the production of these elements in check.

One of the first levels of defense controlling TE expression is promoter strength and availability of the transcription factors necessary to drive RNA production. Members of the SOX and RUNX family of transcription factors as well as steroid hormones can regulate L1 expression dictating tissue specificity of L1 transcription (Morales *et al.*, 2002; Tchenio *et al.*, 2000; Yang *et al.*, 2003). YY1 transcription factor does not change transcription efficiency, but it is important for the proper site of transcription initiation at the L1 promoter (Athaniar *et al.*, 2004). Genomic sequences immediately upstream of the L1 integration site can stimulate L1 promoter strength potentially rendering some L1 loci the ability to contribute more molecules than others to the combined pool of L1 transcripts (Lavie *et al.*, 2004). A similar situation is reported for Alu elements; while transcription of Alu promoter alone by Pol-III is relatively weak but upstream genomic sequences can stimulate it by several fold (Roy *et al.*, 2000). Expression of SVA elements is reported to be driven by cellular promoters present upstream of the SVA integration sites suggesting that these elements are likely to exhibit significant variation in their total expression, with different SVA loci differentially contributing to the combined pool of SVA transcripts (Damert *et al.*, 2009).

Promoter function can be significantly regulated through the promoter methylation and epigenetic modifications. Utilization of promoter methylation to repress TE expression is thought to be a common cellular defense mechanism against TE-associated damage [reviewed in Yoder *et al.*, 1997]. The promoters of the majority of the L1 and Alu loci are

hypermethylated in normal cells (Alves *et al.*, 1996; Florl *et al.*, 1999; Hata and Sakaki, 1997; Lees-Murdock *et al.*, 2003; Takai *et al.*, 2000; Thayer *et al.*, 1993; Tsutsumi, 2000). The importance of the maintenance of their methylation status is confirmed in the transgenic mice deficient for DnmtL3 protein required for upholding of DNA methylation (Bourc'h and Bestor, 2004). Significant upregulation of the expression of endogenous retroviruses and L1 elements accompanied the lack of proper DNA methylation in these transgenic animals. The effect of deficient methylation of genomic DNA on the expression of SINEs has not been tested in this experimental model. Retrotransposon methylation is also regulated by Piwi family members MILI and MIWI2 (Kuramochi-Miyagawa *et al.*, 2008). Despite hypermethylation, there is a small portion of loci, at least in the case of L1 elements that escape methylation (Coufal *et al.*, 2009). This is consistent with the ongoing endogenous L1 expression in human somatic tissues (Belancio *et al.*, 2010b). Methylation control of TE transcription is likely released during the course of embryogenesis when methyl groups are removed in the global manner to reestablish genome methylation (Hellmann-Blumberg *et al.*, 1993). The loss of L1 and Alu promoter methylation is also reported in the majority of human cancers (Takai *et al.*, 2000; Tsutsumi, 2000) and it is routinely used as a marker of transformation.

Once transcription occurs, mRNA processing, specifically as it relates to L1 expression, plays an important role in dictating how much of the functional L1 mRNA is made. This processing involves both RNA splicing and premature polyadenylation at the polyadenylation (pA) sites abundant within L1 ORFs (Belancio *et al.*, 2006; Han *et al.*, 2004; Perepelitsa-Belancio and Deininger, 2003). Both processes exhibit tissue specificity and can lead to the production of L1-related transcripts that are retrotranspositionally incompetent at the expense of generation of full-length L1 (Belancio *et al.*, 2010b). This level of control exists in both normal and cancer cells (Belancio *et al.*, 2010b). It is currently unclear whether L1 RNA processing in cancer cells changes or remains the same as it was in the normal cells that they originated from. Nevertheless, mRNA processing can limit L1 expression by as much as 10-fold among normal cells and it is responsible for at least fivefold difference in the full-length L1 mRNA expression between cancer cells that otherwise support the same steady-state total L1-related mRNA (Belancio *et al.*, 2010b). Similar to L1, SVA elements also contain functional splice sites that are efficiently used during SVA transcription (Damert *et al.*, 2009). Alu elements that are transcribed by the Pol-III machinery are observed to be immune to the effects of RNA processing observed for L1 and SVA elements.

Once functional TE RNAs are produced, the next defense mechanism targets transcripts stability. The implementation of the cellular machinery to reduce steady-state L1 mRNA levels has been shown through several experimental systems. One such system implicated siRNA production from the antisense L1 transcripts generated by the reverse L1 promoter itself. Endogenously expressed L1 siRNAs were detected in human cells. This regulation through RNA interference was reported to exhibit a twofold effect on L1 expression. Another example of L1 mRNA regulation was shown in transgenic animals lacking PIWI protein. The lack of functional piRNA (or rasiRNA) pathway resulted in multifold upregulation of endogenous retroviruses and L1 elements through their effect on

methylation (Kuramochi-Miyagawa *et al.*, 2008). In later studies, siRNA against PIWI in cultured human cells was shown to lead to the upregulation of endogenously expressed L1 transcripts (Lin *et al.*, 2009).

Because very little is known about production of human L1 proteins, particularly the ORF2 protein, some conflicting mechanisms controlling L1 translation have been put forward (Alisch *et al.*, 2006; Dmitriev *et al.*, 2007; Leibold *et al.*, 1990; Li *et al.*, 2006; McMillan and Singer, 1993). It is, however, known that wild-type L1 sequence composition favors suboptimal codon usage (Han and Boeke 2004; Wallace *et al.*, 2008b). Codon optimization of the mouse L1 element that also eliminated the majority of polyadenylation sites present in L1 sequence resulted in a significant increase in L1 mRNA, presumably L1 proteins, and L1 retrotransposition rate (Han and Boeke 2004). Codon optimization of the human L1 ORF2 also lead to a measurable increase in protein production (Wallace *et al.*, unpublished data). The L1 life cycle presumably involves steps that take place in different cellular compartments (nucleus and cytoplasm). Almost nothing is known about the trafficking of L1 mRNA, proteins, and/or RNPs and as a result little is known about targeted intracellular trafficking (Goodier *et al.*, 2004, 2007). While the intermediate steps in L1-mediated integration still remain poorly understood, the involvement of several cellular proteins associated with antiviral defense and DNA repair has been reported. The APOBEC family of proteins that function as RNA editing enzymes exhibit an effect on L1 and Alu retrotransposition that seems to be independent of their enzymatic activity through which they reduce viral RNA (Bogerd *et al.*, 2006; Chiu *et al.*, 2006; Hulme *et al.*, 2007; Muckenfuss *et al.*, 2006; Stenglein and Harris, 2006).

An emerging role of cellular DNA repair machinery in L1 retrotransposition is exciting because it introduces an interesting dilemma into the TE– host relationship. On one hand, DNA repair proteins (ATM) are required for L1 mobilization (Gasior *et al.*, 2006), that is, the host is assisting parasite invasion. On the other hand, ERCC1 protein of the cellular nucleotide excision repair (NER) pathway negatively regulates L1 retrotransposition (Gasior *et al.*, 2008) serving a protective role against L1 assault. Hence the dependence on host proteins for the completion of retrotransposition may serve to provide an important avenue through which the host cell can regulate TE proliferation levels.

Very little is known about posttranscriptional regulation of Alu and SVA elements. In the case of Alu it is mainly due to the technical difficulties associated with the detection of the endogenous Alu transcripts produced through Pol-III vs. Pol-II transcription. Alu inserts are enriched within human genes (Lander *et al.*, 2001), and, as a result, they are transcribed as a part of numerous cellular mRNAs. This makes it extremely challenging to delineate authentic Pol-III Alu-promoted transcripts from Pol-II generated transcripts. It is known that, following transcription, many Alu transcripts appear to be processed into small cytoplasmic (scAlu) transcripts, which, based on genomic sequence data, exhibit greatly reduced retrotransposition efficiency (Shaikh *et al.*, 1997). Although a subset of these scAlus can be attributable to the introduction of cryptic Pol-III termination sites, it is unclear whether the remainder are generated by simple degradation or a more active process (Shaikh *et al.*, 1997). Another reported mechanism that acts downstream of transcription but at the unidentified step in the TE life cycle is down regulation of TE retrotransposition by the

APOBEC family of proteins (Chiu *et al.*, 2006; Hulme *et al.*, 2007; Muckenfuss *et al.*, 2006; Stenglein and Harris, 2006). Members of this family of proteins exhibit differential effects on L1 and Alu retrotransposition, highlighting the differences in some yet unknown requirements between mobilizations of the two human retroelements.

The complicated and dynamic nature of the TE–host coexistence is further supported by the observation that artificial upregulation of L1 expression through stable L1 transfection in HeLa cells significantly decreases retrotransposition of both L1 and Alu elements in these cells (Wallace *et al.*, 2010). The phenomenon appears to correlate with the levels of stably expressed L1 ORF2 protein and the presence of the functional L1 EN domain (Wallace *et al.*, 2010). Adaptation of the cellular DNA repair machinery to handle the L1-induced DNA double-strand breaks is implicated as a primary cause of the reduced L1 and Alu retrotransposition in cancer cells overexpressing functional L1. Because of the usage of artificial levels of L1 expression in this experimental system it is not immediately clear what threshold levels of endogenous L1 expression need to be bypassed to institute a similar phenomenon *in vivo*. Nevertheless, these data strongly suggest that the interplay between TEs and their host is likely an ever-changing process and should be viewed as such, particularly when considering TE expression and activity in cells with genetic defects (Morrish *et al.*, 2002) or cells exposed to environmental stimuli known to affect TE activity (Kale *et al.*, 2005, 2006).

It is obvious from the above examples that keeping TE expression at bay is an important priority for the normal cellular existence. This significance is further supported by the data collected in cultured cells that show that ectopic L1 expression causes significant toxicity in both normal and cancer cells (Belancio *et al.*, 2010b; Gasior *et al.*, 2006; Wallace *et al.*, 2008b), indicating that each cell type may have a threshold for tolerable L1 expression. This limit may reflect the ability of cells to deal with the TE-induced genomic insult. As a result, unplanned upregulation of endogenous L1 expression may result in elimination of normal—and even some cancerous—cells through apoptosis (Belgnaoui *et al.*, 2006; Bourc'h and Bestor, 2004) or senescence (Belancio *et al.*, 2010b; Wallace *et al.*, 2008b). Thus, the collective TE expression and activity in any given cell type is likely determined by the combination of the factors established to limit TE expression and by the amount of the TE-associated toxicity that can be endured by these cells.

D. Genetic variation and polymorphism of TE loci

Significant progress has been made in understanding the complexity associated with the fact that TEs are families of elements individual members of which are distributed throughout the genome and often harbor significant sequence variations relative to the rest of the group. Several causes are currently known to contribute to this variation. One of them is infidelity of the L1 RT which is estimated to introduce at least one mutation per every 6000 incorporated nucleotides. This corresponds to the size of the human L1. The biological implication of this finding is that every *de novo* full-length L1 insertion generated by the same active L1 locus has a high likelihood of containing at least one point mutation. Another reason for diversity among family members is the natural accumulation of mutations within any given TE locus due to genomic drift.

Other frequent on-arrival or time-inflicted changes include significant alterations in the length of the polyA tail present at the end of each element. Longer A-tails are associated with younger and more active Alu and L1 elements (Lander *et al.*, 2001; Roy-Engel *et al.*, 2002). However, because of the difference in the generation of this A-tail between L1 and Alu elements, the encoded length of the polyA stretch has a particularly dramatic effect on the rate of Alu retrotransposition (Comeaux *et al.*, 2009; Roy-Engel *et al.*, 2002). An A-tail of an individual element can shrink or grow by 10s of nucleotides during, or shortly following retrotransposition. This translates in the several fold difference in its retrotransposition efficiency among individual insertions (Comeaux *et al.*, 2009; Dewannieux *et al.*, 2003; Roy-Engel *et al.*, 2002). The same consequences for Alu retrotransposition can also arise from mutations interrupting the continuity of the A-tail (Comeaux *et al.*, 2009). The variations generated during transposition along with accumulated random mutations influence the relative activity of individual TE loci (Aleman *et al.*, 2000; Bennett *et al.*, 2008). The disparity in the efficiency of retrotransposition is observed not only among the loci of the same TE family but also between the same TE locus isolated from different individuals (Brouha *et al.*, 2003; Lutz *et al.*, 2003; Seleme *et al.*, 2006). Depending on the presence of modifying mutations, the same insertion locus could be highly active in one individual but quiescent in another. This phenomenon was first reported for L1 elements and is likely to be the case for Alu and SVA elements as well.

Regardless of their origin, mutations within TE sequences can have a dual effect on their activity, as already alluded to above regarding A-tail length. On one hand, these mutations can inactivate or reduce the activity of one of the important L1 protein functions required for retrotransposition of L1 or all three human non-LTR retrotransposons. They can also disrupt other critical elements, such as promoter activity, RNA folding, RNA/protein, protein/protein interactions, etc., each one of which may eliminate or diminish retrotransposition. On rare occasions, however, these mutations may potentially lead to a beneficial change in any of the above mentioned activities or interactions leading to an improvement in retrotransposition. L1 elements with highest retrotransposition potential are referred to as “hot” elements (Brouha *et al.*, 2003). “Hot” Alu and SVA elements are likely present in the human genome as well, but are yet to be characterized.

TE loci in any given human genome can be either fixed or polymorphic in the population. Fixed loci are evolutionary older and are present in all members of the population (excluding rare genomic rearrangement events that can remove a TE locus after fixation took place). Because of their evolutionary age they are more likely to harbor inactivating mutations and, as a result, have lost their activity (Aleman *et al.*, 2000; Lander *et al.*, 2001; Moran *et al.*, 1996). The vast majority of TE loci found in the human genome is represented by the fixed-present elements. Polymorphic TE loci are found only in the subset of individuals in the population and their allele frequencies may vary significantly among the populations of different origin (Cordaux *et al.*, 2007; Wang *et al.*, 2005; Watkins *et al.*, 2003). Polymorphic TE loci are likely to be the most active and they are expected to account for the bulk of TE activity in any given genome (Beck *et al.*, 2010; Brouha *et al.*, 2002, 2003; Ewing and Kazazian, 2010; Huang *et al.*, 2010; Iskow *et al.*, 2010). Initial screening suggested that polymorphic L1s are rare and a very small number of L1 loci are “hot”

(Brouha *et al.*, 2002, 2003). Advances in the whole genome analysis significantly broadened the estimation of the polymorphic L1 loci per genome. In fact, the most recent estimation is that any two human genomes differ on average by 285 sites with respect to presence or absence of L1 insertion (this includes truncated and full-length L1s; Ewing and Kazazian, 2010). Additionally, over half of the identified polymorphic L1s exhibit levels of activity that can be characterized as “hot” (Beck *et al.*, 2010). The observation that there is most likely a group of polymorphic TE loci that are private, that is unique to individual genomes, further expands the TE-associated genetic diversity between individual human genomes (Beck *et al.*, 2010; Ewing and Kazazian, 2010). There is also a significant increase in the L1-associated variation between normal and some cancer genomes (Iskow *et al.*, 2010). These findings strongly support that there is a considerable interindividual genetic variation due to TE presence or absence and as a consequence, considerable interindividual variation in the combined TE activity. These observations combined with the significant tissue-specific variation in the endogenous L1 expression in somatic human tissues (Belancio *et al.*, 2010b) suggest that TE-associated genetic variation is likely to exist among somatic tissues of the same individual.

IV. GENOMIC INSTABILITY

One of the main conflicts between TEs and their host genomes arises from the fact that multiple aspects of TE activity pose a challenge to the structural integrity of the genome. These include L1-induced DSBs, L1, Alu, and SVA retrotransposition and retrotransposition-mediated rearrangements (deletions, insertions, inversions; Gilbert *et al.*, 2002, 2005; Moran *et al.*, 1999; Ostertag and Kazazian, 2001), and TE-associated recombination [reviewed in Hedges and Deininger, 2007] (Fig. 6.2).

A. DSBs

The discovery of DSB generation by the L1 EN (Belgnaoui *et al.*, 2006; Farkash *et al.*, 2006; Gasior *et al.*, 2006) introduced a possibility of the existence of genomic damage that is likely to impose additional strain on the TE/host relationship. The L1 EN that allows for L1, Alu, and SVA mobilization within the human genome introduces DNA strand breakage which, when not properly resolved, can have deleterious consequences on the host genome (Gasior *et al.*, 2006; Wallace *et al.*, 2008b). These breakages can involve one or both strands of the DNA backbone. While single stranded DNA breaks can typically be repaired by the cell without catastrophic consequences, they can, when in combination with a replication fork or some secondary chemical insult, lead to the generation of DSBs. DSBs are known to have detrimental effects on cells [reviewed in van Gent *et al.*, 2001; Vilenchik and Knudson, 2003]. Consequently, organisms have involved an impressive protein-based surveillance system to detect and mark such breaks for repair (Abraham, 2003; Bakkenist and Kastan, 2003; Lieber, 2010). In humans, both H2AX foci staining and comet assays have been used to demonstrate that L1 EN activity results in the formation of DSBs in both normal and cancer cells (Belancio *et al.*, 2010b; Belgnaoui *et al.*, 2006; Farkash *et al.*, 2006; Gasior *et al.*, 2006). Such breaks were observed to occur in excess of the number of successful insertions (Gasior *et al.*, 2006). The accumulation of DSBs can initiate an arrest of the cell cycle while repair is attempted (Gasior *et al.*, 2006). If the damage is too extensive,

apoptosis or senescence will be triggered (Belancio *et al.*, 2010b; Belgnaoui *et al.*, 2006; Gasior *et al.*, 2006; Wallace *et al.*, 2008b). Even though the mutagenic potential of the L1-induced DSBs is not known, they are expected to contribute to genomic instability at the sites of L1 expression that, in addition to germline and tumors (Branciforte and Martin, 1994; Bratthauer and Fanning, 1993; Bratthauer *et al.*, 1994; Martin and Branciforte, 1993; Ostertag *et al.*, 2002), also includes normal human somatic tissues (Belancio *et al.*, 2010b). This assumption is based on the reports that even when dsDNA repair is successful, genomic rearrangements resulting from the repair process can lead to the loss or gain of genomic information and/or the disruption of regulatory signals [reviewed in van Gent *et al.*, 2001; Vilenchik and Knudson, 2003]. Thus, it is likely that L1 EN is responsible for “hidden” L1 damage previously not associated with the activity of these elements. As discussed in more detail in the *TEs and human health* section below, when these effects occur in the germline, they have implications for heritable genetic diseases. When they occur in the soma, there are potential consequences for both cancer and aging.

B. Retrotransposition

While the mutagenic potential of the L1-associated DSBs has not been characterized, the contribution of human TEs to genomic instability through retrotransposition is well established [reviewed in Belancio *et al.*, 2009, 2010a; Cordaux and Batzer, 2009]. All active human retroelements (L1, Alu, and SVA) contributed with different frequencies to the variety of human disease ranging from hemophilia to cancer [reported human diseases are summarized in Belancio *et al.*, 2008a, 2010a]. Despite the continuously growing list of the TE-induced human diseases, the rate of TE-associated mutagenesis, including retrotransposition and recombination, in humans is likely significantly underestimated due to the difficulty in detection of TE-caused mutations by conventionally used diagnostic methods of screening [reviewed in Belancio *et al.*, 2009]. New studies, however, leveraging 2nd generation sequencing are improving our estimates of baseline transposition rates (Beck *et al.*, 2010; Ewing and Kazazian, 2010; Huang *et al.*, 2010; Iskow *et al.*, 2010).

In addition to insertional mutagenesis resulting in disease, retrotransposition of L1, Alu, and most likely SVA elements can have less drastic phenotypic effects on cell survival or function. Integration of these elements within human intronic regions most of the time does not prevent gene expression, but may reduce cell fitness or function under certain stress conditions or in combination with other genomic alterations [reviewed in Kines and Belancio, 2011]. Because L1 and SVA elements contain functional splice and polyadenylation sites (Belancio *et al.*, 2006, 2008b; Hancks *et al.*, 2009; Perepelitsa-Belancio and Deininger, 2003) and Alu elements are prone to acquiring functional splice sites through random mutations long after the integration process is completed. Intronic integration events are known to interfere with expression of genes in which they integrate (Belancio *et al.*, 2006; Han *et al.*, 2004; Lin *et al.*, 2008; Perepelitsa-Belancio and Deininger, 2003; Sorek *et al.*, 2004; Ustyugova *et al.*, 2006; Wheelan *et al.*, 2005). Full-length L1 insertions in forward orientation have the most pronounced effect on gene expression most likely due to the presence of the functional promoter in addition to the highest content of the splice and polyA sites (Boissinot *et al.*, 2001, 2006; Chen *et al.*, 2006; Ustyugova *et al.*, 2006). Intronic Alu integration events can influence alternative splicing

(Sorek *et al.*, 2002) and the presence of inverted Alu sequences within 3'UTRs of genes can alter nuclear retention of mRNAs and strongly represses gene expression (Chen and Carmichael, 2008, 2009; Chen *et al.*, 2008). As discussed below, occasionally, retrotransposition provides genomic rearrangements or cis-acting signals that convey positive benefits to the host organism (Babushok *et al.*, 2007; Xing *et al.*, 2006)

C. Issues in repair and recombination

Another significant contribution of human TEs to genomic instability is through nonallelic homologous recombination (NAHR). Due to their very nature, TEs typically exist in multiple interspersed copies within a given host genome. Depending on their age, the various genomic copies will diverge at the level of nucleotide identity from one another. The longer elements reside in the genome, the more likely that random genetic mutation will level the level of nucleotide identity that they share with their family consensus sequence. The presence of interspersed sequence homology poses a serious challenge for the maintenance of genomic integrity [reviewed in Hedges and Deininger, 2007]; this is particularly the case for higher eukaryotes, such as mammals, that have larger genomes containing higher repetitive element content (Lander *et al.*, 2001). The cellular machinery involved in both recombination and DNA repair can be led astray by homologous sequences present at nonallelic positions (Jasin, 2000). In the case of recombination, this is most often observed in the context of NAHR. The cellular protein machinery that initiates and monitors the strand invasion process cannot always discern interspersed TE sequence homology from truly allelic sequences. As a consequence of NAHR, genetic sequence can be duplicated on one chromosome and lost on the other. In terms of TE/host interactions this result can often have negative fitness consequences due to the loss or alteration of critical genetic information. A significant number of human genetic disorders have resulted from the nonallelic recombination of nearby Alu elements (Callinan and Batzer, 2006; Deininger and Batzer, 1999).

In addition to meiotic recombination, diploid (and higher ploidy) organisms typically rely on the set of homologous chromosomes as templates for the homology-driven (HR) DNA repair. The HR DNA repair system is related to the recombination system and subject to the same limitations in discerning true allelic sequence from interspersed homology. Products of misaligned DNA repair—or repair events in which the only homologous templates available are nearby repetitive sequences—can also generate rearrangements. These rearrangements are, more often than not, intrachromosomal. Nevertheless, this same process can instigate interchromosomal rearrangements (Elliott and Jasin, 2002; Elliott *et al.*, 2005).

While the focus of the deleterious effects of TEs has often centered on the detrimental effects of unchecked insertional mutagenesis, there is evidence that their ability to instigate nonallelic recombination may have an even greater impact on organismal short- and long-term fitness. To date, the larger fraction of Alu-related genetic disease has been observed to arise through mutagenic recombination (Callinan and Batzer, 2006; Deininger and Batzer, 1999; Elliott *et al.*, 2005). Song and Boissinot (2006) demonstrated evidence that negative selection in the human genome principally acts upon TEs as a function of TE length. This suggests that mutagenic NAHR, as opposed to insertional mutagenesis, may have the more

substantial negative consequences that retrotranspositional activity. Similar evidence for the role of ectopic, nonallelic recombination in determining the fitness consequences of individual TE insertions was found in *Drosophila* (Petrov *et al.*, 2003). In addition to causing disease, TE-associated reshuffling of the host genetic material has had a significant impact on the host genome architecture during the course of evolution (Han *et al.*, 2005, 2007, 2008; Konkel and Batzer, 2010; Lee *et al.*, 2008; Xing *et al.*, 2009).

The variety of mutations induced by TEs in the human genome argues for the importance of the multitude of defense mechanisms present within human cells that target almost every step of the TE life cycle. It also underscores the importance of the balanced long-term evolutionary existence between the host genome and TEs.

V. IMPACT ON HUMAN HEALTH

L1 has long been regarded as one of the intrinsic factors contributing to genomic instability; however, its effects, such as insertional mutagenesis and NAHR between interspersed L1 or Alu repeats (L1-, Alu-, and SVA-induced diseases are summarized in Belancio *et al.*, 2008a, 2010a), were considered to be restricted to cancer and mendelian genetic disorders, such as hemophilia A & B. While TE-associated risk to human health is accepted, it is extremely challenging to estimate TE contribution to human disease because of the randomness of the TE-associated mutagenesis, the diversity of mutations arising from TE activity, and the lack of adequate high throughput methods for the identification and analysis of these mutations [reviewed in Belancio *et al.*, 2009]. The discovery of the ongoing endogenous L1 expression in human tissues (Asch *et al.*, 1996; Belancio *et al.*, 2010b; Ergun *et al.*, 2004), L1's ability to damage human DNA not only through retrotransposition and recombination but also via induction of DNA DSBs (Belgnaoui *et al.*, 2006; Farkash *et al.*, 2006; Gasior *et al.*, 2006), and L1-induced toxicity (Gasior *et al.*, 2006; Wallace *et al.*, 2008b) provides the basis for reevaluation of the L1 role in human somatic and germline diseases.

Based on the significant estimated variation in the combined L1 activity in any given human genome (Seleme *et al.*, 2006), there is likely a gradient of endogenous L1 activity in the population and as a result, a spectrum of the L1-associated burden imposed on the host genome (Fig. 6.3). This estimated discrepancy in the L1 activity is associated with the presence of polymorphic L1 loci and a significant variation in the activity of the same L1 locus among individuals on account of mutations that modulate their activity (Beck *et al.*, 2010; Brouha *et al.*, 2003; Huang *et al.*, 2010; Iskow *et al.*, 2010; Seleme *et al.*, 2006). Persons lacking polymorphic L1 elements (that are most likely to be the most active L1 loci in the genome) and containing inactive or the least active fixed L1 loci would likely experience the least damage from the endogenous L1 activity (Fig. 6.3). On the other hand, genomes expressing very active and polymorphic L1 loci would probably be exposed to the highest L1 damage. Even though this assumption is relatively straight forward, and it is supported by experimental evidence, the transition from knowing the spectrum of functional L1 loci in any given genome to predicting their impact on human health is not trivial. The difficulty predominantly arises from the existence of numerous defense mechanisms implemented by the host to combat the insult from TEs (see above sections). A whole new chapter in human TE research has been opened by the discovery of the intimate connection

between the cellular DNA repair machinery and L1-associated DNA damage (Gasior *et al.*, 2006; Gasior *et al.*, 2008; Suzuki *et al.*, 2009). Thus, adequate DNA repair in individual genomes is expected to control L1-associated damage. On the other hand, any decline in the efficacy of the relevant DNA repair pathways, whether due to the loss of function or to the decreased activity associated with certain genotypes, would likely lead to an increased rate of accumulation of the L1-induced DNA damage.

The same assumption is true for a long list of cellular processes, ranging from DNA methylation, transcription, and RNA processing to antiviral defense, reported to control various steps of the L1 life cycle. The existence of these additional dimensions influencing potential L1-associated health risk suggests a scenario under which individuals with the same cumulative endogenous L1 activity may endure drastically different burden from the L1-induced DNA damage. On the other hand, persons with relatively low endogenous L1 activity may have comparable load of the L1-associated mutagenesis to individuals with significantly higher endogenous L1 activity depending on the strength of their anti-TE shield. Thus, the consideration of the status of all the players of the TE-controlling network is necessary for the proper assessment of the TE-associated individual health risk and further thorough evaluation of the complex relationship between TEs and their hosts need to take place.

Inherited defects in any of the pathways controlling TE expression and/ or activity would result in systemic or tissue- or development-specific increase in the TE-associated damage, while acquired somatic mutations are likely to be limited to the cells in which they occurred. Even though little is known about the combined effect of L1 activity with either somatic or germline mutations, it can potentially be a significant contributor to human diseases. For example, high risk of breast cancer associated with inactivating BRCA1 mutations is proposed to be linked the mutagenic repair of DSBs induced by response to estrogen (Fu *et al.*, 2003). Despite the fact that L1 expression and the role of the L1-induced DSBs in mammary gland tumorigenesis remain unexplored, it may be one of the factors contributing to genomic instability during breast cancer development.

A. The potential for L1 to contribute to cancer and aging

The finding of extensive endogenous L1 expression in normal human somatic tissues combined with the L1's ability to induce DSBs not only in human cancer but also normal cells brings a possibility that ongoing endogenous L1 expression can be a continuous source of DNA damage in somatic tissues (Belancio *et al.*, 2010b). DNA damage is known to promote genomic instability, which is one of the contributing factors of the normal aging process and age-associated diseases particularly cancer (Campisi and Vijg, 2009; Coppede and Migliore, 2010; Erol, 2010). One of the accepted theories of mammalian aging is destabilization of the genome through accumulation of DNA damage over the course of life span of an organism (Alexander, 1967). Low levels of endogenous L1 expression in the majority of human somatic tissues and adult stem cells strongly suggests that ongoing L1 expression can contribute to genomic instability over the life span of an organism through retrotransposition of L1, SINEs, and SVA elements, recombination between interspersed copies of L1 and SINEs, and DNA DSBs induced by L1 EN. Endogenous L1 expression

varies significantly among different human tissues and adult stem cells (Belancio *et al.*, 2010b) suggesting tissue-specific variation in the L1-induced DNA damage. Our findings suggest that adult stem cells as well as some human tissues may be protected from the endogenous L1-induced DNA damage relative to other tissues due to the extensive L1 RNA processing that leads to the production of retrotranspositionally defective L1-related mRNAs (Belancio *et al.*, 2010b). Accumulation of the L1-related DNA damage in somatic and adult stem cells with time may result in reduced performance, or, in the case of adult stem cells, in the decreased tissue-renewal capacity and production of differentiated progeny cells that inherit genomic defects accumulated in stem cells (Fig. 6.4). L1-induced DNA damage may also increase with time due to age-related alterations in DNA repair and other cellular pathways that control L1 activity (Barbot *et al.*, 2002; Richardson, 2003; Seluanov *et al.*, 2004; Singhal *et al.*, 1987). Additionally, the significant variation in the endogenous L1 activity estimated to exist in the human population (Seleme *et al.*, 2006; Fig. 6.3) correlates with the remarkable variation in the rate of individual aging, the length of life span, and the onset and severity of the age-associated diseases.

With the exception of an increasing number of evolutionary advantageous examples of acquisition of diverse beneficial functions by TEs within their host genomes, the largely supported view of the TE influence on the genome stability is that of a negative nature. While this view is applicable most of the time to the genome of a single cell, it may only represent one side of the coin when considering the fate of the cells that support TE expression in terms of the overall health of the tissue and organism in which they reside. The discovery of the L1-associated toxicity (Gasior *et al.*, 2006; Wallace *et al.*, 2008b) that manifests itself in the form of apoptosis and cellular senescence in both cancer and normal (Belancio *et al.*, 2010b) human cells suggests that the outcome of TE expression needs to be considered in the context of the specific genomic environment. While TE activity promotes genomic instability through a plethora of mechanisms, TE-associated genomic rearrangements that can lead to the onset or progression of disease may also trigger cell-cycle arrest, followed by elimination of aberrant cell(s) from the population. Based on the reported experimental evidence, deregulated L1 expression, which often happens in response to some external or internal stimuli, in cells with fully or partially functional cell cycle check points can trigger apoptosis or senescence leading to the removal of these cells from further propagation. Interestingly, accumulation of senescent cells in normal tissues with age has been reported (Herbig *et al.*, 2006; Jeyapalan *et al.*, 2007). On the contrary, cells defective in the DNA damage surveillance would escape elimination and continue to exist with accumulated genetic defects. The double-edged-sword hypothesis of L1 expression (Belancio *et al.*, 2010a) implies that L1, through precisely the same types of DNA damage, may play a role in human diseases such as cancer or normal biological processes such as aging.

B. A positive note

The previous section highlighted several avenues by which TE activity can potentially compromise genomic integrity and, as a consequence, human health. At the same time, it is also safe to say that evolution of humans would have taken a dramatically different course—if it could even have been possible at all—were it not for the periodic “intervention” of TE

activity. There are now numerous established examples of TEs being coopted for functional roles across many different taxa. Perhaps most famously, the vertebrate adaptive immune system was made possible, in part, by the invasion of immunoglobulin-coding genes by a RAG transposon (reviewed in Flajnik and Kasahara, 2010). The recombination signal sequences (RSS), introduced into these genes by the transposon conferred the ability for directed somatic recombination, increasing the ability to rapidly generate antibody diversity. A second example is that of the syncytin gene, which is derived from the envelop gene of the human endogenous retrovirus (HERV-W; Mi *et al.*, 2000). Yet another important example of the contribution of TEs to human evolution was the discovery that the SETMAR gene was derived from the fusion of the SET histone methyltransferase gene with a mariner-like Hsmar1 transposon (Cordaux *et al.*, 2006). SETMAR was observed to have inherited several biochemical properties from its transposon source, although its molecular function remains poorly understood (Liu *et al.*, 2007). In addition to the creation of novel genes by the direct contribution of genetic sequence, TEs activity can result in the rearrangement, duplication, and/or reshuffling of existing exons or entire genes (Goodier *et al.*, 2000; Moran *et al.*, 1999; Pickeral *et al.*, 2000). A demonstration of this process in humans was provided by Xing *et al.* (2006), where the authors observed that three copies of the AMAC gene in primates had been generated by L1 transduction (Xing *et al.*, 2006). Several lines of evidence, including intact ORFs and active expression, suggest these newly emerged copies play a functional role in the genome. These are but a few examples from an increasing number of examples of TE domestication events. Across the entire genome, it has been reported that TE sequences are incorporated in as much as 4% of human protein-coding genes (Kim *et al.*, 2010a,b). As our knowledge of the genome increases, demonstrations of TE contributions to organismal biology will continue to mount. There nevertheless remains a large theoretical and empirical leap to be made between the observation of extensive evidence of TE domestication, to the conclusion that TE lineage activity is maintained by natural selection fulfill such an evolutionary role. Once again, it is important to be reminded of Kidwell and Lisch's observation that the role of TEs in the evolution of its host is not at odds with their having an essentially parasitic relationship to the genome. As discussed below, how precisely to characterize that relationship depends, in part, on empirical data that remains to be collected.

VI. CONCLUSIONS

A. The human model

The human organism, in many respects, serves as an exemplary system within which to consider the interplay between host genomes and their TE inhabitants. As detailed above, we find in our own species clear examples of both the positive and negative attributes of TE activity. We have discussed at length the extensive systems the host organism has evolved to curtail TE activity. This serves a strong indicator that TE activity, left unchecked, presents a significant danger to the host organism. We must, however, consider that the unchecked activity of many "functional" biological processes would also be detrimental to the organism. When such processes are themselves curtailed, we are apt to say they are "kept in balance" or "regulated," as opposed to "defended against." The presence of host repression

mechanisms, in and of themselves, is not sufficient to exclude a possible functional role for persistent TE activity, at least among some taxa.

Perhaps what distinguishes TEs from these more essential systems is the ambiguity concerning the extent to which they are components of the “organismal machine,” as opposed to separate entities pursuing their own agenda. As Dawkins illustrated in his *The Selfish Gene* (1976), any gene can be conceived of as pursuing its own agenda of propagation within subsequent generation, potentially at the expense of what might appear to be the wellbeing of its host organism. When the gene happens to be fundamental to a species’ survival, however, we tend to have a more tolerant view towards this sort of scheming. The ability of some TEs to be horizontally transferred across species boundaries strongly suggests that their presence can be nonessential for organismal function. Presumably the host species receiving the horizontally transferred TE was getting along well enough before the TE arrived on the scene. At the same time, we perceive TEs to be diligently pursuing their own agenda of proliferation, occasionally generating negative consequences for the organism. From this perspective, their behavior is more like a parasite, one that just so happens to reside in our own germline. But, as discussed above, TEs have occasionally made positive functional contributions to organismal biology during evolution. The overall frequency of such contributions, however, remains to be determined. Is it sufficiently high enough to pay the rent, so to speak. For further insight, we can look to the study of the human microbiome. The increasing realization that mammals and other multicellular eukaryotes function as super-organisms, comprised of a number of interacting entities exhibiting various degrees of interdependence. In any such system, the line between self and nonself is not so easily delineated. In the case of the human organism, a spectrum emerges wherein, on the one extreme, there is the mitochondria organelle, an entity fully integrated and essential to organismal survival. On the other side are members of the microbial and viral community that give and take to various extents from the entity as a whole. Where along this gradient TEs will ultimately be placed has yet to be determined, as the relative degree of positive and negative contributions to the human organism remain to be elucidated.

B. The road ahead

The direction of TE research in the next decades, as with most scientific endeavors, will be driven largely by technological and methodological improvements. Figuring heavily in these changes will be the availability of relatively inexpensive whole genome sequencing. Low cost, ubiquitous genomic sequencing will allow for better evaluation of “natural”, unmodified, TE activity, decreasing reliance on artificially tagged retrotransposition constructs. While such constructs have been of tremendous importance for mobile element research and will continue to be critical for evaluating certain molecular hypotheses, sequencing-based experiments will be able to address questions surrounding how closely modified TE constructs mirror the biological activity their natural counterparts. We further expect that the sheer volume of sequencing data anticipated over the coming few years will also bring changes to the focus of TE research. While novel TE biology no doubt awaits discovery amidst the multitude of genomes that have yet to be sampled, saturation with TE diversity information will likely lead to a deemphasis of TE discovery and annotation as an

end-in-itself, and the increasing relegation of identification and taxonomical assignment to automated computational pipelines. Emerging technologies will also allow us to peer more deeply into somatic activity of TEs within different organisms, including the mechanisms of their genetic and epigenetic regulation. As discussed above, new evidence suggesting that TE activity in the soma is higher than once believed opens several avenues for exploration in the arena of impacts on organismal fitness and on human health in particular. The ultimate consequences of somatic TE activity on both cancer and aging is currently unknown, and the elucidation of TE impact in these areas will require detailed analysis across diverse tissues. Arguably, the most important findings of the past several years have revolved around the discovery novel cellular regulatory systems for controlling TE activity. These include the regulation of retrotransposition by posttranscriptional processing of TE expression and by members of the APOBEC protein family, the evidence of small RNA pathways playing a role in repressing mammalian retrotransposons, and the involvement of DNA repair in regulation of TE activity. The presence of these systems across diverse taxa speaks to the threat posed by unchecked TE proliferation. Further understanding of these regulatory mechanisms, and how they may interact with gene regulatory networks, will continue to be fertile areas of investigation. In addition, the ability to survey epigenetic markers at a genomic scale, in multiple tissues and developmental stages, will likely provide novel insights into the relationship between interspersed TEs sequences, chromatin structure, and genetic regulation. We anticipate that, despite the scientific advances that await us, no simplistic, pithy description of the host–element relationship will emerge. At best, we only be able to more deeply appreciate the complex relationship dynamic, where cooperation, antagonism, and serendipity each have their roles.

Acknowledgments

VPB is supported by P20RR020152, NIH/NIA 5K01AG030074-02 and The Ellison Medical Foundation New Scholar in Aging award.

References

- Abraham RT. Checkpoint signaling: Epigenetic events sound the DNA strand-breaks alarm to the ATM protein kinase. *Bioessays*. 2003; 25:627–630. [PubMed: 12815717]
- Aleman C, Roy-Engel AM, Shaikh TH, Deininger PL. Cis-acting influences on Alu RNA levels. *Nucleic Acids Res*. 2000; 28:4755–4761. [PubMed: 11095687]
- Alexander P. The role of DNA lesions in the processes leading to aging in mice. *Symp. Soc. Exp. Biol*. 1967; 21:29–50. [PubMed: 4860956]
- Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev*. 2006; 20:210–224. [PubMed: 16418485]
- Alves G, Tatro A, Fanning T. Differential methylation of human LINE-1 retro-transposons in malignant cells. *Gene*. 1996; 176:39–44. [PubMed: 8918229]
- An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD. Active retrotransposition by a synthetic L1 element in mice. *Proc. Natl. Acad. Sci. USA*. 2006; 103:18662–18667. [PubMed: 17124176]
- An W, Han JS, Schrum CM, Maitra A, Koentgen F, Boeke JD. Conditional activation of a single-copy L1 transgene in mice by Cre. *Genesis*. 2008; 46:373–383. [PubMed: 18615728]
- Athanikar JN, Badge RM, Moran JV. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res*. 2004; 32:3846–3855. [PubMed: 15272086]

- Asch AL, Eliacin E, Fanning TG, Connolly JL, Bratthauer G, Asch BB. Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* 1996; 8(6): 239–247. [PubMed: 8895199]
- Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr. L1 integration in a transgenic mouse model. *Genome Res.* 2006; 16:240–250. [PubMed: 16365384]
- Babushok DV, Ohshima K, Ostertag EM, Chen X, Wang Y, Mandal PK, Okada N, Abrams CS, Kazazian HH Jr. A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Res.* 2007; 17:1129–1138. [PubMed: 17623810]
- Bakkenist CJ, Kastan MB. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature.* 2003; 421:499–506. [PubMed: 12556884]
- Barbot W, Dupressoir A, Lazar V, Heidmann T. Epigenetic regulation of an IAP retrotransposon in the aging mouse: Progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res.* 2002; 30:2365–2373. [PubMed: 12034823]
- Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV. LINE-1 retrotransposition activity in human genomes. *Cell.* 2010; 141:1159–1170. [PubMed: 20602998]
- Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.* 2006; 34:1512–1521. [PubMed: 16554555]
- Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Res.* 2008a; 18:343–358. [PubMed: 18256243]
- Belancio VP, Roy-Engel AM, Deininger P. The impact of multiple splice sites in human L1 elements. *Gene.* 2008b; 411:38–45. [PubMed: 18261861]
- Belancio VP, Deininger PL, Roy-Engel AM. LINE dancing in the human genome: Transposable elements and disease. *Genome Med.* 2009; 1:97. [PubMed: 19863772]
- Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know 'bout retroelements in cancer. *Semin. Cancer Biol.* 2010a; 20(4):200–210. [PubMed: 20600922]
- Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* 2010b; 38(12):3909–3922. [PubMed: 20215437]
- Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.* 2006; 6:13. [PubMed: 16670018]
- Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE. Active Alu retrotransposons in the human genome. *Genome Res.* 2008; 18:1875–1883. [PubMed: 18836035]
- Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci. USA.* 2006; 103:8780–8785. [PubMed: 16728505]
- Boissinot S, Furano AV. Adaptive evolution in LINE-1 retrotransposons. *Mol. Biol. Evol.* 2001; 18:2186–2194. [PubMed: 11719568]
- Boissinot S, Entezam A, Furano AV. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* 2001; 18:926–935. [PubMed: 11371580]
- Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl. Acad. Sci. USA.* 2006; 103:9590–9594. [PubMed: 16766655]
- Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature.* 2004; 431:96–99. [PubMed: 15318244]
- Bowen NJ, Jordan IK. Exaptation of protein coding sequences from transposable elements. *Genome Dyn.* 2007; 3:147–162. [PubMed: 18753790]
- Branciforte D, Martin SL. Developmental and cell type specificity of LINE-1 expression in mouse testis: Implications for transposition. *Mol. Cell. Biol.* 1994; 14:2584–2592. [PubMed: 8139560]
- Bratthauer GL, Fanning TG. Active line-1 retrotransposons in human testicular cancer. *Oncogene.* 1992; 7:507–510. [PubMed: 1312702]
- Bratthauer GL, Fanning TG. Line-1 retrotransposon expression in pediatric germ-cell tumors. *Cancer.* 1993; 71:2383–2386. [PubMed: 8384068]
- Bratthauer GL, Cardiff RD, Fanning TG. Expression of LINE-1 retrotransposons in human breast cancer. *Cancer.* 1994; 73:2333–2336. [PubMed: 8168038]

- Brookfield JFY. The ecology of the genome—Mobile DNA elements and their hosts. *Nat. Rev. Genet.* 2005; 6:128–136. [PubMed: 15640810]
- Brouha B, Badge R, Schustak J, Moran J, Kazazian H. Active L1 retrotransposons in the human genome. *Am. J. Hum. Genet.* 2002; 71:410.
- Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA.* 2003; 100:5280–5285. [PubMed: 12682288]
- Callinan PA, Batzer MA. Retrotransposable elements and human disease. *Genome Dyn.* 2006; 1:104–115. [PubMed: 18724056]
- Campisi J, Vijg J. Does damage to DNA and other macromolecules play a role in aging? If so, how? *J. Gerontol. A Biol. Sci. Med. Sci.* 2009; 64:175–178. [PubMed: 19228786]
- Cantrell MA, Scott L, Brown CJ, Martinez AR, Wichman HA. Loss of LINE-1 activity in the megabats. *Genetics.* 2008; 178:393–404. [PubMed: 18202382]
- Casavant NC, Scott L, Cantrell MA, Wiggins LE, Baker RJ, Wichman HA. The end of the LINE?: Lack of recent L1 activity in a group of South American rodents. *Genetics.* 2000; 154:1809–1817. [PubMed: 10747071]
- Chen LL, Carmichael GG. Gene regulation by SINES and inosines: Biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle.* 2008; 7:3294–3301. [PubMed: 18948735]
- Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: Functional role of a nuclear noncoding RNA. *Mol. Cell.* 2009; 35:467–478. [PubMed: 19716791]
- Chen J, Rattner A, Nathans J. Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: Lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements. *Hum. Mol. Genet.* 2006; 15:2146–2156. [PubMed: 16723373]
- Chen LL, DeCerro JN, Carmichael GG. Alu element-mediated gene silencing. *EMBO J.* 2008; 27:1694–1705. [PubMed: 18497743]
- Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, Heidmann T, Greene WC. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc. Natl. Acad. Sci. USA.* 2006; 103:15588–15593. [PubMed: 17030807]
- Christensen SM, Eickbush TH. R2 target-primed reverse transcription: Ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol. Cell. Biol.* 2005; 25:6617–6628. [PubMed: 16024797]
- Clements AP, Singer MF. The human LINE-1 reverse transcriptase: Effect of deletions outside the common reverse transcriptase domain. *Nucleic Acids Res.* 1998; 26:3528–3535. [PubMed: 9671814]
- Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: A critical assessment. *Gene.* 2009; 448:105–114. [PubMed: 19577618]
- Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. Diverse cis factors controlling Alu retrotransposition: What causes Alu elements to die? *Genome Res.* 2009; 19:545–555. [PubMed: 19273617]
- Coppede F, Migliore L. DNA repair in premature aging disorders and neurodegeneration. *Curr. Aging Sci.* 2010; 3:3–19. [PubMed: 20298165]
- Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* 2009; 10:691–703. [PubMed: 19763152]
- Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl. Acad. Sci. USA.* 2006; 103:8101–8106. [PubMed: 16672366]
- Cordaux R, Srikanta D, Lee J, Stoneking M, Batzer MA. In search of polymorphic Alu insertions with restricted geographic distributions. *Genomics.* 2007; 90:154–158. [PubMed: 17498924]
- Cost GJ, Boeke JD. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry.* 1998; 37:18081–18093. [PubMed: 9922177]

- Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 2002; 21:5899–5910. [PubMed: 12411507]
- Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. *Nature.* 2009; 460:1127–1131. [PubMed: 19657334]
- Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, Batzer MA, Lower R, Schumann GG. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res.* 2009; 19:1992–2008. [PubMed: 19652014]
- Deininger PL, Batzer MA. Alu repeats and human disease. *Mol. Genet. Metab.* 1999; 67:183–193. [PubMed: 10381326]
- Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 2003; 35:41–48. [PubMed: 12897783]
- Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN. Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol. Cell. Biol.* 2007; 27:4685–4697. [PubMed: 17470553]
- Dooner HK, Weil CF. Give-and-take: Interactions between DNA transposons and their host plant genomes. *Curr. Opin. Genet. Dev.* 2007; 17:486–492. [PubMed: 17919898]
- Edgar RC, Myers EW. PILER: Identification and classification of genomic repeats. *Bioinformatics.* 2005; 21(Suppl. 1):i152–i158. [PubMed: 15961452]
- Elliott B, Jasin M. Double-strand breaks and translocations in cancer. *Cell. Mol. Life Sci.* 2002; 59:373–385. [PubMed: 11915950]
- Elliott B, Richardson C, Jasin M. Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. *Mol. Cell.* 2005; 17:885–894. [PubMed: 15780943]
- Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Stratling WH, Schumann GG. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J. Biol. Chem.* 2004; 279:27753–27763. [PubMed: 15056671]
- Erol A. Systemic DNA damage response and metabolic syndrome as a premalignant state. *Curr. Mol. Med.* 2010; 10:321–334. [PubMed: 20334625]
- Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* 2010
- Farkash EA, Kao GD, Horman SR, Prak ET. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res.* 2006; 34:1196–1204. [PubMed: 16507671]
- Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell.* 1996; 87:905–916. [PubMed: 8945517]
- Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 2007; 41:331–368. [PubMed: 18076328]
- Finnegan DJ. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 1989; 5:103–107. [PubMed: 2543105]
- Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nat. Rev. Genet.* 2010; 11:47–59. [PubMed: 19997068]
- Flori AR, Lower R, Schmitz-Drager BJ, Schulz WA. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. *Br. J. Cancer.* 1999; 80:1312–1321. [PubMed: 10424731]
- Fu YP, Yu JC, Cheng TC, Lou MA, Hsu GC, Wu CY, Chen ST, Wu HS, Wu PE, Shen CY. Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: A multigenic study on cancer susceptibility. *Cancer Res.* 2003; 63:2440–2446. [PubMed: 12750264]
- Fuhrman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP. Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res.* 1981; 9:6439–6456. [PubMed: 6275362]
- Furano AV, Duvernell DD, Boissinot S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* 2004; 20:9–14. [PubMed: 14698614]

- Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* 2006; 357:1383–1393. [PubMed: 16490214]
- Gasior SL, Preston G, Hedges DJ, Gilbert N, Moran JV, Deininger PL. Characterization of pre-insertion loci of de novo L1 insertions. *Gene.* 2007; 390:190–198. [PubMed: 17067767]
- Gasior SL, Roy-Engel AM, Deininger PL. ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst.).* 2008; 7:983–989. [PubMed: 18396111]
- Gilbert N, Lutz-Prigge S, Moran JV. Genomic deletions created upon LINE-1 retrotransposition. *Cell.* 2002; 110:315–325. [PubMed: 12176319]
- Gilbert N, Lutz S, Morrish TA, Moran JV. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* 2005; 25:7780–7795. [PubMed: 16107723]
- Goodier JL, Ostertag EM, Kazazian HH Jr. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 2000; 9:653–657. [PubMed: 10699189]
- Goodier JL, Ostertag EM, Engleka KA, Seleme MC, Kazazian HH Jr. A potential role for the nucleolus in L1 retrotransposition. *Hum. Mol. Genet.* 2004; 13:1041–1048. [PubMed: 15028673]
- Goodier JL, Zhang L, Vetter MR, Kazazian HH Jr. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.* 2007; 27:6469–6483. [PubMed: 17562864]
- Han JS, Boeke JD. A highly active synthetic mammalian retrotransposon. *Nature.* 2004; 429:314–318. [PubMed: 15152256]
- Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature.* 2004; 429:268–274. [PubMed: 15152245]
- Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.* 2005; 33:4040–4052. [PubMed: 16034026]
- Han K, Lee J, Meyer TJ, Wang J, Sen SK, Srikanta D, Liang P, Batzer MA. Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS Genet.* 2007; 3:1939–1949. [PubMed: 17953488]
- Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA. L1 recombination-associated deletions generate human genomic variation. *Proc. Natl. Acad. Sci. USA.* 2008; 105:19366–19371. [PubMed: 19036926]
- Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res.* 2009; 19:1983–1991. [PubMed: 19635844]
- Hata K, Sakaki Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene.* 1997; 189:227–234. [PubMed: 9168132]
- Hedges DJ, Deininger PL. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.* 2007; 616:46–59. [PubMed: 17157332]
- Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW. Developmental differences in methylation of human Alu repeats. *Mol. Cell. Biol.* 1993; 13:4523–4530. [PubMed: 8336699]
- Herbig U, Ferreira M, Condell L, Carey D, Sedivy JM. Cellular senescence in aging primates. *Science.* 2006; 311:1257. [PubMed: 16456035]
- Huang CR, Schneider AM, Lu Y, Niranjana T, Shen P, Robinson MA, Steranka JP, Valle D, Civin CI, Wang T, et al. Mobile interspersed repeats are major structural variants in the human genome. *Cell.* 2010; 141:1171–1182. [PubMed: 20602999]
- Hulme AE, Bogerd HP, Cullen BR, Moran JV. Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene.* 2007; 390:199–205. [PubMed: 17079095]
- Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell.* 2010; 141:1253–1261. [PubMed: 20603005]
- Jasin M. Chromosome breaks and genomic instability. *Cancer Invest.* 2000; 18:78–86. [PubMed: 10701370]
- Jeyapalan JC, Ferreira M, Sedivy JM, Herbig U. Accumulation of senescent cells in mitotic tissue of aging primates. *Mech. Ageing Dev.* 2007; 128:36–44. [PubMed: 17116315]

- Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA*. 1997; 94:1872–1877. [PubMed: 9050872]
- Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR-a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.* 1996; 20:119–121. [PubMed: 8867843]
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 2005; 110:462–467. [PubMed: 16093699]
- Kale SP, Moore L, Deininger PL, Roy-Engel AM. Heavy metals stimulate human LINE-1 retrotransposition. *Int. J. Environ. Res. Public Health*. 2005; 2:14–23. [PubMed: 16705797]
- Kale SP, Carmichael MC, Harris K, Roy-Engel AM. The L1 retrotranspositional stimulation by particulate and soluble cadmium exposure is independent of the generation of DNA breaks. *Int. J. Environ. Res. Public Health*. 2006; 3:121–128. [PubMed: 16823085]
- Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH Jr. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 2009; 23:1303–1312. [PubMed: 19487571]
- Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. USA*. 1997; 94:7704–7711. [PubMed: 9223252]
- Kim DS, Huh JW, Kim YH, Park SJ, Chang KT. Functional impact of transposable elements using bioinformatic analysis and a comparative genomic approach. *Mol. Cells*. 2010a; 30:77–87. [PubMed: 20652499]
- Kim DS, Huh JW, Kim YH, Park SJ, Kim HS, Chang KT. Bioinformatic analysis of TE-spliced new exons within human, mouse and zebrafish genomes. *Genomics*. 2010b; 96(5):266–271. [PubMed: 20728532]
- Kines KJ, Belancio VP. Expressing genes do not forget their LINEs; interference of transposable elements with mammalian gene expression. *Front. Biosci.* 2011 in press.
- Kolosha VO, Martin SL. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl. Acad. Sci. USA*. 1997; 94:10155–10160. [PubMed: 9294179]
- Kolosha VO, Martin SL. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J. Biol. Chem.* 2003; 278:8112–8117. [PubMed: 12506113]
- Konkel MK, Batzer MA. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Semin. Cancer Biol.* 2010; 20(4):211–221. [PubMed: 20307669]
- Kordis D, Lovsin N, Gubensek F. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. *Syst. Biol.* 2006; 55:886–901. [PubMed: 17345671]
- Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM. The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet.* 2009; 5:e1000458. [PubMed: 19390602]
- Kubo S, Seleme MC, Soifer HS, Perez JL, Moran JV, Kazazian HH Jr, Kasahara N. L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci. USA*. 2006; 103:8036–8041. [PubMed: 16698926]
- Kulpa DA, Moran JV. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* 2005; 14:3237–3248. [PubMed: 16183655]
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* 2008; 22:908–917. [PubMed: 18381894]
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]

- Lavie L, Maldener E, Brouha B, Meese EU, Mayer J. The human L1 promoter: Variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.* 2004; 14:2253–2260. [PubMed: 15520289]
- Lee J, Han K, Meyer TJ, Kim HS, Batzer MA. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE.* 2008; 3:e4047. [PubMed: 19112500]
- Lees-Murdock DJ, De Felici M, Walsh CP. Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics.* 2003; 82:230–237. [PubMed: 12837272]
- Leibold DM, Swergold GD, Singer MF, Thayer RE, Dombroski BA, Fanning TG. Translation of LINE-1 DNA elements in vitro and in human cells. *Proc. Natl. Acad. Sci. USA.* 1990; 87:6990–6994. [PubMed: 1698287]
- Lev-Maor G, Sorek R, Shomron N, Ast G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science.* 2003; 300:1288–1291. [PubMed: 12764196]
- Li R, Ye J, Li S, Wang J, Han Y, Ye C, Wang J, Yang H, Yu J, Wong GK, et al. ReAS: Recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun. *PLoS Comput. Biol.* 2005; 1:e43.
- Li PW, Li J, Timmerman SL, Krushel LA, Martin SL. The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: Implications for retrotransposition. *Nucleic Acids Res.* 2006; 34:853–864. [PubMed: 16464823]
- Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 2010; 79:181–211. [PubMed: 20192759]
- Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL, Xing Y. Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet.* 2008; 4:e1000225. [PubMed: 18841251]
- Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell.* 2009; 139:1069–1083. [PubMed: 19962179]
- Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y, Chalmers R. The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. *Mol. Cell. Biol.* 2007; 27:1125–1132. [PubMed: 17130240]
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell.* 1993; 72:595–605. [PubMed: 7679954]
- Lutz SM, Vincent BJ, Kazazian HH Jr, Batzer MA, Moran JV. Allelic heterogeneity in LINE-1 retrotransposition activity. *Am. J. Hum. Genet.* 2003; 73:1431–1437. [PubMed: 14610717]
- Lynch, M. *The Origins of Genome Architecture.* Sunderland, MA: Sinauer Associates, Inc; 2007.
- Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLoS Genet.* 2006; 2:e2. [PubMed: 16440055]
- Martin SL. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 1991; 11:4804–4807. [PubMed: 1715025]
- Martin SL, Branciforte D. Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 1993; 13:5383–5392. [PubMed: 8395003]
- Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* 2001; 21:467–475. [PubMed: 11134335]
- Martin SL, Li J, Epperson LE, Lieberman B. Functional reverse transcriptases encoded by A-type mouse LINE-1: Defining the minimal domain by deletion analysis. *Gene.* 1998; 215:69–75. [PubMed: 9666081]
- Martin SL, Li J, Weisz JA. Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. *J. Mol. Biol.* 2000; 304:11–20. [PubMed: 11071806]
- Martin SL, Cruceanu M, Branciforte D, Wai-Lun LP, Kwok SC, Hodges RS, Williams MC. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.* 2005; 348:549–561. [PubMed: 15826653]

- Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science*. 1991; 254:1808–1810. [PubMed: 1722352]
- McCLINTOCK B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA*. 1950; 36:344–355. [PubMed: 15430309]
- McMillan JP, Singer MF. Translation of the human LINE-1 element, L1Hs. *Proc. Natl. Acad. Sci. USA*. 1993; 90:11533–11537. [PubMed: 8265584]
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*. 2000; 403:785–789. [PubMed: 10693809]
- Morales JF, Snow ET, Murnane JP. Environmental factors affecting transcription of the human L1 retrotransposon. I. Steroid hormone-like agents. *Mutagenesis*. 2002; 17:193–200. [PubMed: 11971989]
- Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. High frequency retrotransposition in cultured mammalian cells. *Cell*. 1996; 87:917–927. [PubMed: 8945518]
- Moran JV, DeBerardinis RJ, Kazazian HH Jr. Exon shuffling by L1 retrotransposition. *Science*. 1999; 283:1530–1534. [PubMed: 10066175]
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat. Genet*. 2005; 37:997–1002. [PubMed: 16056225]
- Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retro-transposition. *Nat. Genet*. 2002; 31:159–165. [PubMed: 12006980]
- Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C. APOBEC3 Proteins Inhibit Human LINE-1 Retro-transposition. *J. Biol. Chem*. 2006; 281:22161–22172. [PubMed: 16735504]
- Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. 2005; 435:903–910. [PubMed: 15959507]
- Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol*. 2003; 4:R74. [PubMed: 14611660]
- Ostertag EM, Kazazian HH. Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res*. 2001; 11:2059–2065. [PubMed: 11731496]
- Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH. A mouse model of human L1 retrotransposition. *Nat. Genet*. 2002; 32:655–660. [PubMed: 12415270]
- Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet*. 2003; 73:1444–1451. [PubMed: 14628287]
- Pace JK, Feschotte C. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res*. 2007; 17:422–432. [PubMed: 17339369]
- Penzkofer T, Dandekar T, Zemojtel T. L1Base: From functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res*. 2005; 33:D498–D500. [PubMed: 15608246]
- Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet*. 2003; 35:363–366. [PubMed: 14625551]
- Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. Size matters: Non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol. Biol. Evol*. 2003; 20:880–892. [PubMed: 12716993]
- Pickeral OK, Makalowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res*. 2000; 10:411–415. [PubMed: 10779482]
- Piskareva O, Schmatchenko V. DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. *FEBS Lett*. 2006; 580:661–668. [PubMed: 16412437]
- Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*. 2005; 21(Suppl. 1):i351–i358. [PubMed: 15961478]

- Pritham EJ, Feschotte C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc. Natl. Acad. Sci. USA*. 2007; 104:1895–1900. [PubMed: 17261799]
- Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* 2005; 1:166–175. [PubMed: 16110336]
- Richardson B. Impact of aging on DNA methylation. *Ageing Res. Rev.* 2003; 2:245–261. [PubMed: 12726774]
- Roy AM, West NC, Rao A, Adhikari P, Aleman C, Barnes AP, Deininger PL. Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* 2000; 302:17–25. [PubMed: 10964558]
- Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL. Active Alu element “A-tails”: Size does matter. *Genome Res.* 2002; 12:1333–1344. [PubMed: 12213770]
- Seleme MC, Vetter MR, Cordaux R, Bastone L, Batzer MA, Kazazian HH Jr. Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc. Natl. Acad. Sci. USA*. 2006; 103:6611–6616. [PubMed: 16618923]
- Seluanov A, Mittelman D, Pereira-Smith OM, Wilson JH, Gorbunova V. DNA end joining becomes less efficient and more error-prone during cellular senescence. *Proc. Natl. Acad. Sci. USA*. 2004; 101:7624–7629. [PubMed: 15123826]
- Shaikh TH, Roy AM, Kim J, Batzer MA, Deininger PL. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* 1997; 271:222–234. [PubMed: 9268654]
- Singhal RP, Mays-Hoopers LL, Eichhorn GL. DNA methylation in aging of mice. *Mech. Ageing Dev.* 1987; 41:199–210. [PubMed: 3431172]
- Song M, Boissinot S. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene*. 2006; 390(1–2):206–213. [PubMed: 17134851]
- Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. *Genome Res.* 2002; 12:1060–1067. [PubMed: 12097342]
- Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D, Ast G. Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. *Mol. Cell*. 2004; 14:221–231. [PubMed: 15099521]
- Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 2001; 21:1973–1985. [PubMed: 11238933]
- Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.* 2006; 281:16837–16841. [PubMed: 16648136]
- Suzuki J, Yamaguchi K, Kajikawa M, Ichianagi K, Adachi N, Koyama H, Takeda S, Okada N. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.* 2009; 5:e1000461. [PubMed: 19390601]
- Sverdlov ED. Retroviruses and primate evolution. *Bioessays*. 2000; 22:161–171. [PubMed: 10655035]
- Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 1990; 10:6718–6729. [PubMed: 1701022]
- Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell*. 2002; 110:327–338. [PubMed: 12176320]
- Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. Molecular archeology of L1 insertions in the human genome. *Genome Biol.* 2002; 3 research0052.
- Takai D, Yagi Y, Habib N, Sugimura T, Ushijima T. Hypomethylation of LINE1 retrotransposon in human hepatocellular carcinomas, but not in surrounding liver cirrhosis. *Jpn. J. Clin. Oncol.* 2000; 30:306–309. [PubMed: 11007163]
- Tchenio T, Casella JF, Heidmann T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* 2000; 28:411–415. [PubMed: 10606637]
- Thayer RE, Singer MF, Fanning T. Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene*. 1993; 133:273–277. [PubMed: 7693554]

- Tsutsumi Y. Hypomethylation of the retrotransposon LINE-1 in malignancy. *Jpn. J. Clin. Oncol.* 2000; 30:289–290. [PubMed: 11007159]
- Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* 2001; 11:1531–1535. [PubMed: 11591322]
- Ustyugova SV, Lebedev YB, Sverdlov ED. Long L1 insertions in human gene introns specifically reduce the content of corresponding primary transcripts. *Genetica.* 2006; 128:261–272. [PubMed: 17028956]
- Van Duyne, G. Mobile DNA II. In: Craig, NL.; Craigie, R.; Gellert, M.; Lambowitz, AM., editors. In A structural view of tyrosine recombinase site-specific recombination. Washington, DC: ASM Press; 2002.
- van Gent DC, Hoeijmakers JH, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* 2001; 2:196–206. [PubMed: 11256071]
- Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. USA.* 2003; 100:12871–12876. [PubMed: 14566050]
- Volfovsky N, Haas BJ, Salzberg SL. A clustering method for repeat analysis in DNA sequences. *Genome Biol.* 2001; 2 RESEARCH0027.
- Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM. LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene.* 2008a; 419:1–6. [PubMed: 18534786]
- Wallace NA, Belancio VP, Deininger PL. L1 mobile element expression causes multiple types of toxicity. *Gene.* 2008b; 419:75–81. [PubMed: 18555620]
- Wallace NA, Belancio VP, Faber Z, Deininger P. Feedback inhibition of L1 and Alu retrotransposition through altered double strand break repair kinetics. *Mob. DNA.* 2010; 1:22. [PubMed: 20979631]
- Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA elements: A hominid-specific retroposon family. *J. Mol. Biol.* 2005; 354:994–1007. [PubMed: 16288912]
- Wang J, Song L, Grover D, Azrak S, Batzer MA, Liang P. dbRIP: A highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* 2006; 27:323–329. [PubMed: 16511833]
- Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, et al. Genetic variation among world populations: Inferences from 100 Alu insertion polymorphisms. *Genome Res.* 2003; 13:1607–1618. [PubMed: 12805277]
- Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 2001; 21:1429–1439. [PubMed: 11158327]
- Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: A new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* 2005; 15:1073–1078. [PubMed: 16024818]
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 2007; 8:973–982. [PubMed: 17984973]
- Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl. Acad. Sci. USA.* 2006; 103:17608–17613. [PubMed: 17101974]
- Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, et al. Mobile elements create structural variation: Analysis of a complete human genome. *Genome Res.* 2009; 19(9):1516–1526. [PubMed: 19439515]
- Xiong Y, Eickbush TH. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* 1988; 5:675–690. [PubMed: 2464735]
- Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 1990; 9:3353–3362. [PubMed: 1698615]
- Yang N, Zhang L, Zhang Y, Kazazian HH. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* 2003; 31:4929–4940. [PubMed: 12907736]

Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 1997; 13:335–340. [PubMed: 9260521]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 6.1.

The Structure of currently active human TE Lineages. The basic structure of the currently active human TE lineages is represented. The Pol-II transcribed L1 sequence contains an internal bidirectional promoter and encodes a bicistronic transcript that contains two proteins (ORF1 and ORF2). The Alu element consists of two monomeric units united by an A-rich middle region. The SVA element appears to be a chimeric construct, derived from components of both SINE and endogenous retrovirus elements. The structure of HERV elements closely resembles that of the exogenous retroviruses from which they derived from. They are flanked by two long terminal repeats (LTRs) and retain protein-coding regions that are recognizably related to viral *gag*, *pol*, and *env* genes.

Figure 6.2.

L1-associated genomic instability. Human non-LTR transposable elements can trigger a number of events leading to mutagenic alterations in the host genome. Among them are retrotransposition, DNA double-strand breaks (DSBs), and nonallelic homologous recombination (NAHR). Out of the three human non-LTR TEs only L1 elements can induce all of these events. L1, Alu, and SVA retrotransposition causes insertional mutagenesis when the *de novo* integrations reside within or near human genes or functional regulatory elements. Both retrotransposition and NAHR lead to interference with normal gene expression through alterations in the gene regulation and/or processing. NAHR and L1-associated DSBs have a potential to produce gross genomic rearrangements. In addition, L1-induced DSBs are reported to induce cell-cycle arrest, apoptosis, and senescence-like phenotype in normal and cancer cells.

Figure 6.3. Variation in the endogenous L1 activity in the human population. Due to the significant number of the functional L1 loci in any human genome, the variation in the activity of the same L1 locus among individuals, and the presence of polymorphic L1 loci that are likely to be the most active L1s, there is an estimated 300-fold variation in the L1 activity among individuals in the human population. In a simplified model where there is only one fixed and one polymorphic locus, the lowest predicted endogenous L1 activity would be expected in individuals whose genomes harbor fixed L1 loci with low activity (black bars). Individuals heterozygous (one blue/dark gray bar) or homozygous (two blue/ dark gray bars) for polymorphic low or high “hot” (green/light gray bars) activity L1 loci are likely to experience a gradual increase in their total endogenous L1 activity. A further augmentation in the endogenous L1 activity is expected in persons that contain fixed active L1 loci (red/ white bars) and a combination of these fixed L1s and polymorphic low or high activity L1 loci. This gradient of L1 activity and most likely L1-associated mutagenesis may be one of the contributing factors to the discrepancy in the time of the onset and severity of human diseases associated with genomic instability observed within human population.

Figure 6.4.

Model of L1 involvement in aging and age-associated diseases. A stem cell in the bottom left corner of the diagram represents a young adult stem cell that gives rise to healthy differentiated progeny cells. With time endogenous L1 activity most likely in the form of L1 ORF2 expression from SpORF2 mRNA (Belancio *et al.*, 2010a,b) leads to accumulation of somatic mutations (different color stars in the nucleus) that may result in generation of differentiated progeny cells that carry those mutations, decrease in differentiation potential (vertical dashed arrows), or cell death or malignant transformation. Gradual loss of fitness or ability to perform preprogrammed functions due to accumulation of mutations is reflected by the increased intensity of the gray color in the stem cells with age. At the level of differentiated somatic cells, a healthy replicatively young cell (the very left cell in the middle horizontal row) with time will progress into a cell (top right corner of the diagram) that has accumulated a number of mutations in the cellular DNA due to the endogenous L1 activity via insertional mutagenesis of full-length L1, SINEs, and SVA elements and via integration independent mutagenic activity of L1 and SpORF2 expression through unfaithful repair of the L1-induced DSBs (different color stars in the nucleus). L1 activity in somatic cells may also result in cell death, cellular senescence, or malignant transformation. An acceleration of L1-induced mutagenesis with time may result from either increased L1

expression due to the age-associated hypomethylation of the cellular DNA, decrease in the efficiency of the DNA repair machinery with age, and/or age-associated alterations in the expression and/or function of other cellular factors that may play a role in L1 expression.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript