Minireview

Role of long interspersed nuclear element-1 in the regulation of chromatin landscapes and genome dynamics

Kenneth S Ramos[1], Pasano Bojang[2] and Emma Bowers[1]

[1]Institute of Biosciences and Technology, Texas A&M Health, Houston, TX 77030, USA; [2]University of Kentucky College of Medicine, Lexington, KY 40506, USA

Corresponding author: Kenneth S Ramos. Email: kramos@tamu.edu

Impact statement

L1 is an autonomous mobile element that has played a critical role in shaping the human genome over evolutionary time. Here we synthesize critical insights on various aspects of L1 biology, including its structure, function, regulation, and associated biological processes. L1 dysregulation and uncontrolled expression can have dire consequences for host genomes and can precipitate diseases such as cancer. Thus, having a cohesive understanding of L1 biology may be particularly important in understanding the mechanisms of carcinogenesis, and may provide sources of critical information for future development of novel biomarkers and cancer treatments.

Abstract

LINE-1 retrotransposon, the most active mobile element of the human genome, is subject to tight regulatory control. Stressful environments and disease modify the recruitment of regulatory proteins leading to unregulated activation of LINE-1. The activation of LINE-1 influences genome dynamics through altered chromatin landscapes, insertion mutations, deletions, and modulation of cellular plasticity. To date, LINE-1 retrotransposition has been linked to various cancer types and may in fact underwrite the genetic basis of various other forms of chronic human illness. The occurrence of LINE-1 polymorphisms in the human population may define inter-individual differences in susceptibility to disease. This review is written in honor of Dr Peter Stambrook, a friend and colleague who carried out highly impactful cancer research over many years of professional practice. Dr Stambrook devoted considerable energy to helping others live up to their full potential and to navigate the complexities of professional life. He was an inspirational leader, a strong advocate, a kind mentor, a vocal supporter and cheerleader, and yes, a hard critic and tough friend when needed. His passionate stand on issues, his witty sense of humor, and his love for humanity have left a huge mark in our lives. We hope that that the knowledge summarized here will advance our understanding of the role of LINE-1 in cancer biology and expedite the development of innovative cancer diagnostics and treatments in the ways that Dr Stambrook himself had so passionately envisioned.

Keywords: Chromatin landscapes, genome dynamics, LINE-1 (L1), L1 mobilization

***Experimental Biology and Medicine* 2021; 246: 2082–2097. DOI: 10.1177/15353702211031247**

Introduction

An unexpected outcome of the Human Genome Project was the finding that the human genome contains far fewer genes than originally postulated, with nearly half of the genomic DNA content constituted by repetitive sequences.[1] Some of these sequences encode non-coding RNAs, while others encode proteins that support the functionality of repetitive elements.[2] A number of repetitive sequences can move to new locations and as such are referred to as transposable elements (TEs). TEs have been broadly classified into DNA transposons or retrotransposons. DNA transposons mobilize directly via a "cut-and-paste" mechanism executed by transposase and inverted terminal repeats.[3] In contrast, retrotransposons mobilize via a "copy and paste" mechanism catalyzed by a unique TE-encoded reverse transcriptase (RT).[4] Retrotransposons include long terminal repeat (LTR) retrotransposons,[5] non-LTR retrotransposons, such as the Long Interspersed Element-1 (LINE-1 or L1)[6] and Short Interspersed Elements (SINEs), like the human Alu sequences.[7] Collectively, LINEs, SINEs, LTR retrotransposons, SINE-R/variable-number-of-tandem-repeats (VNTR)/Alu (SVA), and DNA transposons comprise at least 45% of the human genome,[8] with recent estimates suggesting that repetitive sequences may in fact account for up to 69% of the human genome.[9]

Among these elements, L1 is the only autonomous mobile element that remains active in the human genome.[10–12] L1 can copy and paste itself into different genomic locations by converting RNA back to DNA through reverse transcription using an RNA intermediate that is necessary for retrotransposition.[13] Its expression can mediate profound changes of genome architecture and function and transmobilize other TEs.[14,15] Here we review L1 biology and highlight the critical role of L1 sequences in genetics, genomics, and human health. We begin with a discussion of L1 architecture, expression, and function and then review the complex interplay between L1 and chromatin landscapes and genome dynamics. Given the scope and sensitivity of these processes, a host of mechanisms have evolved to carefully coordinate cellular L1 expression. We discuss these regulatory mechanisms and how they can be perturbed by environmental injury and become dysregulated in diseases such as cancer.

L1 architecture, expression, and function

Structural considerations

A functional human L1 element is ~6 kb in length and consists of a 5′-untranslated region (5′-UTR), which serves as an internal promoter, two open reading frames (ORF1 and ORF2), and a 3′ UTR terminating in a poly (A) tail (Figure 1).[16] While the 5′ UTR of human L1 contains a bidirectional internal promoter that is 903 bp long, the mouse consists of ~200 bp sequences (called monomers) which are sequentially repeated two to five times at the 5′ end to support promoter activity in a manner that is proportional to the number of monomers present.[17,18] Transient transcription assays have established that the first 100 bp within the 5′UTR of human L1 contain DNA sequences required for transcription,[19] with deletion of this region associated with a massive reduction of promoter activity, compared to deletions between +98 and +525 which only reduce activity ~30-fold. Noticeably, deletions at the 3′ end of the promoter between +662 to +902 increase transcriptional activity, implicating negative DNA regulatory domains within this region.[19] Sequences upstream of the +1 nucleotide or target site duplications (TSD) varying in length up to 13 nucleotides long do not influence L1 transcription.[20–22] Most L1 insertions are truncated at the 5′ end and carry sequence changes that render newly inserted elements unable to retrotranspose.[23] Genome-wide analysis has identified ~100 human and 3000 mouse retrotransposition competent, full-length L1s.[24–26] These sequences have been compiled and are available at http://line1.molgen.mpg.de.[26]

Transcriptional control

Transcription of L1 is mediated by RNA polymerase II,[27,28] but the element functions independent of a canonical TATA box. The TATA-less L1 promoter contains an initiator (Inr) sequence element that directs transcription initiation at the first nucleotide.[22] Several transcription binding sites have been identified within the 5′ UTR of human L1 (Figure 2). The zing finger transcription factor Yin-Yang-1 (YY1) interacts with a perfect core binding site located immediately after the transcription initiation site of the human promoter (between +13 and +21). YY1 functions in transcription initiation site selection,[27,29] a finding consistent with its ability to drive gene expression from Inr elements, and to directly recruit RNA Polymerase II. Putative DNA binding sequences for transcription factors of the Sry-related high-mobility group (HMG) box (SOX) family of proteins,[30] and the runt-domain transcription factor, RUNX3, have also been identified within the human promoter.[31] Intriguingly, SOX2 represses L1 transcription in neuronal stem cells of the rat hippocampus, with decreases in SOX2 during neuronal differentiation correlating with increases in L1 retrotransposition.[32] RUNX3, a key regulator of TrkC dorsal root ganglia (DRG) neurons,[33] is required for L1 transcription and retrotransposition *in vitro*.[34] Mutation of RUNX3 binding sites decreases L1 transcription in both the sense and antisense directions, implicating RUNX factors not only in regulation of the L1 element, but also of genes lying upstream of L1 and under the influence of the L1antisense promoter (ASP).[34] Lastly, a role for methyl-CpG-binding protein 2 (MeCP2) in transcriptional control of L1 has been established based on its ability to repress L1 in neuronal tissues via recruitment of repressor proteins to methylated CpG loci within the L1 promoter.[35]

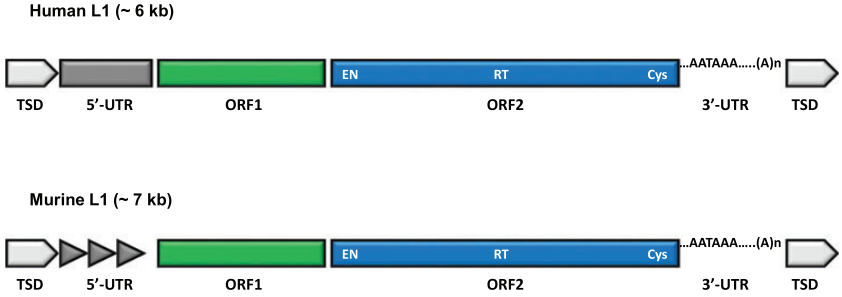

Figure 1. Schematic representation of the full-length human and murine L1. Human L1 is approximately 6 kb long, while murine L1 is approximately 7 kb long. These retroelements are flanked by target site duplications (TSD). Both the 5′-untranslated region (5′-UTR) and the 3′ -UTR of human and mouse are molecularly distinct. Human L1 5′ -UTR is 906 bp long, while the murine L1 5′-UTR is variable in length and contains monomeric units (triangles) approximately 203 bp long that provide the genetic sequences required for transcriptional activation. The monomers are arranged in tandem repeats that confer enhanced transcriptional activation, with increasing number of monomers present within the 5′-UTR. The 3′ UTR contains a polyadenylation signal and a poly(A) tail. However, the length of the 3′-UTR in mice is longer than in humans. The two open reading frames (ORF1 and ORF2) encode p40 protein with RNA-binding and chaperone activity and a protein of approximately 150 kDa with endonuclease (EN) and reverse transcriptase activities (RT). ORF2p also contains a cysteine (Cys)-rich domain near its carboxyl-terminus.

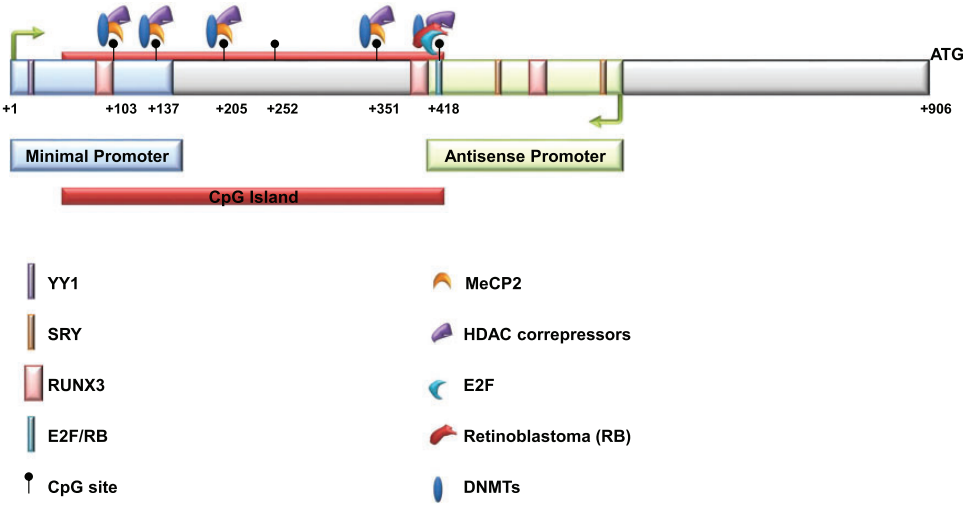

Figure 2. Human L1 promoter regulatory motifs. The human L1 promoter is a TATA-less promoter that begins transcription at bp +1 of the 5′UTR, frequently with the sequence 5′-GGGGG-3′. The 5′-UTR contains a minimal promoter, approximately 155 bp long, and an antisense promoter approximately 200 bp long. Several transcription factors are involved in L1 reactivation and silencing. YY1, SRY-related transcription factor SOX, and RUNX3 have been associated with L1 activation, while E2F/Rb and MeCP2 are believed to interact with the promoter and to direct L1 silencing. RUNX3 exerts strong regulatory control of L1, as denoted by the thickened size of the boxes. L1 silencing is likely orchestrated within the CpG island, which contains several CpG sites that become hypomethylated following exposure to environmental stimuli. It should be noted that not all CpG sites are under identical regulatory control and their influence on L1 expression exhibits cell and contextual specificity. L1 epigenetic silencing also depends on other repressor and corepressor proteins including histone deacetylases (HDACs), DNA methyltransferases (DNMTs), and methyl-binding domain proteins (MBDs). For simplicity, additional mechanisms involving APOBEC proteins, small non-coding RNAs, histone modifications, and epigenetic mechanisms during embryogenesis and germ cell line events have been omitted. The length of the 5′-UTR and the location of each CpG site within the CpG island is numbered below the schematic. ATG corresponds to the start of the ORF1 sequence.

In silico analysis of the L1 5′UTR DNA has identified E2F transcription factor binding sites in both human and murine L1s.[36] In contrast to the human sequence, the E2F site in the murine 5′UTR is located in the antisense orientation. E2F proteins interact with retinoblastoma (RB) to control cell division, differentiation, and senescence, as well as L1 silencing.[36] While both E2F1 and E2F4 bind the L1 promoter, E2F binding does not directly regulate L1 expression and instead, mediates recruitment of RB and associated repressor proteins to the L1 promoter.[36] Given that E2F1 activates transcription, while E2F4 lacks a transactivation domain, the interplay between different E2F factors may dictate key aspects of L1 activity *in vivo.* This prediction is reinforced by the finding that the L1 promoter shows increased histone acetylation and transcriptional activation in the absence of RB proteins (pRB, p107, and p130), coupled with reductions in histone trimethylation marks characteristic of epigenetic silencing.[36] Moreover, the inactivation of RB proteins is associated with the loss of repressive marks and increased expression of L1,[36] consistent with the ability of RB protein family members to recruit corepressors that mediate gene silencing.[37–39] The removal of RB proteins dramatically reduces the localization of histone deacetylase1 (HDAC1) and histone deacetylase2 (HDAC2) within the L1 promoter, suggesting that repetitive elements function as centers of heterochromatin formation[40,41] and that RB proteins function as global regulators of chromatin dynamics.[41,42] Thus, L1s may function as sequence-specific sites for recruitment of protein complexes with heterochromatin nucleation abilities that induce epigenetic silencing and site-specific DNA compaction in response to environmental cues. Of note are previous reports that primary and transformed cells challenged with carcinogenic polycyclic aromatic hydrocarbons (PAHs) activate L1.[18,43–45] Interestingly, the absence of AHR, a highly conserved ligand-activated bHLH transcription factor in mammalian cells that interacts with RB and other transcription factors, prevents reactivation of L1 by PAHs.[45] Whether many of the well-known AHR interacting proteins participate in the regulation of L1 remains to be fully established, though a recent report implicated ARNT (aryl hydrocarbon receptor nuclear translocator) in the regulation of L1. [46]

Translational control

The production of full-length L1 mRNA is highly inefficient due to the A-richness of the RNA, leading to inefficient elongation and premature polyadenylation of the transcript.[47–49] As a result, L1 transcripts often lack one or both of the L1 open reading frames or undergo splicing that precludes synthesis of functional proteins.[50] Full-length L1 transcripts contain the polyadenylation signal AATAA immediately followed by a poly(A) tail (3′ poly-A stretch) but lack the prototypical downstream GU-rich sequence present in mammalian genes. This poly(A) tail, rather than being post-transcriptionally generated by Poly (A) polymerase (PAP1), is encoded within the genomic sequence of L1 itself. Such an unusual structure is believed to ensure that the full-length mRNA contains a poly(A) that serves as a motif during the initial stages of retroelement insertion via target primed reverse transcription (TPRT).[12]

L1 protein functions

The bicistronic L1 transcript serves as a template for the synthesis of two proteins, ORF1p and ORF2p, in the

cytoplasm. ORF1p is a ~40-kDa basic protein with evolutionarily conserved C-terminal and N-terminal coiled-coil domains responsible for multimerization.[51] This protein functions as a non-sequence specific, single-stranded RNA, and DNA-binding protein with chaperone activity essential for retrotransposition.[52–57] ORF2p is a ~150-kDa protein with endonuclease (EN)[58,59] and RT[60,61] activities, and a C-terminal cysteine-rich motif[62] containing a zinc knuckle proposed to mediate interactions with DNA,[12] and to facilitate polymerization of the RT domain.[63] The mobilization of L1 is dependent upon L1-encoded ORF2p, with nicking of DNA at AT-rich sequences within the consensus 5′-TTAAAA-3′/3′-AA↑TTTT-5′ by EN to generate a priming site for reverse transcription of its own RNA transcript.[64] The TPRT model predicts that the poly(A) tail of L1 mRNA interacts with the cleaved DNA to "prime" reverse transcription and integration.[12,65] TPRT-mediated L1 insertion can be inhibited by ERCC/1XPF.[66]

Human ORF1p is believed to recognize the L1 transcript near the 5′ end of ORF2,[52] and to facilitate strand exchanges during DNA priming onto the L1 RNA template.[67] ORF2p is translated separately from ORF1p. Translation of ORF2p is extremely inefficient compared to ORF1p, giving rise to large differences in protein abundance within the cytosolic compartment.[28,68,69] It is estimated that a fully coated L1 mRNA may contain as many as 240 ORF1 proteins, but only a single ORF2.[70] Attempts to define the actual stoichiometric relationship between ORF1 and ORF2 have resulted in estimates ranging between ~6:1–9:1 and 27:1 to 47:1.[71,72]

Steady-state levels of ORF2p are tightly controlled post-translationally, presumably to avoid the toxicity associated with nicking genomic DNA. Indeed, ectopic expression of L1 *in vitro* decreases cell viability and induces double-strand DNA breaks (DSB) and apoptosis.[73–75] L1 transcripts are detected within large cytoplasmic RNPs containing ORF1p[52,76,77] and ORF2p,[76,78] both endogenously in human NTera2D1 and mouse F9 cells, and ectopically in HeLa cells.[52,76–78] The formation of RNPs exhibits cis-preference, a mechanism believed to restrict retrotransposition of other cellular RNAs[65,79] and pseudogene formation.[79] The RNP complex then translocates to the nucleus where the L1 mRNA is reverse transcribed by L1-encoded RT into a cDNA that is integrated into the genome upon nicking of the DNA to complete the L1 life cycle (Figure 3).[80] The dynamics of translocation are strongly influenced by the cell cycle. Mita *et al.* demonstrated that the L1 RNP enters the nucleus at the beginning of mitosis, as the nuclear membrane begins to disintegrate. The L1 RNPs are then retained in the nucleus as the nuclear membrane later reforms. LINE-1 retrotransposition peaks during DNA synthesis, with only LINE-1 RNA and ORF2p being retained within the nucleus.[81] The L1 machinery is responsible for most reverse transcription within the genome,[79,82] as well as the creation of processed pseudogenes.[65,83] L1s are responsible for

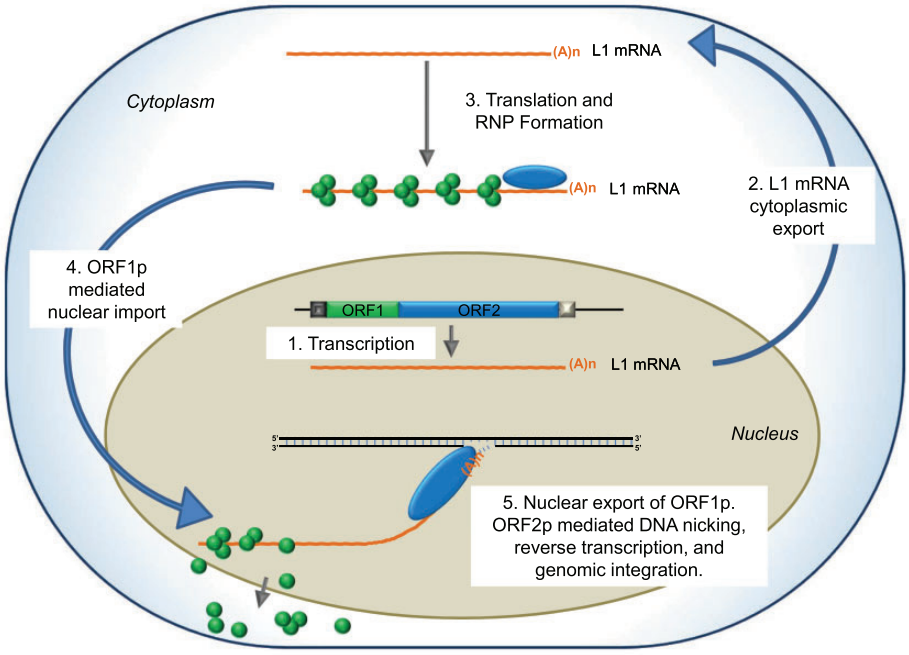

Figure 3. L1 life cycle. The life cycle of a full-length L1 begins with production of a full-length L1 mRNA. In the cytoplasm, L1 proteins are synthesized and they bind in *cis* to L1 mRNA. L1 ORF1p is a trimeric protein that is believed to cover most of the L1 message and ORF2p is hypothesized to bind at a ratio of one molecule per each L1 mRNA. The newly formed ribonucleoprotein particle (RNP) is imported into the nucleus where the endonuclease (EN) domain of ORF2p nicks the DNA to allow the complementary alignment of L1 mRNA poly(A) tail and the nicked DNA, which in turn acts as a primer for cDNA synthesis via the reverse transcriptase domain (RT) domain of ORF2p. During reverse transcription, ORF1 is exported back to the cytoplasm. Once cDNA synthesis is completed, second strand synthesis and integration of L1 into the genome occur.

retrotransposition of non-autonomous SINEs such as human Alu[84] and SVA.[85]

L1 silencing and reactivation

DNA methylation. The presence of retrotransposons in mammalian cells dates back to several hundred million years.[4] Such a long-lived host-parasite relationship has allowed cells to develop a variety of mechanisms to either maintain repetitive elements in a silenced state (host defense hypothesis) or "domesticate" TEs in ways that benefit the host's genetic and adaptive needs (exaptation hypothesis).[86] The latter can be conceptualized as a means to support cellular plasticity[32,87,88] and/or to control genomic regulatory networks.[89,90] These mechanisms are depicted in Figure 4. L1 silencing is paramount to the preservation of genome stability as uncontrolled retrotransposition can be deleterious to the host. Methylation of L1 DNA serves as a primary mechanism to repress genome transcription and induce heterochromatin formation. The characteristic event is the introduction of a methyl group at Carbon 5 in the cytosine ring (5mC). The DNA target sequences are primarily cytosine-phospho-guanidine (CpG) dinucleotides, with CpHpG and CpHpH (where H is A, C or T) also reported as potential targets.[91,92] Of note is the demonstration that mammalian DNA contains two modified cytosine bases: 5mC and 5-hydroxymethylcytosine (5hmC) generated by the action of the Ten-Eleven-Translocation (TET) family of oxygenases.[93] The relative abundance of 5mC and 5hmC is believed to define the dynamics of DNA methylation in many cell types and tissues. Locus-specific methylation of L1 may be differentially influenced depending on where the particular L1 resides within the genome.[94] Moreover, variations of L1 methylation levels in normal cells occur specifically at some loci,[94] suggesting that ectopic expression of L1-regulated genes could be mosaic, similar to the epigenetic influence exerted by the LTR retrotransposon IAP on the coat color in viable yellow agouti (A^{vy}) mouse. In this model, the methylation of the IAP controls the expression of the downstream *Agouti* gene, which influences phenotypic traits such as coat color, obesity, diabetes, and tumorigenesis.[95] The methylation levels of this allele are highly variable across cells and strongly influenced by external factors, such as nutrition and environment.[95]

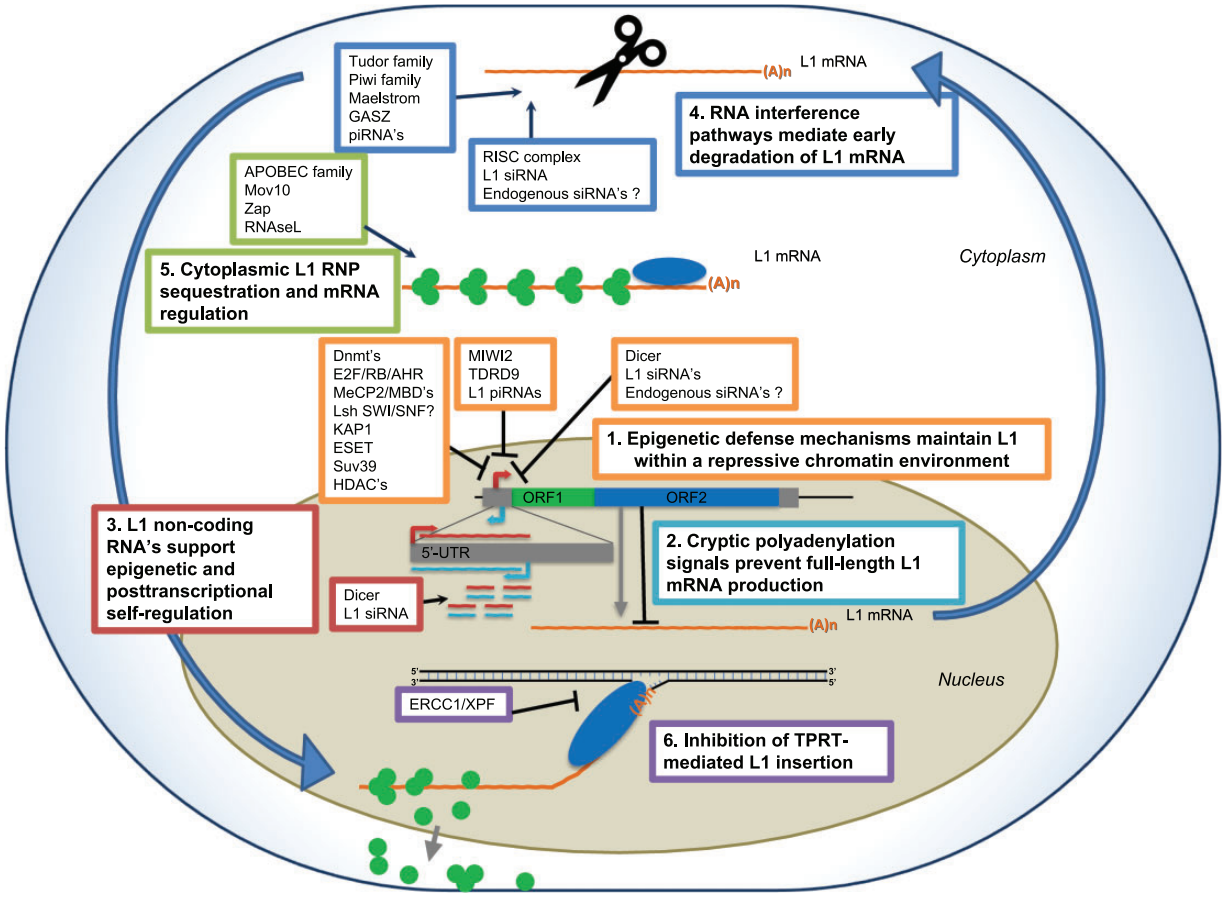

Figure 4. L1 silencing mechanisms. Maintaining L1 silencing is critical for the preservation of genome stability, as uncontrolled L1 expression and retrotransposition can be deleterious to the host genome. As a result, a repertoire of mechanisms have evolved to control L1 at various stages of its lifecycle. 1. Various epigenetic mechanisms suppress L1 transcription by maintaining a repressive chromatin environment (orange). 2. During transcription, cryptic polyadenylation signals inhibit L1 mRNA production (light blue). 3. Non-coding L1 RNAs act to reinforce the formation of repressive epigenetic landscape (red). 4. Proteins involved in microRNA processing facilitate the degradation of L1 mRNA (dark blue). 5. In the cytoplasm, the translation of L1 mRNA into ORF1 and ORF2 can be inhibited, the L1 ribonucleoprotein (RNP) can be sequestered within stress granules. Proteins such as MOV10, UPF1, ZCCHC3, and ZAP1 regulate L1 RNA levels and in the case of UPF1 may also participate in re-integration. L1 mRNA can be degraded, and APOBEC can degrade the protein (green). 6. TPRT-mediated L1 insertion can be inhibited by ERCC/1XPF (purple). Please refer to Pizarro and Cristofari.

The human L1 promoter is heavily methylated at CpG and non-CpG dinucleotides.[96] Indeed, the amount of L1 ORF1p present is inversely correlated with the degree of methylation of the human L1 5′ UTR.[96,97] A reciprocal relationship holds true, with hypomethylation of the L1 promoter associated with increased L1 expression. CpG loci-targeted hypermethylation at the 5′ UTR mediates recruitment of methyl-CpG binding domain (MBD) protein family members including MBD1, MBD2, MBD3, and MBD4, and the X-linked Methyl-CpG-binding domain protein 2 (MeCP2).[35,98] Interestingly, MeCP2 mutations are associated with Rett syndrome, a neurological disorder that involves deceleration of head growth, developmental regression, and encephalopathy.[99] MeCP2 acts as a transcriptional repressor and plays an important role in silencing of L1 in neurons, with Rett syndrome-derived progenitor neuronal cells showing increased retrotransposition.[32] Previous reports have shown that MeCP2 prevents expression and retrotransposition of L1, but not Alu elements.[35] Since MeCP2 interacts with corepressor protein complexes including, Sin3A, as well as DNMT1 and HDAC1, a scenario can be envisioned in which methylation of DNA followed by binding of MeCP2 drives the recruitment of corepressor complexes that enhance DNA methylation of CpG loci along with histone-mediated epigenetic silencing of L1. Other evidence suggests that MeCP2 could function as a transcriptional activator when bound to 5-hydroxy methyl cytosine (5-hmC),[99] further supporting the view that relative 5mC and 5hmC are critical to cellular function. Interestingly, the MeCP2 R133C mutant found in Rett-syndrome patients preferentially inhibits 5-hmC binding.[99] In light of these findings, variations in 5-hmC content of the L1 promoter during development may serve as a mechanism to regulate binding of MeCP2 and associated cofactors to define heterochromatin signatures. Since L1 activity may mediate brain plasticity,[32] MeCP2 may act as a regulator of the kinetics of L1 reactivation. Systematic studies are needed to identify fluctuations between 5-hmC and 5-mC, MeCP2 binding to L1 promoter, and plasticity during development.

Members of the DNA methyltransferases (DNMTs) also play important roles in epigenetic silencing and reactivation of retroelements. We reported earlier that epigenetic reactivation of L1 by DNA-damaging agents involves proteasomal-mediated degradation of DNMT1.[98] Further, genetic silencing of DNMT1 increases the abundance of L1 message in human cervical tumor cells.[98]

Piwi-interacting RNAs. The methylation of retrotransposons also involves small RNAs specific to the germ cell lineage, so called, Piwi-interacting RNAs (piRNAs).[100–102] MILI1 and MIWI2 proteins belong to the well-conserved PIWI family and bind small RNAs of 26–30 nucleotides (piRNAs) to form RNA-based silencing complexes.[103] Piwi/piRNA complexes may serve as sequence-specific guides to direct de novo methylation of TEs (Figure 4). Loss of the piRNA pathway prevents recognition and silencing of L1 by DNMT3L in mouse germ cells.[102] Costa *et al.*[104] have shown that the absence of mouse MAELSTROM, a protein that interacts with both MILI and MWI in the germline-specific structure *nuage* (cloud in French), leads to dramatic derepression of L1 in spermatocytes via a mechanism involving piRNAs and loss of DNA methylation.[104,105] Others have associated members of the Tudor-domain family of proteins, TDRD1 and TDRD9, and GASZ, with the biogenesis of piRNA and the silencing of retrotransposons through de novo methylation in male germ cells.[106–108] While the Piwi/piRNA driven DNA methylation of TEs appears to be specific to germ cells, RNA interference (RNAi) could be a more generalized mechanism of L1 suppression. Murine embryonic stem cells deficient in Dicer (an enzyme that cleaves double-stranded RNAs producing 21–25 nucleotides siRNA) exhibit increased levels of L1 transcripts, providing evidence that RNA interference may control mammalian L1s.[109] Indeed, double-stranded RNA transcribed from a human L1 retrotransposon template is a substrate for *in vitro* cleavage by Dicer, and the resulting siRNAs lead to inhibition of L1 activity.[110] Moreover, siRNAs derived from human L1 itself, due to bidirectional transcription at the 5′ UTR, can suppress retrotransposition by reducing stability of the L1 transcript in cell cultures.[111]

Post-transcriptional protein regulators. L1 activity is also tightly controlled by various post-transcriptional protein regulators. Proteomic studies have identified numerous L1 binding partners that can positively and negatively modulate L1 activity (reviewed in Pizarro and Cristofari).[112] These proteins participate in cellular pathways related to diverse functions such as RNA processing, RNP formation, DNA synthesis, DNA repair, and host defense[70,113,114] .

A large proportion of these protein regulators are part of the interferon-mediated innate immune response, which resists infection from invading pathogens, particularly viruses. APOBEC3 (apolipoprotein B mRNA-editing enzyme catalytic peptide 1-like) proteins are a family of cytidine deaminases that afford intracellular resistance to retroviral replication.[115] All members of this family possess one or two cytidine deaminase domains that catalyze deamination of the cytosine base to uridine during cDNA synthesis, leading to C/G to T/A transition mutations of viral DNA. There are seven APOBEC3 genes denoted hA3A to hA3H that are clustered on human chromosome 22, and only a single gene, mA3 in rodents.[116,117] Previous reports demonstrate that human APOBEC3 proteins inhibit L1 retrotransposition *in vitro*,[65,118–124] while the more ancient APOBEC family members, APOBEC2 and AID (activation-induced deaminase), lack such activities.[119,122] Another study showed that AIDs from multiple species inhibit L1 retrotransposition in a DNA deamination-independent manner,[125] but the functional significance of this finding remains to be established. Other interferon stimulated genes limit L1 activity through other mechanisms. Proteins such as MOV10 and ZAP co-localize with L1 RNPs in stress granules and reduce L1 mRNA levels.[126] RNAseL also preferentially degrades L1 mRNA. Other interferon stimulated genes such as BST2, ISG20, MAVS,

and MX2 strongly inhibit L1 RT activity, but the mechanism remains unclear.[127]

Finally, ORF1p contains several serine/threonine residues that could be putatively phosphorylated by proline-directed protein kinases (PDPKs) such as cyclin-dependent kinases and mitogen-activated protein kinases. Indeed, S18, S27, T203, and T213 were found to be phosphorylated in human cells and mutating these sites has been found to decrease L1 retrotransposition.[128]

Chromatin landscapes and genome dynamics

L1 chromatin landscapes

In mammals, L1 sequences may act as "way stations" in XIST-mediated X-chromosome inactivation (XCI).[129,130] "Way station" was a term coined to define specific repeat sequences that facilitate spreading of XCI.[129] In the early stages of XCI, XIST RNA coats nearly the entire X chromosome thereby depleting RNA Polymerase II and transcription factors.[131,132] This process correlates with the density of L1 sequences in the X chromosome; with high-density regions showing efficient XCI and low-density regions showing escape from XCI.[31,133,134]

Epigenetic silencing of L1 sequences through DNA methylation and histone modifications serve as a mechanism to shape chromatin landscapes (Figure 6). A typical L1 promoter contains approximately 34 CpG sites, all of which are heavily methylated in most somatic cells.[98,136] Treatment of HeLa cells with the PAH carcinogen benzo (a)pyrene leads to enrichment of activation marks (histone-3 lysine 4 trimethylation and histone 3 lysine 9 acetylation), and reduction in the level of DNMT1 at the L1 promoter.[36,98] Delivery of non-L1 sequences in trans by L1 has been shown to influence the architecture of the surrounding chromatin. In fact, *Alu* elements which are transposed in trans by L1s are excluded from human imprinted regions owing to their potential negative effect on methylation.[137] Furthermore, a reporter gene delivered by L1 can be silenced shortly after delivery[138] or after serial passage in primary cells where the L1 reporter continues to be present but is not efficiently expressed (Ramos *et al.*, unpublished). In addition, the endonuclease activity of L1 ORF2 induces double-stranded DNA breaks at target loci leading to a reorganization of chromatin architecture[75,139].

The finding that genetic ablation of RB proteins leads to the reactivation of L1 elements suggests that RB proteins exert silencing roles beyond those of MeCP2 and may in fact involve changes in posttranslational histone modifications. pRB interacts with HDAC1, DNMT1, pRB-associated protein 48, suppressor of variegation 3–9 homolog 1 (Drosophila) (Suv39H1), and suppressor of variegation 4–20 homolog 2 (Drosophila) (Suv420H2). These proteins regulate chromatin conformation and induce signatures characteristic of silenced retroelements.[36] The silencing epigenetic marks H3K9me3 and H4K20me3 are specifically associated with TEs in centromeric regions in undifferentiated mouse cells.[140,141]

The chromatin alterations that follow L1 reactivation and/or mobilization are associated with alterations in cellular differentiation. A connection between L1 and cellular differentiation was established in studies showing that full-length L1 transcripts were not detectable after induction of differentiation in the embryonic stem cell line NTera2D1 with retinoic acid.[142] If heterochromatin formation spreads past the boundaries of L1 into adjacent sequences, it may negatively affect expression of neighboring genes.

Interestingly, a significantly higher density of full-length L1 sequences has been found in regions surrounding mono-allelically expressed autosomal genes, compared to bi-allelic genes in both mouse and human genomes.[143] Alus are enriched in nonimprinted genes compared to imprinted genes in both mouse and human genome.[137,143–146] Incidentally, olfactory receptor genes known to be expressed in a random mono-allelic manner are enriched with L1 sequence in their flanking regions.[143,147] These findings implicate L1 elements in the establishment of non-equivalent chromatin structures and mono-allelic expression at a subset of autosomal genes.

Role of L1 in genome dynamics

L1 sequences are often located within introns[26] (http://line1.molgen.mpg.de) and therefore represented in primary transcripts. As noted earlier, the processivity of the RNA polymerase complex through L1 sequences is inefficient due to A-rich bias leading to inefficient elongation,[48] premature polyadenylation,[49] and generation of different splicing variants.[50] These events appear to be orientation- and length-specific and are observed when the L1 sequence is in the same transcriptional orientation ("sense strand") at a minimum length of 1 kb, with longer sequences leading to stronger effects.[48] The "molecular rheostat" model proposed by Han *et al.* states that L1 sequences can affect genome-wide gene expression due to reduced transcriptional elongation of the genes in which they reside.[48] In support of this model, bioinformatics analyses have shown that the average amount of L1 sequence present per gene is markedly different for highly expressed compared to poorly expressed genes, with poorly expressed genes showing larger amounts of L1 sequence.[48]

A mechanism by which intronic LINE-1 influences genome dynamics relates to the recruitment of RBPs, particularly MATR3 and PTBP1, to nascent mRNA transcripts. Together these proteins repress cryptic splicing and polyadenylation sites around L1 sequences.[126,127] Relatively young L1s, which are closer to exons, are more heavily decorated with these repressive RBPs than older L1s, which lose their insulation over evolutionary time.[126,127]

Insertion orientation also plays an important role in how L1 alters genome dynamics. Although RBPs bind to L1 sequences in both orientations, they tend to be more highly enriched in antisense L1 sites.[126,127] Insertions in both orientations are known to influence genome dynamics. *In vivo*, naturally occurring full-length L1 insertions into introns in the sense orientation significantly reduced RNA levels of target genes in the mouse black-eyed white

gene,[148] and the human retinitis pigmentosa 2 gene.[149] These observations support a model where large "sense strand" insertions can interfere with target gene expression. We have found a similar relationship in nickel-treated lung bronchial epithelial cells undergoing malignant transformation associated with novel L1 insertions (Ramos and Bojang, unpublished).

"Antisense" intronic L1 insertions may impact gene expression through a "gene-breaking" mechanism where the insertion splits the transcript into two pieces by providing a strong polyadenylation site in the antisense ORF2 region (major antisense polyadenylation site, MAPS) of L1.[150,151] This would give rise to transcripts containing upstream exons and terminating in the MAPS, as well as transcripts originating from the L1 antisense promoter (ASP) and including downstream exons of the target gene (Figure 5). Bioinformatic analysis has identified several genes with transcripts that originate from the full-length L1 ASP located in intronic regions containing the correct splicing junctions of the L1 sequence and a sequence from the joined downstream exons.[151,153] Experimental analysis identified transcripts terminating at the MAPS for three of those genes, suggesting that gene-breaking may in fact occur in humans.[151] Over 150 full-length L1s with near exact matches to active human $L1^{RP}$, and thousands more slightly degenerate L1s, have been found in the antisense orientation in introns of human genes,[151] suggesting that gene breaking may afford cells a mechanism through which L1 elements remodel mammalian genomes. The function and biological consequences of such split gene products remain to be determined.

Activation of L1 by environmental injury

Compelling evidence suggests that L1 is activated in somatic tissues by genotoxic stress or tissue injury via mechanisms that involve loss of epigenetic silencing and transcriptional activation.[18,43–45,98] For example, UV irradiation of NTera2D1 cells is associated with substantial increases in L1 RT activity[154] and increased steady state levels of L1 mRNA in human microvascular endothelial cells and human cervical carcinoma HeLa cells.[45] PAHs have been shown to increase L1 expression in vascular smooth muscle cells,[43] microvascular endothelial cells,[45] cervical tumor cells,[44,45] embryonic kidney cells[88] and bronchial epithelial cells.[155]

L1 activation may also be sensitive to oxidative stress. Activation of murine L1 is mediated by proteins that bind in a redox-dependent manner to cis-acting regulatory elements located in the 5′ UTR of the L1Md-A5 retroelement.[18] The presence of electrophile response elements in the murine L1 promoter coupled with the presence of redox-regulated transcription factors and AHR within the promoter region established a central role for oxidative signaling in the activation of murine L1.[18] Thus, the evidence indicates that activation of L1 by stressful environments and disease facilitates recruitment of proteins that mediate L1 reactivation.

Heavy metals[156–158] and γ-irradiation[159] also increase L1 retrotransposition. While PAHs upregulate L1 RNA expression,[18,44,45] heavy metals and ionizing radiation stimulate retrotransposition through mechanisms that likely do not involve changes in L1 transcript levels.[158,159] Mutant cells deficient in non-homologous end-joining (NHEJ) of double-stranded break repair DNA repair lose their dependence on L1 endonuclease for integration and display retrotransposition rates near wild-type levels, most likely by integrating into preexisting breaks in the DNA.[160] Previous studies have also shown that endonuclease deficient L1 can use dysfunctional telomeres as integration substrates.[160] Importantly, contextual differences define L1 inducibility since not all genotoxic stressors reproducibly increase L1 retrotransposition.[159] Other work has revealed that genetic

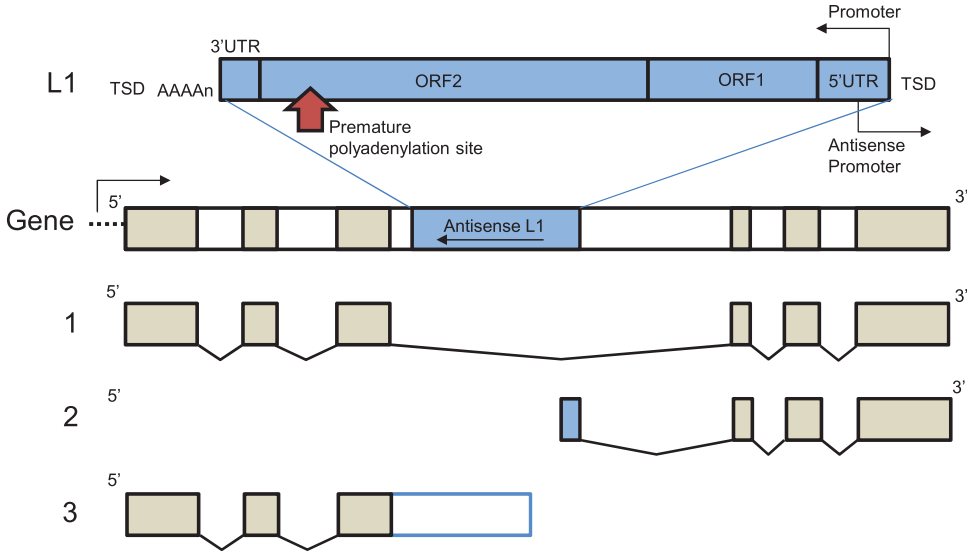

Figure 5. Gene breaking model proposed by Wheelan *et al*.[152] L1 is inserted within a gene's intron in the antisense direction. This could produce three possible transcripts: (1) The native, full-length transcript; (2) A truncated transcript arising from the antisense promoter in which a portion of L1 is spliced together with downstream 3′ exons; (3) A truncated transcript containing the gene's 5′ exons that terminates at the premature polyadenylation site and also contains a portion of the intron and L1 transcript. UTR: untranslated region; TSD: target site duplications.

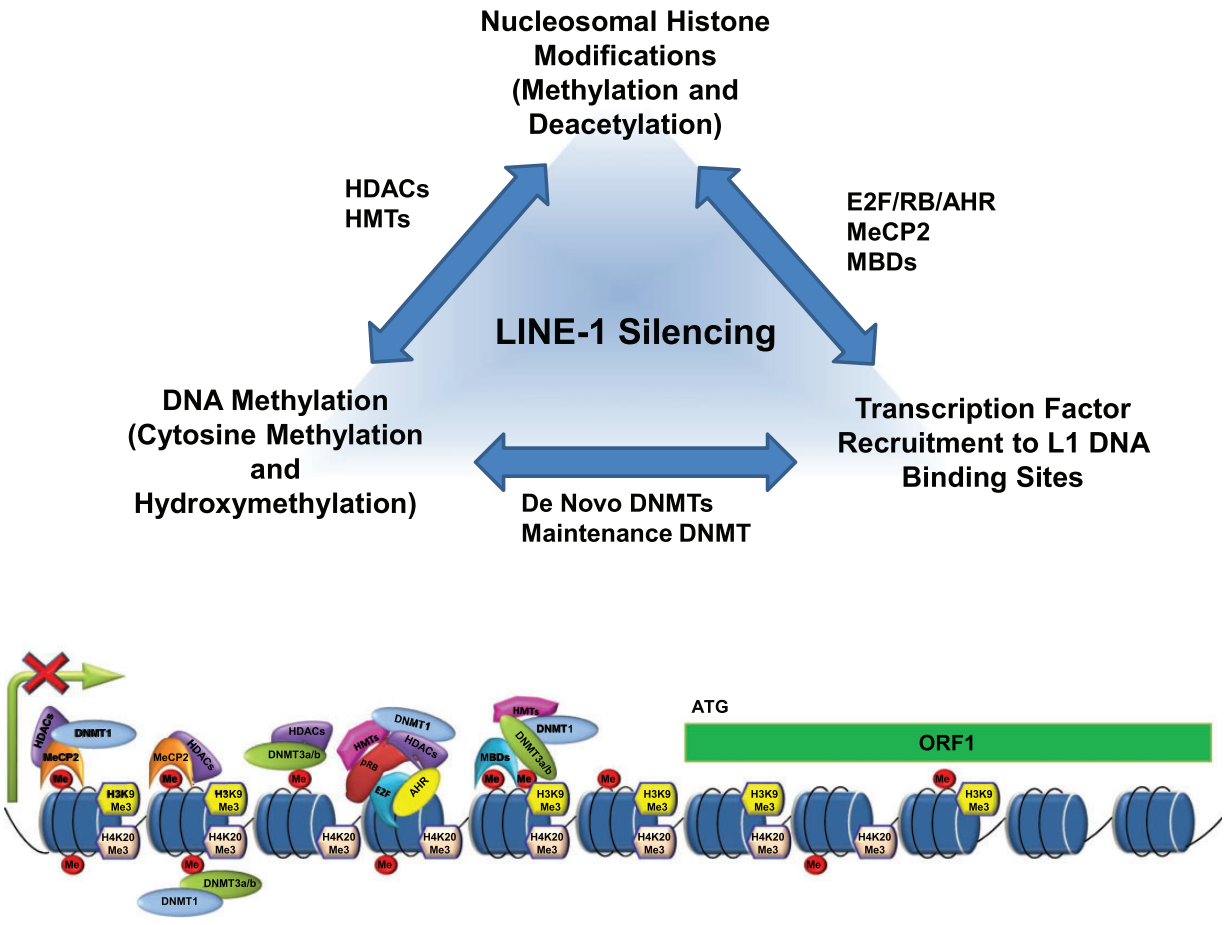

Figure 6. Mediators of L1 epigenetic silencing.[135] Schematic representation of L1 silencing mechanisms. Histone trimethylation by histone methyltransferases (HMTs) and histone deacetylation mediated by histone deacetylases (HDACs) act in concert with DNA methyltransferase (DNMT)-mediated DNA cytosine methylation and hydroxymethylation to create covalent epigenetic marks that induce L1 silencing. Evidence suggests that L1 can be maintained in a silenced state by macromolecular interactions between E2F, retinoblastoma and aryl hydrocarbon receptor (E2F/RB/AHR), methyl CpG binding protein 2 (Rett syndrome) (MeCP2), and DNMTs. This model hypothesizes that molecularly different complexes use a similar group of corepressors (e.g. HDACs, MBDs, HMTs) to bring about the epigenetic modifications required for L1 silencing, among them nucleosomal histone H3 and H4 trimethylation, DNA methylation, and histone deacetylation. *De novo* DNA methylation is hypothesized to play an important role in both L1 methylation during embryogenesis as well during insertion of novel full-length L1 elements in extraembryonic tissues/differentiated cells. It is possible that L1-associated epigenetic silencing marks are also present within the L1 "gene body".

defects leading to deficiency in DNA repair enzymes, such as flap endonuclease, increase L1 retrotransposition,[66] suggesting that L1 may contribute to genetic instability in affected individuals. Exposure of murine and human cells to etoposide, a topoisomerase II inhibitor, γ-irradiation, or UV also leads to transcriptional activation of SINEs (including human Alu sequences) with concomitant activation of endogenous RT activity from L1 elements.[161–164] These findings are clinically relevant given that L1-mediated SINE mobility may contribute to genomic instability after exposure to DNA-damaging chemotherapy,[165] and other stresses.[166]

Health implications

Cancer

Mechanisms involved in mutation and reprogramming of the genome by L1 have been frequently implicated as mediators and biomarkers of cancer. As previously described, hypermethylation of the L1 5′ UTR is critical for the repression of L1 sequences in healthy tissues, with erosion of methylation associated with increased L1 expression. As such, L1 promoter hypomethylation has been widely observed in cancers originating from many different tissues, including breast,[167] testis,[168] kidney,[169] prostate,[169,170] liver,[171,172] chronic lymphocytic leukemia,[173] ovary,[174] colon,[175] and lung.[176] In addition to biomarker studies based on methylation status, L1 hypomethylation has also been associated with functional changes that advance tumorigenesis.[177] Several studies have in fact established a link between hypomethylation of TEs and genomic instability during cancer initiation and progression.[178] L1 measures have also been associated with clinical features such as cancer risk and mortality.[179,180]

Furthermore, epigenetic changes in the L1 promoter have been associated with alterations in the expression of several differentiation genes in embryonic kidney cells,[88] HepG2 cells,[69] and vascular smooth muscle cells. In epithelial cells, these L1 activated gene networks can initiate epithelial to mesenchymal transition,[181] which is key feature of cancer progression, allowing tumors to become more invasive and acquire metastatic phenotypes.

One mechanism by which L1 activation may advance carcinogenesis is by inserting itself into genomic loci and producing deleterious functional changes in gene expression and genomic architecture, as previously discussed. Next generation sequencing has permitted investigation of L1 insertion patterns. We and others have documented that L1 insertions occur in cancer-associated genes and exhibit a bias against transcriptionally active genes, such as housekeeping genes.[178,182] However, more recent reports suggest that sequence and replication timing, particularly the presence on the leading strand of actively synthesized DNA, may be the most influential factors rather than gene content, transcription, or the local epigenetic environment.[183,184]

There have been many instances in which retrotransposon-mediated insertions have been linked to human disease.[185] Most detailed investigations to date have focused on cancer endpoints, with oncogenic transformation associated with elevated expression of retrotransposons,[178] and several L1 insertions documented into tumor suppressor genes.[186]

Autoimmune diseases

In other studies, a link between retroelements and autoimmunity has been established. This connection may involve recognition of nucleic acids by the innate immune system. Viral and retrotransposon nucleic acids can be detected by pathogen recognition receptors leading to the production of type I interferons through NF-κB and interferon regulatory factors (IRFs).[187] For instance, deficiencies in cellular processing of single-stranded DNA (ssDNA) by the 3′ exonuclease Trex1 are associated with inflammatory myocarditis in mice and with Aicardi-Goutiéres syndrome (AGS) in humans. AGS is a rare genetic disorder caused by damage to the myelin sheath severely affecting the brain, spinal cord, and immune system. Stetson *et al.* reported that cells lacking Trx1 accumulate ssDNA in the cytosol, with a striking proportion of DNA corresponding to endogenous retroelements in mouse hearts.[188] In their studies, overexpression of Trx1p suppressed retrotransposition of both human L1 and murine IAP *in vitro*. In contrast, overexpression of catalytically inactive Trex1 mutants associated with AGS had no effect on retrotransposition. Therefore, accumulation of reverse-transcribed DNA may be a key determinant for the involvement of endogenous retroelements in autoimmunity. In addition, the ORF1p is known to be associated with many autoantigens implicated in autoimmune diseases such as Systemic Lupus Erythematosus (SLE). Indeed, a majority of SLE patients have IgG autoantibodies against ORF1p and the degree of reactivity has been found to correlate with serological measures of disease severity.[189,190]

Unraveling the role of L1 in human disease may be made more complex by the fact that active human L1 retrotransposons are highly polymorphic and exhibit variable rates of retrotransposition. In addition to contributing to genetic diversity, this variation also likely contributes to inter-individual differences in the occurrence and severity of disease.[191,192]

Conclusions

L1 is an active mobile element that has shaped the human genome over evolutionary time. L1 has vital roles in development, cell division, differentiation, and the establishment of genome architecture. Numerous mechanisms have evolved to control L1 activity, including epigenetic suppression, transcriptional and translational control measures, and repressive non-coding RNAs. Exposure to genotoxins and agents that alter the epigenome can lead to L1 activation, and uncontrolled L1 activation can lead to cancer initiation and progression through deleterious de novo L1 insertions, changes in gene expression, and remodeling of genome structure and function. As such, markers of L1 activation are common features of many different types of cancers.

Continued efforts will be required to elucidate the role of L1 in the onset and progression of chronic human diseases and the extent to which genetic modifications by L1 define chromatin architecture, genome dynamics, and cellular plasticity (Figure 6). Given the potentially deleterious effects of L1, histone trimethylation and histone deacetylation mediated act in concert with DNA methyltransferases to establish epigenetic marks that mediate L1 silencing. This model hypothesizes that molecularly different complexes use a similar group of corepressors to bring about the epigenetic modifications required for L1 silencing. These relationships are detailed in several of the studies discussed earlier, and reviewed elsewhere.[135] As such, the intersection of genetic and epigenetic mechanisms afforded by L1 poses interesting challenges as genomic medicine efforts continue to be integrated into the diagnosis and treatment of chronic human diseases.

AUTHORS' CONTRIBUTIONS

The manuscript and figures were conceived and prepared by KSR and PB, with additional intellectual contributions, editing and revision by EB.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Ramos and Bowers filed patent 2019 62/787444 "Diagnostic Analyte to Enrich LINE-1 and Detect LINE-1 Driven Oncologic Disease: Systems and Methods for Characterizing and Treating Disease".

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Parts of the work reviewed here were supported in part by grants from the National Institutes of Health [5R01 ES004849; GB090603A1 and ES014443], Astra Zeneca, and Texas Governors Research Initiative to KSR. EB was supported the University of Arizona Bio5 Postdoctoral Fellowship and a Texas Governors Distinguished Research Investigator Award to KSR.

ORCID iDs

Kenneth S Ramos https://orcid.org/0000-0002-5196-8674
Emma Bowers https://orcid.org/0000-0002-2009-2489

REFERENCES

1. International Consortium Completes Human Genome Project, https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp (accessed 1 July 2021)
2. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Kheradpour P, Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJ, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, Zhang J, Wexler JR, Good PJ, Feingold EA, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Giddings MC, Gingeras TR, Guigó R, Hubbard TJ, Kent WJ, Lieb JD, Margulies EH, Myers RM, Stamatoyannopoulos JA, Tenenbaum SA, Weng Z, White KP, Wold B, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Eaton ML, Dobin A, Tanzer A, Lagarde J, Lin W, Xue C, Williams BA, Zaleski C, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Robyr D, Ruan X, Sammeth M, Sandhu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Hayashizaki Y, Reymond A, Antonarakis SE, Hannon GJ, Ruan Y, Carninci P, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Grasfeder LL, Giresi PG, Battenhouse A, Sheffield NC, Showers KA, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Iyer VR, Zheng M, Wang P, Gertz J, Vielmetter J, Partridge EC, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, Muratet MA, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Newberry JS, Levy SE, Absher DM, Wong WH, Blow MJ, Visel A, Pennachio LA, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Davidson C, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Hunt T, Jungreis I, Kay M, Khurana E, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tapanari E, Tress ML, Van Baren MJ, Washietl S, Wilming L, Zadissa A, Zhang Z, Brent M, Haussler D, Valencia A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyengar S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Lamarre-Vincent N, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Raha D, Ramirez L, Reed B, Shi M, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Struhl K, Weissman SM, Penalva LO, Karmakar S, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Victorsen A, Auer T, Centanin L, Eichenlaub M, Gruhl F, Heermann S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Jain G, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Johnson AK, Johnson EM, Kutyavin TV, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sanchez ME, Sandstrom RS, Shafer AO, Stergachis AB, Thomas S, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JM, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Beal K, Brazma A, Flicek P, Johnson N, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Miller W, Bickel PJ, Banfai B, Boley NP, Huang H, Li JJ, Noble WS, Bilmes JA, Buske OJ, Sahu AD, Kharchenko PV, Park PJ, Baker D, Taylor J, Lochovsky L. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74
3. Reznikoff WS. Tn5 as a model for understanding DNA transposition. *Mol Microbiol* 2003;**47**:1199–206
4. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 2009;**10**:691–703
5. Havecker ER, Gao X, Voytas DF. The diversity of LTR retrotransposons. *Genome Biol* 2004;**5**:225
6. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends Genet* 2007;**23**:183–91
7. 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**:1875–83
8. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Hong ML, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, De La Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran J V, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM,

Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A, Morgan MJ. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860–921

9. de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 2011;**7**:e1002384
10. Burns KH, Boeke JD. Human transposon tectonics. *Cell* 2012;**149**:740–52
11. Kazazian HH. Mobile elements: drivers of genome evolution. *Science* 2004;**303**:1626–32
12. Ostertag EM, Kazazian HH, Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001;**35**:501–38
13. Dombroski BA, Feng Q, Mathias SL, Sassaman DM, Scott AF, Kazazian HH, Boeke JD. An in vivo assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1994;**14**:4485–92
14. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, Löwer J, Strätling WH, Löwer R, Schumann GG. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* 2012;**40**:1666–83
15. Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 2011;**12**:615–27
16. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. Isolation of an active human transposable element. *Science* 1991;**254**:1805–8
17. Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, Edgell MH, Hutchison Iii CA. The sequence of a large LlMd element reveals a tandemly repeated 5′ end and several features found in retrotransposons. *Mol Cell Biol* 1986;**6**:168–82
18. Lu KP, Ramos KS. Redox regulation of a novel L1Md-A2 retrotransposon in vascular smooth muscle cells. *J Biol Chem* 2003;**278**:28201–9
19. Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* 1990;**10**:6718–29
20. Goodier JL, Ostertag EM, Kazazian HH. Transduction of 3′-flanking sequences is common in L1 retrotransposition. *Hum Mol Genet* 2000;**9**:653–7
21. Pickeral OK, Makaowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by line-1 retrotransposition. *Genome Res* 2000;**10**:411–5
22. 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**:2253–60
23. Ramos KS, Teneng I, Montoya-Durango DE, Bojang P, Haeberle MT, Ramos IN, Stribinskis V, Kalbfleisch T. The intersection of genetics and epigenetics: reactivation of mammalian LINE-1 retrotransposons by environmental injury. In: Jirtle R, Tyson F (eds) *Environmental epigenomics in health and disease: epigenetics and disease origins*. Berlin: Springer, 2013, pp. 127–60
24. DeBerardinis RJ, Goodier JL, Ostertag EM, Kazazian HH. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet* 1998;**20**:288–90
25. 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**:research00521
26. Penzkofer T, Dandekar T, Zemojtel T. L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res* 2005;**33**:D498
27. Athanikar JN, Badge RM, Moran JV. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* 2004;**32**:3846–55
28. 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**:4685–97
29. Becker KG, Swergold G, Ozato K, Thayer RE. Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum Mol Genet* 1993;**2**:1697–702
30. Tchénio T, Casella JF, Heidmann T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res* 2000;**28**:411–5
31. Yang F, Babak T, Shendure J, Disteche CM. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* 2010;**20**:614–22
32. Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH. L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 2010;**468**:443–6
33. Levanon D, Bettoun D, Harris-Cerruti C, Woolf E, Negreanu V, Eilam R, Bernstein Y, Goldenberg D, Xiao C, Fliegauf M, Kremer E, Otto F, Brenner O, Lev-Tov A, Groner Y. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. *EMBO J* 2002;**21**:3454–63
34. Yang N, Zhang L, Zhang Y, Kazazian HH. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res* 2003;**31**:4929–40
35. Yu F, Zingler N, Schumann G, Strätling WH. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not alu transcription. *Nucleic Acids Res* 2001;**29**:4493–501
36. Montoya-Durango DE, Liu Y, Teneng I, Kalbfleisch T, Lacy ME, Steffen MC, Ramos KS. Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. *Mutat Res* 2009;**665**:20–8
37. Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, O'Carroll D, Firestein R, Cleary M, Jenuwein T, Herrera RE, Kouzarides T. Rb targets histone H3 methylation and HP1 to promoters. *Nature* 2001;**412**:561–5
38. Pradhan S, Kim GD. The retinoblastoma gene product interacts with maintenance human DNA (cytosine-5) methyltransferase and modulates its activity. *EMBO J* 2002;**21**:779–88
39. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with RB, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 2000;**25**:338–42
40. Montoya-Durango DE, Ramos KS. L1 retrotransposon and retinoblastoma: molecular linkages between epigenetics and cancer. *Curr Mol Med* 2010;**10**:511–21
41. Montoya-Durango DE, Ramos KS. Retinoblastoma family of proteins and chromatin epigenetics: a repetitive story in a few LINEs. *Biomol Concepts* 2011;**2**:233–45
42. Chinnam M, Goodrich DW. RB1, development, and cancer. *Curr Top Dev Biol* 2011;**94**:129–69
43. Lu KP, Hallberg LM, Tomlinson J, Ramos KS. Benzo(a)pyrene activates L1Md retrotransposon and inhibits DNA repair in vascular smooth muscle cells. *Mutat Res* 2000;**454**:35–44
44. Stribinskis V, Ramos KS. Activation of human long interspersed nuclear element 1 retrotransposition by benzo(a)pyrene, an ubiquitous environmental carcinogen. *Cancer Res* 2006;**66**:2616–20
45. Teneng I, Stribinskis V, Ramos KS. Context-specific regulation of LINE-1. *Genes Cells* 2007;**12**:1101–10
46. Okudaira N, Iijima K, Koyama T, Minemoto Y, Kano S, Mimori A, Ishizaka Y. Induction of long intersperse nucleotide element-1 (L1) retrotransposition by 6-formylindolo[3,2-b]carbazole (FICZ), a tryptophan photoproduct. *Proc Natl Acad Sci U S A* 2010;**107**:18487–92
47. Han JS, Boeke JD. A highly active synthetic mammalian retrotransposon. *Nature* 2004;**429**:314–8
48. Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 2004;**429**:268–74
49. Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet* 2003;**35**:363–6
50. Belancio VP, Roy-Engel AM, Deininger P. The impact of multiple splice sites in human L1 elements. *Gene* 2008;**411**:38–45
51. Holmes SE, Singer MF, Swergold GD. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading

frame of an active human LINE-1 transposable element. *J Biol Chem* 1992;**267**:19765–8
52. Hohjoh H, Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J* 1996;**15**:630–9
53. Hohjoh H, Singer MF. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J* 1997;**16**:6034–43
54. 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 U S A* 1997;**94**:10155–60
55. 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**:8112–7
56. Martin SL, Cruceanu M, Branciforte D, Li PWL, Kwok SC, Hodges RS, Williams MC. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J Mol Biol* 2005;**348**:549–61
57. Martin SL. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol* 2010;**7**:706–11
58. Martin F, Marañón C, Olivares M, Alonso C, López MC. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from Trypanosoma cruzi: homology of the first ORF with the ape family of DNA repair enzymes. *J Mol Biol* 1995;**247**:49–59
59. Feng Q, Moran JV, Kazazian HH, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 1996;**87**:905–16
60. Hattori M, Kuhara S, Takenaka O, Sakaki Y. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature* 1986;**321**:625–8
61. Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science* 1991;**254**:1808–10
62. Fanning TG, Singer MF. LINE-1: a mammalian transposable element. *Biochim Biophys Acta* 1987;**910**:203–12
63. 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**:661–8
64. 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
65. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000;**24**:363–7
66. Gasior SL, Roy-Engel AM, Deininger PL. ERCC1/XPF limits L1 retrotransposition. *DNA Repair* 2008;**7**:983–9
67. Basame S, Wai-Lun Li P, Howard G, Branciforte D, Keller D, Martin SL. Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *J Mol Biol* 2006;**357**:351–7
68. Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev* 2006;**20**:210–24
69. Bojang P, Roberts RA, Anderton MJ, Ramos KS. Reprogramming of the HepG2 genome by long interspersed nuclear element-1. *Mol Oncol* 2013;**7**:812–25
70. Taylor MS, LaCava J, Mita P, Molloy KR, Huang CRL, Li D, Adney EM, Jiang H, Burns KH, Chait BT, Rout MP, Boeke JD, Dai L. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* 2013;**155**:1034–48
71. Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* 2013;**41**:e28
72. Hernan R, Heuermann K, Brizzard B. Multiple epitope tagging of expressed proteins for enhanced detection. *Biotechniques* 2000;**28**:789–93
73. 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
74. Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 2006;**357**:1383–93
75. Wallace NA, Belancio VP, Deininger PL. L1 mobile element expression causes multiple types of toxicity. *Gene* 2008;**419**:75–81
76. Kulpa DA, Moran JV. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum Mol Genet* 2005;**14**:3237–48
77. Martin SL. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol* 1991;**11**:4804–7
78. Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N. Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet* 2010;**6**:1–19
79. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. Human L1 retrotransposition: cisPreference versus trans complementation. *Mol Cell Biol* 2001;**21**:1429–39
80. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH. L1 integration in a transgenic mouse model. *Genome Res* 2006;**16**:240–50
81. Mita P, Wudzinska A, Sun X, Andrade J, Nayak S, Kahler DJ, Badri S, LaCava J, Ueberheide B, Yun CY, Fenyö D, Boeke JD. LINE-1 protein localization and functional dynamics during the cell cycle. *eLife* 2018;**7**: e30058
82. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *EMBO J* 2002;**21**:5899–910
83. Garcia-Perez JL, Doucet AJ, Bucheton A, Moran JV, Gilbert N. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res* 2007;**17**:602–11
84. Dewannieux M, Heidmann T, Yaniv M. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J Mol Biol* 2005;**349**:241–7
85. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* 2003;**73**:1444–51
86. Gould SJ, Vrba ES. Exaptation – a missing term in the science of form. *Paleobiology* 1982;**1**:4–15
87. Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, Kokubo N, Kimura-Yoshida C, Matsuo I, Sumiyama K, Saitou N, Shimogori T, Okada N. Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci U S A* 2008;**105**:4220–5
88. Ramos KS, Montoya-Durango DE, Teneng I, Nanez A, Stribinskis V. Epigenetic control of embryonic renal cell differentiation by L1 retrotransposon. *Birth Defects Res A Clin Mol Teratol* 2011;**91**:693–702
89. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 2008;**9**:397–405
90. Ramos KS, He Q, Kalbfleisch T, Montoya-Durango DE, Teneng I, Stribinskis V, Brun M. Computational and biological inference of gene regulatory networks of the LINE-1 retrotransposon. *Genomics* 2007;**90**:176–85
91. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009;**462**:315–22
92. Tomizawa SI, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* 2011;**138**:811–20
93. Dahl C, Grønbæk K, Guldberg P. Advances in DNA methylation: 5-hydroxymethylcytosine revisited. *Clin Chim Acta* 2011;**412**:831–6
94. Phokaew C, Kowudtitham S, Subbalekha K, Shuangshoti S, Mutirangura A. LINE-1 methylation patterns of different loci in normal and cancerous cells. *Nucleic Acids Res* 2008;**36**:5704–12
95. Dolinoy DC. The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr Rev* 2008;**66**(Suppl 1):S7–11
96. Woodcock DM, Linsenmeyer ME, Warren WD. DNA methylation in mouse A-repeats in DNA methyltransferase-knockout ES cells and in

normal cells determined by bisulfite genomic sequencing. *Gene* 1998;**206**:63–7
97. Thayer RE, Singer MF, Fanning TG. Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1 -encoded protein. *Gene* 1993;**133**:273–7
98. Teneng I, Montoya-Durango DE, Quertermous JL, Lacy ME, Ramos KS. Reactivation of L1 retrotransposon by benzo(a)pyrene involves complex genetic and epigenetic regulation. *Epigenetics* 2011;**6**:355–67
99. Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 2012;**151**:1417–30
100. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 2007;**316**:744–7
101. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T. DNA methylation of retrotransposon genes is regulated by piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* 2008;**22**:908–17
102. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 2008;**31**:785–99
103. Girard A, Hannon GJ. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol* 2008;**18**:136–48
104. Costa Y, Speed RM, Gautier P, Semple CA, Maratou K, Turner JMA, Cooke HJ. Mouse MAELSTROM: the link between meiotic silencing of unsynapsed chromatin and microRNA pathway? *Hum Mol Genet* 2006;**15**:2324–34
105. Soper SFC, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P, Bortvin A. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* 2008;**15**:285–97
106. Ma L, Buchold GM, Greenbaum MP, Roy A, Burns KH, Zhu H, Han DY, Harris RA, Coarfa C, Gunaratne PH, Yan W, Matzuk MM. GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet* 2009;**5**:e1000635
107. Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. Loss of the mili-interacting tudor domain-containing protein-1 activates transposons and alters the mili-associated small RNA profile. *Nat Struct Mol Biol* 2009;**16**:639–46
108. Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, Kondoh G, Okawa K, Chujo T, Suzuki T, Hata K, Martin SL, Noce T, Kuramochi-Miyagawa S, Nakano T, Sasaki H, Pillai RS, Nakatsuji N, Chuma S. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell* 2009;**17**:775–87
109. Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* 2005;**19**:489–501
110. Soifer HS, Zaragoza A, Peyvan M, Behlke MA, Rossi JJ. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res* 2005;**33**:846–56
111. Yang N, Kazazian HH. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* 2006;**13**:763–71
112. Pizarro JG, Cristofari G. Post-transcriptional control of LINE-1 retrotransposition by cellular host factors in somatic cells. *Front Cell Dev Biol* 2016;**4**:14
113. Goodier JL, Cheung LE, Kazazian HH. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res* 2013;**41**:7401–19
114. Moldovan JB, Moran JV. The zinc-finger antiviral protein ZAP inhibits LINE and alu retrotransposition. *PLoS Genet* 2015;**11**:e1005121
115. Harris RS, Liddament MT. Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* 2004;**4**:868–77
116. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 2005;**22**:367–77
117. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* 2002;**79**:285–96
118. 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 U S A* 2006;**103**:8780–5
119. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol* 2006;**16**:480–5
120. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* 2007;**35**:2955–64
121. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Löwer J, Cichutek K, Flory E, Schumann GG, Münk C. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* 2006;**281**:22161–72
122. Niewiadomska AM, Tian C, Tan L, Wang T, Sarkis PTN, Yu X-F. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-Body association. *J Virol* 2007;**81**:9577–83
123. Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* 2006;**281**:16837–41
124. Tan L, Sarkis PTN, Wang T, Tian C, Yu X. Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. *FASEB J* 2009;**23**:279–87
125. Macduff DA, Demorest ZL, Harris RS. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res* 2009;**37**:1854–67
126. Attig J, Agostini F, Gooding C, Chakrabarti AM, Singh A, Haberman N, Zagalak JA, Emmett W, Smith CWJ, Luscombe NM, Ule J. Heteromeric RNP assembly at LINEs controls lineage-specific RNA processing. *Cell* 2018;**174**:1067–81
127. Attig J, Ule J. Genomic accumulation of retrotransposons was facilitated by repressive RNA-Binding proteins: a hypothesis. *Bioessays* 2019;**41**:e1800132
128. Cook PR, Jones CE, Furano AV. Phosphorylation of ORF1p is required for L1 retrotransposition. *Proc Natl Acad Sci U S A* 2015;**112**:4298–303
129. Gartler SM, Riggs AD. Mammalian X-chromosome inactivation. *Annu Rev Genet* 1983;**17**:155–90
130. Buzin C, Mann J, Singer-Sam J. Quantitative RT-PCR assays show xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. *Development* 1994;**120**:3529–36
131. Chaumeil J, Le Baccon P, Wutz A, Heard E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* 2006;**20**:2223–37
132. Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E, Greally JM, Voinnet O, Heard E. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* 2010;**141**:956–69
133. Bailey JA, Carrel L, Chakravarti A, Eichler EE. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A* 2000;**97**:6634–9
134. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 2005;**434**:400–4
135. Khalid M, Bojang P, Hassanin AAI, Bowers EC, Ramos IN, Ramos KS. LINE-1: implications in the etiology of cancer, clinical applications and pharmacological targets. *Mutat Res* 2018;**778**:51–60
136. Hata K, Sakaki Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 1997;**189**:227–34
137. Greally JM. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A* 2002;**99**:327–32
138. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, Moran JV.

Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 2010;**466**:769–73
139. Sen SK, Huang CT, Han K, Batzer MA. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res* 2007;**35**:3741–51
140. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;**448**:553–60
141. Martens JHA, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 2005;**24**:800–12
142. Skowronski J, Singer MF. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc Natl Acad Sci U S A* 1985;**82**:6050–4
143. Allen E, Horvath S, Tong F, Kraft P, Spiteri E, Riggs AD, Marahrens Y. High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc Natl Acad Sci U S A* 2003;**100**:9940–5
144. Ke X, Thomas NS, Robinson DO, Collins A. The distinguishing sequence characteristics of mouse imprinted genes. *Mamm Genome* 2002;**13**:639–45
145. Weidman JR, Murphy SK, Nolan CM, Dietrich FS, Jirtle RL. Phylogenetic footprint analysis of IGF2 in extant mammals. *Genome Res* 2004;**14**:1726–32
146. Walter J, Hutter B, Khare T, Paulsen M. Repetitive elements in imprinted genes. *Cytogenet Genome Res* 2006;**113**:109–15
147. Chess A, Simon I, Cedar H, Axel R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* 1994;**78**:823–34
148. Yajima I, Sato S, Kimura T, Yasumoto KI, Shibahara S, Goding CR, Yamamoto H. An L1 element intronic insertion in the black-eyed white (mitf(mi-bw)) gene: the loss of a single mitf isoform responsible for the pigmentary defect and inner ear deafness. *Hum Mol Genet* 1999;**8**:1431–41
149. Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, Van Duijnhoven G, Kirschner R, Hemberger M, Bergen AAB, Rosenberg T, Pinckers AJLG, Fundele R, Rosenthal A, Cremers FPM, Ropers HH, Berger W. Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet* 1998;**19**:327–32
150. Han JS, Boeke JD. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays* 2005;**27**:775–84
151. Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* 2005;**15**:1073–8
152. Wheelan SJ. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* 2005;**15**:1073–8
153. Nigumann P, Redik K, Mätlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 2002;**79**:628–34
154. Deragon JM, Sinnett D, Labuda D. Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. *EMBO J* 1990;**9**:3363–8
155. Bojang P, Ramos KS. Epigenetic reactivation of LINE-1 retrotransposon disrupts NuRD corepressor functions and induces oncogenic transformation in human bronchial epithelial cells. *Mol Oncol* 2018;**12**:1342–57
156. Kale SP, Moore L, Deininger PL, Roy-Engel AM. Heavy metals stimulate human LINE-1 retrotransposition. *Int J Eviron Res Public Health* 2005;**2**:14–23
157. Kale S, Carmichael M, Harris K, Roy-Engel A. 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**:121–8
158. El-Sawy M, Kale SP, Dugan C, Nguyen TQ, Belancio V, Bruch H, Roy-Engel AM, Deininger PL. Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. *J Mol Biol* 2005;**354**:246–57
159. Farkash EA, Kao GD, Horman SR, Prak ETL. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res* 2006;**34**:1196–204
160. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 2002;**31**:159–65
161. Rudin CM, Thompson CB. Transcriptional activation of short interspersed elements by DNA-damaging agents. *Genes Chromosom Cancer* 2001;**30**:64–71
162. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked alu sequences. *Nat Genet* 2003;**35**:41–8
163. Hulme AE, Bogerd HP, Cullen BR, Moran JV. Selective inhibition of alu retrotransposition by APOBEC3G. *Gene* 2007;**390**:199–205
164. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM. LINE-1 ORF1 protein enhances alu SINE retrotransposition. *Gene* 2008;**419**:1–6
165. Hagan CR, Sheffield RF, Rudin CM. Human alu element retrotransposition induced by genotoxic stress. *Nat Genet* 2003;**35**:219–20
166. Liu W, Man Chu W, Ming Choudary PV, Schmid CW. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* 1995;**23**:1758–65
167. Bratthauer GL, Cardiff RD, Fanning TG. Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* 1994;**73**:2333–6
168. Bratthauer GL, Fanning TG. Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* 1992;**7**:507–10
169. Florl AR, Löwer R, Schmitz-Dräger 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**:1312–21
170. Santourlidis S, Florl A, Ackermann R, Wirtz HC, Schulz WA. High frequency of alterations in DNA methylation in adenocarcinoma of the prostate. *Prostate* 1999;**39**:166–74
171. Lin CH, Hsieh SY, Sheen IS, Lee WC, Chen TC, Shyu WC, Liaw YF. Genome-wide hypomethylation in hepatocellular carcinogenesis. *Cancer Res* 2001;**61**:4238–43
172. 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**:306–9
173. Dante R, Dante-Paire J, Rigal D, Roizes G. Methylation patterns of long interspersed repeated DNA and alphoid repetitive DNA from human cell lines and tumors. *Anticancer Res* 1992;**12**:559–63
174. Menendez L, Benigno BB, McDonald JF. L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. *Mol Cancer* 2004;**3**:12
175. Suter CM, Martin DI, Ward RI. Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. *Int J Colorectal Dis* 2004;**19**:95–101
176. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis V, Field JK, Liloglou T. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* 2009;**124**:81–7
177. Schulz WA. L1 retrotransposons in human cancers. *J Biomed Biotechnol* 2006;**2006**:1–12
178. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ. Landscape of somatic retrotransposition in human cancers. *Science* 2012;**337**:967–71
179. Barchitta M, Quattrocchi A, Maugeri A, Vinciguerra M, Agodi A. LINE-1 hypomethylation in blood and tissue samples as an epigenetic marker for cancer risk: a systematic review and meta-analysis. *PLoS One* 2014;**9**:e109478
180. Saito K, Kawakami K, Matsumoto I, Oda M, Watanabe G, Minamoto T. Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. *Clin Cancer Res* 2010;**16**:2418–26
181. Reyes-Reyes EM, Aispuro I, Tavera-Garcia MA, Field M, Moore S, Ramos I, Ramos KS. LINE-1 couples EMT programming with acquisition of oncogenic phenotypes in human bronchial epithelial cells. *Oncotarget* 2017;**8**:103828–42

182. Bojang P, Anderton MJ, Roberts RA, Ramos KS. De novo LINE-1 retrotransposition in HepG2 cells preferentially targets gene poor regions of chromosome 13. *Genomics* 2014;**104**:96–104
183. Flasch DA, Macia Á, Sánchez L, Ljungman M, Heras SR, García-Pérez JL, Wilson TE, Moran J V. Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell* 2019;**177**:837–51.e28
184. Sultana T, van Essen D, Siol O, Bailly-Bechet M, Philippe C, Zine El Aabidine A, Pioger L, Nigumann P, Saccani S, Andrau J-C, Gilbert N, Cristofari G. The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection. *Mol Cell* 2019;**74**:555–70.e7
185. Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res* 2008;**18**:343–58
186. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* 1992;**52**:643–5
187. Akira S. TLR signaling. *Curr Top Microbiol Immunol* 2006;**311**:1–16
188. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* 2008;**134**:587–98
189. Carter V, LaCava J, Taylor MS, Liang SY, Mustelin C, Ukadike KC, Bengtsson A, Lood C, Mustelin T. High prevalence and disease correlation of autoantibodies against p40 encoded by long interspersed nuclear elements in systemic lupus erythematosus. *Arthritis Rheumatol* 2020;**72**:89–99
190. Crow MK. Reactivity of IgG with the p40 protein encoded by the long interspersed nuclear element 1 retroelement: comment on the article by carter. *Arthritis Rheumatol* 2020;**72**:374–76
191. Hedges DJ, Deininger PL. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* 2007;**616**:46–59
192. Del Carmen Seleme M, Vetter MR, Cordaux R, Bastone L, Batzer MA, Kazazian HH. Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc Natl Acad Sci U S A* 2006;**103**:6611–6