

## REVIEW

## Open Access



# Roles for retrotransposon insertions in human disease

Dustin C. Hancks<sup>1\*</sup> and Haig H. Kazazian Jr.<sup>2\*</sup>

## Abstract

Over evolutionary time, the dynamic nature of a genome is driven, in part, by the activity of transposable elements (TE) such as retrotransposons. On a shorter time scale it has been established that new TE insertions can result in single-gene disease in an individual. In humans, the non-LTR retrotransposon Long Interspersed Element-1 (LINE-1 or L1) is the only active autonomous TE. In addition to mobilizing its own RNA to new genomic locations via a “copy-and-paste” mechanism, LINE-1 is able to retrotranspose other RNAs including Alu, SVA, and occasionally cellular RNAs. To date in humans, 124 LINE-1-mediated insertions which result in genetic diseases have been reported. Disease causing LINE-1 insertions have provided a wealth of insight and the foundation for valuable tools to study these genomic parasites. In this review, we provide an overview of LINE-1 biology followed by highlights from new reports of LINE-1-mediated genetic disease in humans.

**Keywords:** Retrotransposon, LINE-1, Alu, SVA, LINE-1, Disease, Retrotransposition, Autoimmunity, Cancer

## Background

### A brief history

Transposable elements (TEs) are pieces of nucleic acid that encode the inherent ability to mobilize from one genomic location to another. This ability to “jump” is mediated by element-encoded proteins such as DNA transposase or reverse transcriptase. These TEs are referred to as autonomous. In other instances, non-coding TEs -typically referred to as non-autonomous- contain sequence features (e.g. sequence motifs, RNA structural elements), which are recognized by autonomous TE proteins that ultimately result in *trans*-mobilization of these sequences. Collectively, autonomous and non-autonomous transposable elements often comprise greater than 50 % of genomic real estate in mammals. For humans, approximately two-thirds of our genome can be annotated as TE-derived [1–6]; however, it is likely that the actual percentage is greater but due to sequence decay no sequence identity can be assigned.

Almost 70 years ago, Barbara McClintock laid the foundation for TE research with her initial work and discoveries in

maize of what she termed “controlling elements [7].” Since that time, several discoveries have been made leading to an active research community investigating the impact of transposable elements on the human genome and their role in disease. Although work by Britten and Davidson in the 1960s provided hints that the human genome was largely repetitive [8, 9], it wasn’t until the Human Genome Project [4–6] that the true origin and extent of the repeats in our genome became evident. The initial human genome draft sequence estimated that roughly 45 % of our genomic sequence is derived from TE sequence. The Human Genome and other genome projects [1, 3, 6] significantly transformed TE biology by providing the ability to answer questions including 1) Which TEs have been the most active?, 2) Where are specific TEs maintained in the genome?, 3) Which elements and how many have been recently active?

A pivotal transformation in TE biology occurred less than 10 years after the publication of the Human Genome Project. Next-generation sequencing has empowered researchers to interrogate longstanding and previously intractable questions regarding TE biology [7, 10, 11]. Examples include the frequency and location of new insertions and the contribution of TEs to gene regulation genome-wide at an unprecedented resolution [8, 9, 12, 13]. New studies will likely unveil novel ways by which these selfish

\* Correspondence: [dhancks@genetics.utah.edu](mailto:dhancks@genetics.utah.edu); [hkazazi@jhmi.edu](mailto:hkazazi@jhmi.edu)

<sup>1</sup>Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA

<sup>2</sup>McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins School of Medicine, Baltimore, MD, USA

genetic elements may actually be altruistic or even co-opted by the host genome [14] along with new insights into mechanisms by which they can cause disease. Here we provide an update of human TE biology, with a specific emphasis on LINE-1-mediated retrotransposition and disease-causing insertions.

### Human transposable elements

TEs are historically subdivided into two major classes defined by their mobilization intermediate. Class I TEs, also known as retrotransposons, encompass elements that move via a “copy-and-paste” mechanism involving an RNA intermediate [15, 16], while Class II TEs, referred to as DNA transposons, represent TEs that mobilize by a “cut-and-paste” mechanism. DNA transposons are currently thought to be transpositionally inactive in most mammals with bats being the exception [17, 18]; however, several genes in the human genome are derived from DNA transposons [6]. Three of these genes (recombination activating gene 1 (*RAG1*) [19], PiggyBac transposable element-derived protein 5 (*PGBD5*) [20], and THAP domain containing 9 (*THAP9*) [21]) are evolutionarily conserved and can carry out DNA transposition in cell culture or perform reactions reminiscent of DNA transposition. In contrast, retrotransposons (Fig. 1) remain quite active in humans [22–24]; any two human beings differ on average by ~285 different LINE-1 insertions [25].

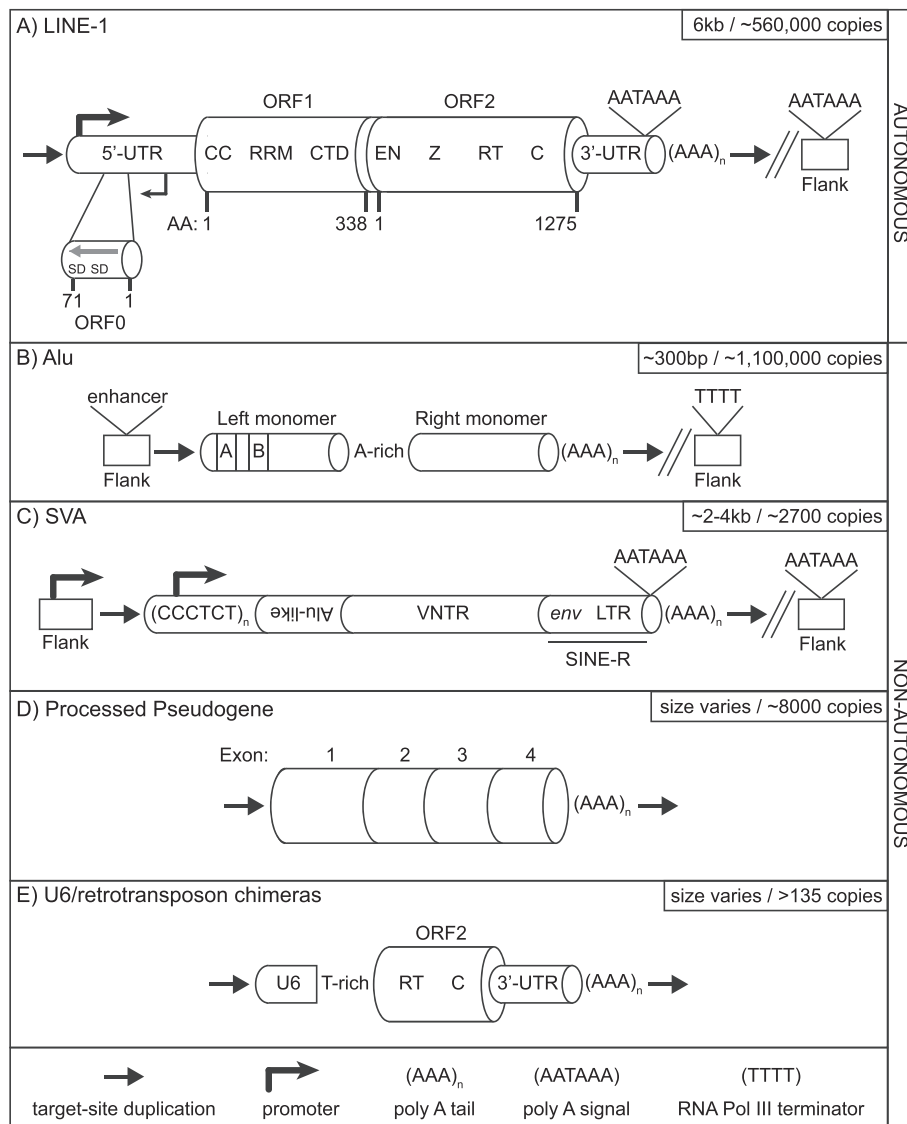
Retrotransposons can be further subdivided into two subclasses: those with Long-Terminal Repeats (LTR) and those without (non-LTR). LTR elements, also known as endogenous retroviruses (ERVs), comprise ~8 % of the human genome [6]. Many of these elements lack a majority of the viral genes and exist only as single LTRs, often referred to as solo LTRs. Similar to DNA transposons, LTR elements are thought to be inactive in the human lineage, although rare polymorphic ERVs in the human population indicate that mobilization has occurred following the human-chimpanzee divergence [26–28]. Very recently, several unfixated HERV-K elements were identified across human genomes including an intact insertion that still may be infectious [29]. In contrast, ERVs have been active recently in the chimpanzee and gorilla lineages [30]. Most ERVs are speculated to be exogenous viruses that integrated into the host germline in the distant past [31, 32]. There is some evidence that endogenous viral elements (EVEs) may have escaped the cell by acquiring a functional envelope gene and that these genetic elements are the ancestors of modern-day retroviruses [33]. Certain hints already exist, but as more genomes are analyzed one might predict that formation of infectious viruses from endogenous elements followed by re-endogenization of exogenous elements might be more common than previously appreciated [34].

### LINE-1

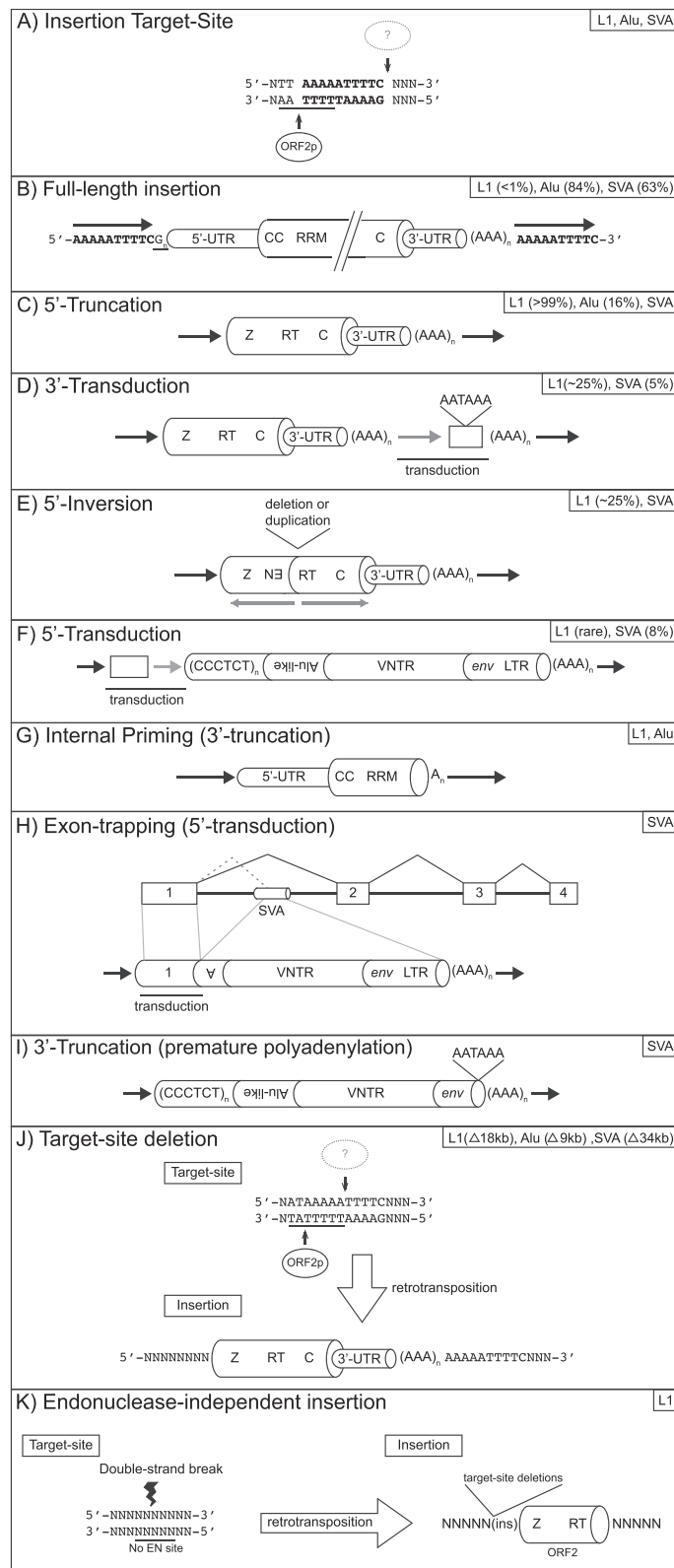
Long Interspersed Element-1 (LINE-1 or L1), a non-LTR element, is the only active autonomous TE in man. Despite the fact that the human genome contains more than 500,000 LINE-1 sequences, most are inactive due to rearrangements, point mutations, and 5'-truncation [6, 35–37]. Only a small subset, 80–100 LINE-1 s, are thought to be active in any given individual [38, 39], with each set of active elements differing between individuals [40]. An active LINE-1 residing in the genome is 6 kb in length [41] (Fig. 1a) contains a 5'- and 3'-UTR, encodes two proteins (i.e. bicistronic), ORF1p and ORF2p, separated by a 63 bp inter-ORF spacer and ends in a long polyA tail. Cell culture retrotransposition assays indicate that both proteins are absolutely required for LINE-1 mobilization in *cis* [42]. ORF1p is a ~40 kDa protein [43] with RNA binding [44, 45] and chaperone activities [46]. Although structural analysis and biochemical studies [47] have revealed that ORF1p forms a series of trimers with nucleic acids [48, 49] via rapid polymerization mediated by coiled-coiled domain interactions, its precise function remains poorly understood; however, new work indicates that phosphorylation of ORF1p is required for retrotransposition [50]. ORF2p is a 150 kDa protein with endonuclease (EN) [51] and reverse transcriptase (RT) [52] activities.

LINE-1 is transcribed from its own promoter [53] located in the ~900 bp 5' UTR presumably by RNA Pol II. LINE-1 RNAs are thought to be capped as evidenced by untemplated guanosines at the 5'-end of full-length genomic insertions [54]. Several transcription factors have been implicated in LINE-1 transcription including ying yang 1 (YY1) [55], T-cell factor/lymphoid enhancer factor (TCF/LEF) [56], p53 [57], and runt related transcription factor 3 (RUNX3) [58]. LINE-1 also contains an antisense promoter in the 5'-UTR [59]. Recently, a novel ORF termed *ORF0*, which is 70 amino acids in length, was identified on the antisense strand of primate LINE-1 5'UTRs [60]. As *ORF0* has two splice donor sites, *ORF0* has the ability to form fusion proteins with downstream exons [60]. Interestingly, overexpression of ORF0p *in trans* results in a 41 % increase in engineered LINE-1 retrotransposition in cell culture [60]. Future research will reveal the role of ORF0p and whether functional homologs have been independently derived in other species.

Transcription of LINE-1 is terminated by an internal weak polyA signal (AATAAA) [42, 61, 62] present in the ~200 bp 3'-UTR. Frequently, LINE-1 transcription will read through its polyA signal in favor of a polyA signal located downstream of the genomic LINE-1 [62–64]. This downstream non-LINE-1 sequence is frequently retrotransposed to new genomic locations, a phenomena referred to as 3'-transduction (Fig. 2). 3'-transductions are an



**Fig. 1** Retrotransposons active in humans. **a** An autonomous active LINE-1. A full-length LINE-1 ~6 kb in length is shown [36, 41, 239]. LINE-1 encodes three proteins, two of which (ORF1p and ORF2p) are absolutely required for retrotransposition *in cis* [42, 146]. Currently, the role of ORF0p is unclear [60]; interestingly, it may form fusion proteins with downstream coding sequences by utilizing internal splice donor sites (SD) [60]. LINE-1 transcription is driven from its own promoter (big black bent arrow) [53, 54] located in the 5'-UTR. The 5'-UTR also encodes a weaker antisense promoter (ASP, small black bent arrow) [59]. It has been postulated that the LINE-1 ASP in conjunction, with splice acceptors located on the antisense strand of LINE-1, may contribute to new gene formation via a mechanism termed "gene-breaking" [240]. Termination of LINE-1 transcription is mediated by a polyA signal (AATAAA) located in the 3'-UTR. Occasionally, transcription proceeds past the internal polyA signal and terminates at a downstream one [139, 241]. Such chimeric transcripts, if retrotransposed, may result in 3'-transductions [42, 62–64, 176]. Majority of insertions end in a polyA tail (AAA)<sub>n</sub> of variable length [37]. In addition, most insertions are characterized by flanking target-site duplications (4–20 bp in length, black horizontal arrows) [35]. CC-coiled coiled domain [47], RRM-RNA recognition motif [44], CTD-C-terminal domain, EN-endonuclease [51], Z domain [242], RT-reverse transcriptase [52], C-cysteine-rich. AA-amino acid. **b** The Alu SINE. Alus are small Pol III transcribed RNAs derived from 7SL RNA [243]. An Alu element consists of a left and right monomer, which are derived from an ancient duplication event, separated by an internal A-rich sequence. Alus contain their own transcriptional signals, an A and B box located in the left monomer. Efficient Alu transcription requires a strong enhancer element in the upstream flanking sequence [103, 104]. Transcription termination of an Alu typically occurs at a Pol III terminator (TTTT) located in the downstream flanking sequence [244]. Similar to LINE-1, Alu insertions end in a polyA tail and are flanked by a target-site duplication. **c** A canonical SINE-VNTR-Alu (SVA) element consisting of its primary domains: CCCTCT hexamer, Alu-like, VNTR, SINE-R derived from the *env* gene and right LTR from a HERV-K is shown [126]. SVA transcription can initiate upstream (black bent arrow) or in the CCCTCT hexamer (black bent arrow) [126, 127]. Like LINE-1, SVA transcription typically terminates at its own [127] or a downstream polyA signal [24, 65]. **d** A processed pseudogene (PP) is shown. Note the lack of introns and the presence of a target-site duplication and a 3'-polyA tail similar to LINE-1, Alu, and SVA. **e** U6 chimera insertion. A U6 snRNA fused with the 3'-end of an LINE-1 sequence formed by "template-switching" [84, 140, 144] is shown. Although the site where ORF2p switches templates varies across the U6 chimera insertions, the junction where the two sequences are joined is typically T-rich [144]



**Fig. 2** (See legend on next page.)

(See figure on previous page.)

**Fig. 2** Anatomy of retrotransposon insertions. A variety of structures for retrotransposon insertions (**a-k**) identified by genomic studies, cell culture retrotransposition assays, and disease-causing insertions that have been reported is shown. Reported frequencies, either from genomic analysis or cell-culture retrotransposition assays, for each structure is located in the upper right hand corner of each panel. If no frequency data has been reported only the element's name is shown. These structures have provided key insights into the mechanism of target-primed reverse transcription [77], retrotransposon transcript structure [127], and the mechanism by which LINE-1-mediated retrotransposition events contribute to genome evolution [62]. **A**) LINE-1 target-site. Most insertions occur at asymmetric AT-rich sequences [6, 37, 51, 86]. The first step of TPRT is cleavage of the bottom-strand by ORF2p endonuclease activity at a motif resembling 5'-TTTT/AA-3' [245]. The nuclease responsible for top-strand cleavage is currently unknown. The nature of the staggered cleavage events generates a target-site duplication (TSD, sequence in bold). **a**) TSD (black horizontal arrows) is used to define the boundaries of an insertion and considered a hallmark of LINE-1-mediated retrotransposition events. **b**) Full-length insertion. It is generally accepted that in order for an element to be retrotransposition-competent it must be full-length. **c**) 5'-truncated insertions. Most LINE-1 s in the human genome are grossly truncated at their 5'-end [6, 36, 37]. In contrast, most Alu [243] and SVA elements are full-length [123, 127]. To date, no consensus sequence has been identified in LINE-1 or SVA insertions regarding the mechanism of 5'-truncation. However, a new report implicates stem-loop structures as a factor driving 5'-truncation in recent Alu insertions [114]. **d**) 3'-transduction. Although the first report of a 3'-transduction was an LINE-1 insertion into the *dystrophin* gene resulting in Duchenne's muscular dystrophy in 1994 [176], it would be several years before the significance of this chimeric insertion was uncovered. Several years later, as one of the first insights gained from insertions recovered from cell-culture retrotransposition assays, it was reported that LINE-1 frequently bypassed its own polyA signal (AATAAA) in favor of a downstream one (AATAAA) [42]. Subsequently, elegant experimental analysis revealed that utilizing a downstream polyA signal could result in LINE-1-mediated exon-shuffling [62]. An insertion containing a 3'-transduction will typically contain two homopolymer stretches (AAA<sub>n</sub>) and contain the 3'-TSD from the source locus (gray horizontal arrow) as part of the transduced sequence. Notably, insertions containing serial 3'-transductions have been reported and can be used to track the evolutionary history of an element [246]. **e**) 5'-end inversions. Another hallmark of LINE-1-mediated retrotransposition events is the inversion of the 5'-end (gray horizontal arrow) of the retrotransposon sequence [35]. Small indels are typically identified at the inversion breakpoint [88]. Inversions have only been reported for LINE-1 s, SVAs, and processed pseudogenes [196]. 5'-end inversion is presumed not to occur for Alus due to their short length. It has been hypothesized that a phenomenon referred to as twin-priming may account for the frequent inversions associated with LINE-1-mediated retrotransposition events [88]. **f**) 5'-transduction. In some instances, LINE-1 [82] or SVA transcription [126, 127] may initiate upstream of the internal promoter generating a chimeric transcript. Retrotransposition of this sequence results in duplication of the sequence 5'- of the source locus at a new genomic location. It has been speculated that 5'-transductions are relatively common for SVA elements due to their weaker internal promoter compared to LINE-1, which has a very strong internal promoter, where only a handful of 5'-transductions have been reported [82]. **g**) Internal priming. Occasionally following bottom-strand cleavage, internal A-rich sequences upstream in the retrotransposon RNA may basepair with the T-rich overhang at the target-site instead of the 3'-polyA tail, followed by first-strand cDNA synthesis by ORF2p [247–249]. These insertions can be deemed a type of 3'-truncation. **h**) Exon-trapping. Retrotransposons are dispersed throughout the genome including intronic sequence. LINE-1, Alu, SVA all have been reported to contain numerous splice sites and be incorporated into the transcriptome [105, 127, 128, 134, 250]. Interestingly, LINE-1 internal splicing can generate a transcript lacking ORF1 but maintaining a functional ORF2 [251]. In some instances, at least for SVA, retrotransposition of chimeric transcripts containing upstream exons may occur [127, 128, 132]. Notably, SVA itself is thought to have originated from alternative splicing from genomic repeats [126] and SVA-related elements (e.g. LAVA, PVA) appear to have acquired distinct 3'-domains via splicing in gibbons [125, 135, 136, 138]. **i**) 3'-truncation. Premature polyadenylation using either canonical or non-canonical polyadenylation sites results in LINE-1 or SVA RNAs lacking 3'-sequence [127, 252]. If this RNA is retrotransposed, it will result in a 3'-truncated insertion. Consistent with the dispensability of SVA domains [130], 3'-truncations may be more frequent for SVA compared to LINE-1. In principle, 3'-truncated LINE-1 RNAs containing ORF1 coding sequence might be actively retrotransposed as in the case of ORF1 *mNEO*i in cell culture [144] and the presence of half-LINE-1 (HAL1) insertions in mammalian genomes [253]. **j**) Target-site deletion. Another surprise from cell culture retrotransposition assays was the discovery of large deletions associated with new retrotransposition events [82, 83]. Genomic deletions up to 1 MB have been associated with LINE-1 mediated retrotransposition events in vivo [153]. These insertions occur at a LINE-1 EN cleavage site, are generated by ORF2 reverse-transcriptase activity, and end in a 3-polyA tail. Currently, the mechanism driving 5-target-site deletions is unclear; yet, it is tempting to speculate that chromatin looping along with cleavage by LINE-1 or another nuclease may play important roles [82, 83]. **k**) Endonuclease-independent (EN<sub>i</sub>) insertion. EN<sub>i</sub> insertions were discovered by the Moran lab when carrying out retrotransposition assays in different Chinese Hamster Ovary (CHO) cell lines lacking key DNA repair factors [213]. Frequent retrotransposition was observed for an engineered LINE-1 element construct, with a catalytically inactive EN, in these cells but not HeLa cells. Characterization of recovered insertions revealed LINE-1 integration at genomic sites not resembling the LINE-1 EN consensus cleavage site. In addition, the insertions were typically truncated at both the 5'- and 3'-ends [213]. These data suggest that LINE-1 can serve as a "molecular band-aid" [254] at double-stranded DNA breaks [213–215] and that LINE-1 s lacking a functional EN domain may be able to retrotranspose in certain contexts. Building on these studies it was later reported that LINE-1 s can also integrate at dysfunctional telomeres in an endonuclease-independent manner [216].

additional mechanism by which LINE-1 contributes to genomic expansion and a means to shuffle protein-coding exons throughout the genome [62, 65].

Following transcription from a genomic locus, the LINE-1 RNA is transported to the cytoplasm for protein translation and LINE-1 ribonucleoprotein (RNP) assembly. Although, the exact nature of LINE-1 ORF1p and ORF2p translation is not entirely resolved, significant insight comes from application of the cell culture retrotransposition assay.

This work suggests that ORF2p is translated via an unconventional mechanism involving translation termination of ORF1 and reinitiation [66]. Surprisingly, this study demonstrated that the codon for any amino acid could serve as the +1 codon for ORF2p.

The next step in the LINE-1 lifecycle is RNP assembly [67]. While the number of ORF1p trimers is thought to be several, the number of ORF2p molecules in an active LINE-1 RNP is unknown but its abundance is thought

to be significantly less when compared to ORF1p in the RNP [68]. In vitro analyses of non-LTR retrotransposon integration predict that at least 2 molecules of ORF2p are present in any given retrotranspositionally-competent (RC) LINE-1 RNP [69]. In addition, a new study has reported that the polyA tail of LINE-1 RNA is required in *cis* for formation of a RC-RNP presumably by serving to recruit ORF2p to the RNP [70]. Similarly, the polyA tail of Alu is also required for reverse transcription [70, 71]. Thus, the basal LINE-1 RNP contains ORF1p trimers, ORF2p, and the LINE-1 RNA. An active area of current research involves determining other components of the LINE-1 RNP, specifically which cellular RNAs [72] and non-LINE-1 proteins [73–76] are present.

LINE-1 insertions occur via a coupled reverse-transcription integration mechanism referred to as target-primed reverse-transcription (TPRT) [77, 78]. TPRT has been characterized in great detail biochemically by Eickbush and colleagues using the *Bombyx mori* non-LTR R2 element as a model. Although R2 differs from LINE-1 in that it only encodes one ORF, this ORF contains endonuclease [79] and reverse transcriptase activities [77]. How LINE-1 identifies a genomic neighborhood for integration remains of great interest. It is highly probable that chromatin states [80] and perhaps protein-protein interactions with nuclear factors dictate target-site preference.

The LINE-1 integration target-site (Fig. 2a) is determined by the ORF2p-encoded endonuclease [51, 81]. Biochemical [51], cell culture retrotransposition assays [42, 82–84], and genomic analysis [6] have revealed the LINE-1 EN consensus site to be 5'-TTTT/AA-3' on the bottom-strand where "/" indicates the site of cleavage. The EN cleavage site is not absolute as variations are common and thus the site can better be defined as 5'-YYYY/RR-3' where Y = pyrimidine and R = purine. The asymmetry of a pyrimidine followed by a purine at the cleavage site is almost always observed. See Table 1 for additional variations (YYRY/RR, YYYY/RR, etc).

The cleavage of the DNA bottom-strand liberates a 3'-OH which will serve as the primer used by ORF2p for reverse-transcription. It is postulated that the T-rich bottom-strand basepairs with the LINE-1 RNA polyA tail and perhaps in some instances, a nuclease activity associated with the LINE-1 RNP processes the 3'-bottom strand to obtain a better primer. 3'-processing can be observed biochemically and for genomic insertions where the cleavage site appears to be absent by annotation (e.g. YYYY/YR), the actual site is merely obscured by this activity [85, 86].

Following bottom-strand cleavage, ORF2p initiates reverse-transcription of the LINE-1 RNA to generate the first strand of LINE-1 cDNA [68, 78]. Cleavage of the DNA top-strand seems to occur following the bottom-

strand nick in a stepwise manner after initiation of first-strand cDNA synthesis [69]. That said, sequence features in some LINE-1 insertions, namely 5'-inversions and target-site deletions, suggest that top-strand cleavage may occur prior to completion of bottom-strand cDNA synthesis. While sequence-specificity for top-strand cleavage has yet to be defined, sequence distance likely plays a role as the majority of target-site duplications are within 4-20 bp in length [6, 37, 84, 86]. A potential suspect for top-strand cleavage could be the additional nuclease activity observed in vitro in LINE-1 RNPs [78, 85]. Next, top-strand cDNA synthesis ensues probably by ORF2p which displays DNA-dependent DNA synthesis activity in vitro [87].

In contrast to DNA transposon and ERV insertions, most LINE-1 insertions are not a full 6 kbp in length (Fig. 2b). The majority of genomic LINE-1 s (>99 %) are grossly truncated at their 5'-end (Fig. 2c) or contain a 5'-inversion (Fig. 2e) of the LINE-1 sequence [37, 88]. Although ~ one-third of the human-specific LINE-1 s are full-length, indicating most full-length elements have been selected against throughout primate evolution [89] and even recently since the human-chimpanzee divergence [90], some LINE-1-containing loci display signatures of positive selection [91]. The lack of LINE-1 RT processivity during cDNA synthesis is unlikely to contribute to short insertions as non-LTR RTs, including ORF2p, are highly processive in vitro [87, 92]. Currently, it is speculated that conflict with host factors, that serve as defenders of the genome against LINE-1 parasites [93, 94] such as apolipoprotein B mRNA editing enzyme catalytic subunit 3A (APOBEC3A) or DNA repair factors [44, 45, 95] like ataxia telangiectasia mutated (ATM), limit the size of a LINE-1 insertion [95]. Ongoing studies will determine whether this conflict interferes with cDNA synthesis resulting in a shortened first-strand cDNA or whether some unknown factor attacks and perhaps cleaves a full-length first-strand cDNA basepaired with the LINE-1 RNA. Conversely, it has not escaped notice that LINE-1 sequences contain numerous sequence motifs resembling the LINE-1 EN cleavage site on what would be deemed the bottom-strand. Although no ribonuclease H (RNaseH) activity has been associated with LINE-1 proteins and perhaps it is counterintuitive, it may be possible that LINE-1 cleaves itself as part of a multifaceted molecular arms-race including but not limited to: 1) AT-rich codon optimization in the ORFs to limit DNA methylation which has the potential to alter coding via cytidine deamination, 2) low ORF2 protein expression to escape innate immunity, and 3) maintenance of only the minimum number of full-length insertions deployed throughout the genome in order to survive and to cloak itself from purifying selection. In contrast, a yet unidentified host-encoded nuclease, such as a factor distantly related to ORF2p with a

**Table 1** Retrotransposition events associated with human disease

	Insertion	Gene	CHR	Reference	Disease	Subfamily	Size	polyA tail length	Truncation	Transduction (bp)	Strand	Exon/ Intron/ Mechanism	Target-site duplication (TSD)	L1 EN site (5'-TTTT/AA-3')	Note
1	Alu	<i>ABCD1</i>	X	Kutsche et al. 2002 [255]	ALD	AluYb9	98	20	Y/5' TR	N	S	4.7 kb Deletion	No TSD	ATTT/GT	
2	Alu	<i>ATP7A</i>	X	Gu et al. 2007 [256]	Menkes Disease	AluYa5a2	282	89	N	N	AS	E	AAAAAGGACAGC	TTTT/AT	
3	Alu	<i>BTK</i>	X	Lester et al. 1997 [257]	XLA	AluY	N/A	N/A	N/A	N	AS	E	N/A	N/A	
4	Alu	<i>BTK</i>	X	Conley et al. 2005 [258]	XLA	AluY	281	74	N	N	S	E	AGAAATGTATGAGTAAAGT	TTCT/AT	Same insertion site Conley et al. SVA
5	Alu	<i>CD40LG</i>	X	Apoil et al. 2007 [259]	HIGM	AluYb8	292	8	N	N	AS	E	AAAAATTTTC	TTTT/AT	
6	Alu	<i>CLCN5</i>	X	Claverie-Martin et. al. 2003 [260]	Dent's Disease	AluYa5	281	50	N	N	S	E	AGAAAATGCTCGAAAGA	TTCT/AT	
7	Alu	<i>CTRC</i>	1	Masson et. al. 2013 [160]	Chronic pancreatitis	Alu	31	11	Y/5' TR	N	AS	53.9 kb Deletion	N/A	TCTT/AT	Deletes entire CTRC and ELA2A genes
8	Alu	<i>PKLR</i>	1	Lesmana et. al. 2015 [159]	Severe Hereditary Nonspherocytic Hemolytic Anemia	Yb8	288	70	N	N	S	E	AAGATCATCAGCAA	TCTT/GA	consanguinity, consensus Yb8
9	Alu	<i>FVIII</i>	X	Sukarova et. al. 2001 [261]	Hemophilia A	AluYb8	290	47	N	N	AS	3 nt Deletion	No TSD	TTTC/AT	
10	Alu	<i>FVIII</i>	X	Ganguly et. al. 2003 [262]	Hemophilia A	AluYb9	288	37	N	N	AS	I/Splicing	AAAAACCAACAGG	TTTT/AT	Consensus Yb9
11	Alu	<i>FVIII</i>	X	Green et. al. 2008 [263]	Hemophilia A	AluYb8	FL	N/A	N	N	AS	E	N/A		
12	Alu	<i>FIX</i>	X	Vidaud et al. 1993 [264]	Hemophilia B	AluYa5a2	244	78	Y/5' TR	N	S	E	AAGAATGGCAGATGCGA	TCTT/AA	Same insertion site as Wulff et al. Alu
13	Alu	<i>FIX</i>	X	Wulff et al. 2000 [265]	Hemophilia B	AluYa5a2	237	39	Y/5' TR	N	S	E	AAGAATGGCAGATGC	TCTT/AA	Same insertion site as Vidaud et al. Alu
14	Alu	<i>FIX</i>	X	Li et al. 2001 [266]	Hemophilia B	AluY	279	40	Y/5' TR	N	AS	E	AAGAAACTGGTCCC	TCTT/AA	
15	Alu	<i>GK</i>	X	Zhang et al. 2000 [267]	GKD	AluYc1	241	74	Y/5' TR	N	AS	I	AAAAATAAG	TTTT/AA	
16	Alu	<i>IL2RG</i>	X	Lester et al. 1997 [257]	XSCID	AluYa5	N/A	N/A	N/A	N	AS	I	N/A	N/A	

**Table 1** Retrotransposition events associated with human disease (Continued)

17	Alu	<i>CRB1</i>	1	den Hollander et al. 1999 [268]	RP	AluY	244	70	Y/5'TR	N	AS	E	AAGAGTAAAGATGA	TCTT/GA	
18	Alu	<i>SERPINC1</i>	1	Beauchamp et al. 2000 [269]	Type 1 ATP	Alu	6	40	Y/5'TR	N	AS	1.4 kb Deletion	N/A	TTCT/AT	Shortest Alu insertion
19	Alu	<i>ALMS1</i>	2	Taşkesen et al. 2012 [270]	Alström syndrome	AluYa5	257	76	Y/5'TR	N	S	E	AAAAGCCTAGAGAA	TTTT/AA	
20	Alu	<i>MSH2</i>	2	Kloor et al. 2004 [271]	HNPCC	AluJ	85	40	Y/5'TR	N	S	E	N/A	N/A	Contains extra 99 nt 3'-of Alu, may be transduction or recombination
21	Alu	<i>MSH2</i>	2	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	N/A	Pan-cancer panel testing
22	Alu	<i>ZFX1B</i>	2	Ishihara et al. 2004 [272]	MWS	AluYa5	281	93	N	N	S	E	AAAATTAACA	TTTT/AA	
23	Alu	<i>BCHE</i>	3	Muratani et al. 1991 [273]	Cholinesterase deficiency	AluYb9	289	38	N	N	S	E	AAAAATTTTTTCC	TTTT/AA	
24	Alu	<i>CASR</i>	3	Janicic et al. 1995 [274]	FHH and NSHPT	AluYa5	280	93	N	N	AS	E	GAAAGCGTGAGCTGC	TTTC/AA	
25	Alu	<i>HESX1</i>	3	Sobrier et al. 2005 [275]	Anterior Pituitary Aplasia	AluYb8	288	30	N	N	S	E	AGAAAATGCTTTAGA	TTCT/AA	
26	Alu	<i>OPA1</i>	3	Gallus et al. 2010 [276]	ADOA	AluYb8	289	25	N	N	AS	I/Splicing	AAAAATTTAAAAAGTT	TTTT/AC	
27	Alu	<i>MLV12</i>	5	Economou-Pachnis and Tschlis 1985 [277]	Associated with leukemia	AluYa5	280	26	N	N	AS	I	GAAAATGT	TTTC/AT	
28	Alu	<i>APC</i>	5	Halling et al. 1999 [278]	Hereditary desmoid disease	AluYb8	278	40	Y/5'TR	N	S	E	AAGAATAATG	TCTT/AA	Same insertion site as Miki et al. L1
29	Alu	<i>APC</i>	5	Su et al. 2000 [279]	FAP	AluYb9	93	60	Y/5'TR	N	AS	I/Splicing	No TSD	TTTT/AA	1.6 kb intronic deletion
30	Alu	<i>APC</i>	5	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	N/A	Pan-cancer panel testing
31	Alu	<i>MAK</i>	6	Tucker et al. 2011 [280], Edwin Stone, personal communication	RP	AluYb8	281	57	N	N	AS	E	AAAGAAAAA	CTTT/AA	Identified by exome resequence
32	Alu	<i>NT5C3</i>	7	Manco et al. 2006 [281], Leticia Ribeiro, personal communication	Chronic hemolytic anemia	Alu Ya5	281	36	N	N	S	E	AAGAAATGGCAGATGG	TCTT/AA	



**Table 1** Retrotransposition events associated with human disease (*Continued*)

33	Alu	<i>CFTR</i>	7	Chen et al. 2008 [282]	Cystic Fibrosis	AluY	46	57	Y/5'TR	N	AS	E	AAGAATCCCACCTATAAT	TCTT/AA	
34	Alu	<i>CFTR</i>	7	Chen et al. 2008 [282]	Cystic Fibrosis	AluYa5	281	56	N	N	S	E	AATAGAAATGATTTTTGTC	TCTC/AT	3'-Processing of (5'-CTC-3')
35	Alu	<i>EYA1</i>	8	Abdelhak et al. 1997 [283]	BOR syndrome	AluYa5	n/a	97,31	N/A	N	AS	E	AAAAAATAAATGTGTG	TTTT/AA	PolyA tail shortening between generations
36	Alu	<i>LPL</i>	8	Okubo et al. 2007 [284]	LPL deficiency	AluYb9	150	60	Y/5'TR	N	AS	2.2 kb Deletion	No TSD	TTTT/AA	
37	Alu	<i>CHD7</i>	8	Udaka et al. 2007 [285]	CHARGE syndrome	AluYa5/8	75	100	Y/5'TR	N	S	10 kb Deletion	No TSD	ATTT/AA	
38	Alu	<i>POMT1</i>	9	Bouchet et al. 2007 [286]	Walker Warburg syndrome	AluYa5	290	53	N	N	AS	E	AAAAAGAGATGTACTG	TTTT/AC	
39	Alu	<i>FGFR2</i>	10	Oldridge et al. 1999 [287]	Apert syndrome	AluYa5	283	69	N	N	AS	I/Splicing	AGAAAACAAGGGAAGCA	TTCT/AG	
40	Alu	<i>FGFR2</i>	10	Oldridge et al. 1999 [287]	Apert syndrome	AluYb8	288	47	N	N	AS	E	AGAATTACCCGCCAAG	TTCT/AT	
41	Alu	<i>FGFR2</i>	10	Bochukova et al. 2009 [288]	Apert syndrome	AluYk13	214	12	Y/5'TR	N	AS	E	AAAAGTTACATTCGG	TTTT/GA	
42	Alu	<i>FAS</i>	10	Tighe et al. 2002 [289]	ALPS	AluYa5	281	33	N	N	AS	I	AGAATATTCTAAATGTG	TTCT/AA	
43	Alu	<i>SERPING1</i>	11	Stoppa-Lyonnet et al. 1990 [290]	HAE	AluYc1	285	42	N	N	S	I	AAAAATACAAAAATTAG	TTTT/AG	
44	Alu	<i>HMBS</i>	11	Mustajoki et al. 1999 [291]	AIP	AluYa5	279	39	N	N	AS	E	AAGAATCTTGTC	TCTT/GA	
45	Alu	<i>ATM</i>	11	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing
46	Alu	<i>GNPTAB</i>	12	Tappino et al. 2008 [292]	ML II	AluYa5	279	17	N	N	AS	E	AAAAACAACAACCTGAG	TTTT/GA	
47	Alu	<i>BRCA2</i>	13	Miki et al. 1996 [293]	Breast Cancer	AluYc1	281	62	N	N	S	E	AATCACAGGC	GATT/AT	
48	Alu	<i>BRCA2</i>	13	Teugels et al. 2005 [294]	Breast Cancer	AluYa5	285	N/A	N	N	S	E	AAGAATCTGAACAT	TTCT/GC	3' Processing 2 nt (5'-CT-3')
49	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing
50	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing
51	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing

**Table 1** Retrotransposition events associated with human disease (*Continued*)

52	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
53	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
54	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
55	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
56	Alu	<i>BRCA2</i>	13	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
57	Alu	<i>PMM2</i>	16	Schollen et al. 2007 [295]	CDG-Ia	AluYb8	263	10	Y/5'TR	N	AS	28 kb Deletion	No TSD	TTTT/AA	
58	Alu	<i>PALB2</i>	16	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
59	Alu	<i>BRCA1</i>	17	Peixoto et al. 2013 [161]	Breast and Ovarian Cancer Family	AluYc	191	60	Y/5'TR	N	AS	23.3 kb Deletion	No TSD	CTTT/AG	
60	Alu	<i>BRCA1</i>	17	Teugels et al. 2005 [294]	Breast Cancer	AluS	286	N/A	N	N	S	E	GAAAAAGAATCTGCTTT	TTTC/GA	
61	Alu	<i>BRCA1</i>	17	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing	
62	Alu	<i>NF1</i>	17	Wallace et al. 1991 [23]	NF1	AluYa5	282	40	N	N	AS	I/Splicing	AAAAAAAAAACAT	TTTT/AA	First report of de novo Alu insertion
63	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluY	280	N/A	N	N	S	I	AAAAAATTCAG	TTTT/AA	Same insertion site as Wimmer et al. <sup>a</sup>
64	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluY	281	N/A	N/A	N	AS	I	N/A		
65	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYa5	282	60	N	N	S	E	ATAAATAGCCTGGA	TTAT/AA	
66	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYa5	284	120	N	N	AS	E	AAAAAACTTGCT	TTTT/GA	Same insertion site as Wimmer et al. <sup>c</sup>
67	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYa5	281	N/A	N	N	AS	E	AAAAAACTTGCTGATGG	TTTT/GA	Same insertion site as Wimmer et al. <sup>c</sup>
68	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYa5	284	110	N	N	AS	E	AATAAACCTAAAGA	TATT/GA	
69	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYa5	279	N/A	N	N	S	E	AAAAGAAGACATAT	TTTT/GT	Same insertion site as Wimmer et al. <sup>b</sup>
70	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYa5	264	60-85	Y/5'TR	N	AS	E	AAGAAGTCCGGTACCT	TCTT/GA	

**Table 1** Retrotransposition events associated with human disease (*Continued*)

71	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYb8	249	121	Y/5'TR	N	S	E	AAAGCAGTGC	CTTT/AT	
72	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYb8	288	N/A	N	N	AS	I	AAAAAAGAGAAAGACAA	TTTT/AA	Same insertion site as Wimmer et al. <sup>a</sup>
73	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYb8	289	120	N	N	AS	E	AACAATGGTCTT	TGTT/AA	
74	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYb8	288	78-178	N	N	S	E	AAACAATGATGTTA	TTTC/AA	3' Processing of 1 nt (C)
75	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYb8	288	118	N	N	S	E	AAAAGAAGACATAT	TTTT/GT	Same insertion site as Wimmer et al. <sup>b</sup>
76	Alu	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	AluYb8	268	121	Y/5'TR	N	AS	I	AAAAACAACAAACA	TTTT/GT	
77	L1	<i>CYBB</i>	X	Meischl et al. 1998 [297], Brouha et al. 2002 [181]	CGD	L1 Ta	1722	101	Y/5'TR	Y (280)	S	E	AA	TGTT/GA	Maternal Meiosis I
78	L1	<i>CYBB</i>	X	Meischl et al. 2000 [298]	CGD	L1 Ta	836	69	Y/5'TR/INV	N	S	I/Splicing	AGAAATAACTATTTAA	TTCT/AA	
79	L1	<i>CHM</i>	X	van den Hurk et al. 2003 [177]	Choroideremia	L1 Ta	6017	71	FL	Y (119/406)	AS	E	AGAAGATCAATTAG	TTCT/AA	Insertion in Early Development
80	L1	<i>DMD</i>	X	Musova et al. 2006 [299]	DMD	L1 Ta	452	41	Y/5'TR/INV	N	AS	E	AAATATCTTTATATCA	ATTT/AA	
81	L1	<i>DMD</i>	X	Narita et al. 1993 [164]	DMD	L1 Ta	608	16	Y/5'TR	N	AS	E	No TSD	TCTT/AA	2 nt deletion
82	L1	<i>DMD</i>	X	Holmes et al. 1994 [176]	DMD	L1 Ta	1400	38	Y/5'TR/INV	Y(489)	S	E	AAATCATCTGCTGCT	ATTT/AA	First Report of L1 3'TR
83	L1	<i>DMD</i>	X	Yoshida et al. 1998 [300]	XLDCM	L1 Ta	530	73	Y/5'TR	N	AS	5'-UTR/Loss of mRNA	AAAAAAAACCTGGTAAA	TTTT/AT	Tissue specific loss of mRNA
84	L1	<i>DMD</i>	X	E Bakker & G van Omenn, personal communication	DMD	N/A	878	N/A	Y/5'TR	N	S	N/A	N/A	N/A	
85	L1	<i>DMD</i>	X	Awano et al. 2010 [301], Solyom et al. 2011 [302]	DMD	L1 Ta	212	118	Y/5'TR	Y (212)	AS	E	GAA	TTTC/AA	Orphan 3'-transduction
86	L1	<i>FVIII</i>	X	Kazazian et al. 1988 [22]	Hemophilia A	L1 Ta	3800	54	Y/5'TR	N	S	E	AAAGACAAACAAAC	CTTT/AA	First report of de novo L1 insertion
87	L1	<i>FVIII</i>	X	Kazazian et al. 1988 [22]	Hemophilia A	L1 preTa	2300	77	Y/5'TR/INV	N	AS	E	AATGTTTCCTTCTTTTC	CATT/AA	
88	L1	<i>FIX</i>	X	Li et al. 2001 [266]	Hemophilia B	L1 Ta	463	68	Y/5'TR	N	S	E	AAAAATAGTGCTGATA	TTTT/AC	
89	L1	<i>FIX</i>	X	Mukherjee et al. 2004 [303]	Hemophilia B	L1 Ta	163	125	Y/5'TR	N	S	E	GAAAAATGGATTGT	TTTC/AT	

**Table 1** Retrotransposition events associated with human disease (*Continued*)

90	L1	<i>RP2</i>	X	Schwahn et al. 1998 [304]	XLRP	L1 Ta	6000	64	FL	N	S	I/Loss of mRNA	AAGACTGTAAGGTG	TCTT/AA	Interrupted polyA
91	L1	<i>RPS6KA3</i>	X	Martinez-Garay et al. 2003 [305]	Coffin-Lowry syndrome	L1 Hs	2800	Yes	Y/5*TR/INV	N	AS	E	AAGAAAACCTGCATT	TCTT/AG	
92	L1	<i>ABDH5</i>	3	Samuelov et al. 2011 [306], Eli Sprecher, personal communication	CDS	N/A	FL	N/A	N	N/A	N	I/Splicing	N/A	N/A	
93	L1	<i>MLH1</i>	3	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing
94	L1	<i>MLH1</i>	3	Qian et al. 2015 [158]	Hereditary Cancer	N/A	N/A	N/A	N/A	N/A	N/A	E	N/A	N/A	Pan-cancer panel testing
95	L1	<i>APC</i>	5	Miki et al. 1992 [200]	Colon cancer	L1Ta	520	222	Y/5*TR/INV	N	S	E	AAGAATAATG	TCTT/AA	Somatic Insertion/same insertion site as Halling et al. Alu
96	L1	<i>EYA1</i>	8	Morisada et al. 2010 [247]	BOR syndrome	L1 Hs	3756	None	Y/3*TR	N	AS	17 kb Deletion	No TSD	TCTC/AG	Internal Priming
97	L1	<i>FKTN</i>	9	Kondo-Iida et al. 1999 [307]	FCMD	L1Ta	1200	59	Y/5*TR	N	S	I/Splicing/6 nt Deletion	No TSD	TTTT/AA	
98	L1	<i>SETX</i>	9	Bernard et al. 2009 [308], Christine Zühlke, personal communication	AOA2	L1 Hs	1300	42	Y/5*TR/INV	N	S	E	GGAAGAATGTGAACTGGCTA	TTCC/AG	3'-processing 2 nt (5'-CC-3')
99	L1	<i>PTEN</i>	10	Helman et al. 2014 [201]	endometrial carcinoma	L1 Hs	90	22	Y/5*TR	N	S	E	AAAGAATCATCTGGATTATAG	CTTT/AA	Somatic Insertion
100	L1	<i>HBB</i>	11	Divoky et al. 1996 [309]	$\beta$ -thalassemia	L1 Ta	6000	107	FL	N	AS	I	AAAATAAAGCAGA	TTTT/AT	
101	L1	<i>PDHX</i>	11	Mine et al. 2007 [310]	PDHc deficiency	L1 Hs	6086	67	FL	N	S	46 kb Deletion	No TSD	TTTT/AT	
102	L1	<i>SLCO1B3</i>	12	Kagawa et al. 2015 [157]	Rotor syndrome	L1 Ta-1d	5989	100	Near FL	N	S	I/Splicing	AAGAATTAATAGTGACAGT	TCTT/AC	0.054 Japanese Allele Frequency, may be "Hot L1"
103	L1	<i>RB1</i>	13	Rodriguez-Martin et al. 2016 [202]	Familial Retinoblastoma	L1 Ta-1d	6044	33	FL	N	S	I/Splicing	AAATTATCTGTTTC	ATTT/AA	N/A
104	L1	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	L1 preTa	1800	N/A	Y/5*TR	N	S	E	AAAAACGAAACTGTGT	TTTT/AT	Untemplated 3'-T?
105	L1	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	L1 Ta	6000	N/A	FL	N	S	E	AAAAATCGAGGG	TTTT/AA	Untemplated 3'-T?
106	L1	<i>NF1</i>	17	Wimmer et al. 2011 [296]	NF1	N/A	2200	N/A	Y/5*TR/INV	N	AS	I/Splicing	AAGAAAATGGT	TCTT/AA	

**Table 1** Retrotransposition events associated with human disease (Continued)

107	SVA	<i>BTK</i>	X	Rohrer et al. 1999 [311], Conley et al. 2005 [258]	XLA	N/A	251	92	Y/5' TR	N	S	E	AGAAATGTATGAGTAA	TTCT/AT	Same insertion site as Conley et. al. Alu
108	SVA	<i>TAF1</i>	X	Makino et al. 2007 [312]	XDP	F	2627	62	FL	N	AS	I	AAAAAAAAAAAAATGAAATAG	TCCT/AT	3'-Processing 3 nt (5'-CCT-3')
109	SVA	<i>FIX</i>	X	Nakamura et al. 2015 [156]	Hemophilia B	F	2524	28	FL	N	AS	E	AAATGGCACTAGAA	TTCC/AT	3'-Processing 1 nt (5'-C-3')
110	SVA	<i>LDRAP1</i>	1	Wilund et al. 2002 [313]	ARH	E	2600	57	FL	N	S	I/Splicing	GAAACCTGTTTTCTC	TTTC/AA	
111	SVA	<i>SPTA1</i>	1	Hassoun et al. 1994 [314], Ostertag et al. 2003 [24]	HE and HPP	E	632	50	Y/5' TR/INV	Y (183/599)	S	E	GAAATTTGAAGACTTCCAAGT	TTTC/AA	Orphan 3'-transduction
112	SVA	<i>CASP8</i>	2	Stacey et al. 2016 [203]	Breast Cancer Susceptibility	E	2782	N/A	FL	N	AS	I/Decreased RNA	AAGAATTTGA	TCTT/AT	Protective against prostate cancer; active locus?
113	SVA	<i>A4GNT</i>	3	Nazaryan et al. 2015 [155]	Chromothripsis	E	502	None	Y/5' TR (VNTR)	N	AS	I	N/A	TTTT/GA	First report of large scale rearrangement and an insertion. Implicates retrotransposition in germline chromothripsis.
114	SVA	<i>HLA-A</i>	6	Takasu et al. 2007 [315]	Leukemia	F1	2000	45	FL	N/A	AS	14 kb Deletion	N/A	CCTT/AG	Novel SVA subfamily (F1)
115	SVA	<i>PMS2</i>	7	van der Klift et al. 2012 [154]	Lynch syndrome	F	2200	64	Y/5' TR (VNTR)	N	S	I/Splicing	AAGAATGTGCCATGTGA	TCTT/AA	SVA exonization
116	SVA	<i>FKTN</i>	9	Kobayashi et al. 1998 [162]	FCMD	E	3023	32	FL	N	S	3'UTR/Splicing	AAGAAAAAAAAAATTGT	TCTT/AA	
117	SVA	<i>PNPLA2</i>	11	Akman et al. 2010 [316]	NLSDM	E	1800	44	Y/5' TR	N	S	E	AAAGAGGCCCGG	CTTT/AG	
118	SVA	<i>SUZ1P</i>	17	Vogt et al. 2014 [153]	NF1	F1	1700	23	Y/5' TR (VNTR)	Y (282/160)	AS	I/Deletion of NF1	N/A	TTTT/AC	Largest reported insertion associated deletion (~1 Mb), somatic
119	SVA	<i>SUZ1P</i>	17	Vogt et al. 2014 [153]	NF1	F	1300	40	Y/5' TR (VNTR)	N	AS	I/Deletion of NF1	N/A	CTTT/AC	867 kb deletion, somatic
120	Processed Pseudogene	<i>CYBB</i>	X	de Boer et al. 2014 [152]	CGD	N/A	5739	100	FL	No	AS	I/Splicing	AAAACCTCAAAGACTC	TTTT/AA	First reported de novo processed pseudogene (TMF1)

**Table 1** Retrotransposition events associated with human disease (*Continued*)

121	pA	COL4A6	X	Segal et al. 1999 [317]	Alport syndrome	N/A	N/A	70	N/A	N/A	AS	13.4 kb Deletion	No TSD	TTCT/AT
122	pA	AGA	4	Jalanko et al. 1995 [318]	AGU	N/A	N/A	37	N/A	N/A	AS	2 kb Deletion	No TSD	TTCT/AA
123	pA	BRCA2	13	Wang et al. 2001 [319]	Breast Cancer	N/A	N/A	35	N/A	N/A	S	6.2 kb Deletion	No TSD	TTCT/AA
124	pA	NF1	17	Wimmer et al. 2011 [296]	NF1	N/A	130	120	N/A	N/A	AS	E	AAGAAA	TCTTNA

Data for this table were compiled from the primary references listed and reports prior to 2009 are reviewed in the following: Ostertag and Kazazian 2001 [35], Chen et al. 2006 [150], Belancio et al. 2008 [151], Hancks and Kazazian 2012 [86]

A few insertions were left off the list as they were common polymorphisms or did not cause disease. The following websites and databases were used in the analysis: <http://www.repeatmasker.org/>, Repbase (<http://www.girinst.org/>), <http://dbrip.brocku.ca/>, The following symbols, <sup>a</sup>, <sup>b</sup>, <sup>c</sup>, indicate same insertion site in Wimmer et al. [296]

**Abbreviations:** TR truncation, INV inversion, E exon, FL full-length, I intron

**Disease acronyms:** ADOA Autosomal dominant optic atrophy, AGU Aspartylglucosaminuria, AIP Acute intermittent porphyria, ALD Adrenoleukodystrophy, ALPS Autoimmune lymphoproliferative syndrome, AOA2 Ataxia with oculomotor apraxia 2, ARH Autosomal recessive hypercholesterolemia, BOR Branchio-oto-renal syndrome, CDG-Ia Congenital disorders of glycosylation type Ia, CDS Chananin-Dorfman syndrome, CGD Chronic granulomatous disease, DMD Duchenne muscular dystrophy, FAP Familial adenomatous polyposis, FCMD Fukuyama-type congenital muscular dystrophy, FHH and NSHPT Familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism, GKD Glycerol kinase deficiency, HAE Hereditary form of angioedema, HE and HPP Hereditary elliptocytosis and hereditary pyropoikilocytosis, HIGM Hyper-immunoglobulin M syndrome, HNPCC Hereditary non-polyposis colorectal cancer syndrome, LPL Lipoprotein lipase, MLIH Mucopolipidosis Type II, MWS Mowat-Wilson syndrome, NF1 Neurofibromatosis Type I, PDHc Pyruvate dehydrogenase complex deficiency, NLSDM Neutral lipid storage disease with subclinical myopathy, RP Retinitis pigmentosa, Type 1 ATP Type 1 antithrombin deficiency, XDP X-linked dystonia-parkinsonism, XLA X-linked agammaglobulinemia, XLDCM X-linked dilated cardiomyopathy, XLRP X-linked retinitis pigmentosa, XSCID X-linked severe combined immunodeficiency

preference for AT-rich motifs may in part explain LINE-1 5'-truncations. Overall, many of the key steps in LINE-1 retrotransposition have been defined; yet, gaps still exist in our understanding of this selfish gene's lifecycle.

### **Trans-mobilization of non-autonomous elements**

#### **Alu elements**

In addition to mobilizing its own RNA, LINE-1 proteins retrotranspose a myriad of other RNAs. For instance, the most abundant retrotransposon in the human genome by copy number is the Short INterspersed Element (SINE) Alu [6]. Its name originates from human DNA renaturation studies which identified an abundant ~300 repetitive nucleotide sequence that contained the *AluI* restriction endonuclease cleavage site [96]. Alu RNAs are primate-specific [97, 98] non-coding RNA Pol III transcripts [98] derived from the 7SL RNA [99, 100], a component of the signal recognition particle. While Alu elements contain their own transcriptional signals [101] (A and B box, and Pol III terminator (TTTT)) [102] and end in a polyA tail of varying length, transcriptional robustness is largely dictated by the presence of upstream enhancer elements [103, 104] (Fig. 1b). The Alu polyA tail, which is part of the element, differs from the LINE-1 polyA, which is presumably added via the canonical polyadenylation pathway. Along with being transcribed via internal signals, these elements are frequently incorporated into the transcriptome via exonization [105]. An antisense Alu element contains certain sequence features that poise this SINE for splicing. In particular, a pyrimidine-rich tract is generated by the reverse complement of the polyA tail along with a CAG trinucleotide motif, which together generate a very strong splice acceptor motif.

The Alu's evolutionary origins provide insight into how it has become the most abundant retrotransposon in the human genome. Namely, evolution from 7SL RNA [100] followed by monomer duplication [106–108], which increased SRP9/14 binding sites, coupled with increased protein levels of SRP 14 due to triplet repeat expansion seeded by a point mutation in the anthropoid ancestor [109], enhanced localization of this non-coding RNA to the ribosome where it can hijack the LINE-1 protein machinery [110]. Cell-culture retrotransposition assays and mutational analysis by Devine and colleagues have shown that Alus with less secondary structure similarity to 7SL have decreased LINE-1-mediated retrotransposition [111]. Structural analysis has revealed Alu in complex with the SRP 9/14 proteins [110, 112].

Following incorporation into the LINE-1 RNP, Alu integration likely follows in a fashion similar to LINE-1. Based on *trans*-mobilization cell culture assays, engineered Alu elements require a polyA tail and appear to only need transfected LINE-1 ORF2 for retrotransposition [71]. Subsequently, it was demonstrated that transfecting increasing

amounts of an ORF1 plasmid enhances Alu retrotransposition [113]. One might infer from these data that endogenous ORF1 and ORF2 generated from distinct LINE-1 elements could serve to retrotranspose Alus.

In contrast to LINE-1, most Alus are full-length. However, 5'-truncated Alu elements have been identified in human genomes [114] and as *de novo* insertions resulting in disease (Table 1). Over evolutionary time, Alus appear to be more tolerated than LINE-1 in introns [115], which may be due to their decreased effectiveness over evolutionary time in mediating ectopic homologous recombination when compared to LINE-1. Alus are also commonly found in 3'-UTRs where they may serve as small RNA binding sites [116, 117] or serve as substrates for the RNA editing enzyme adenosine deaminase acting on RNA (ADAR) when at least two inverted Alus are present [118–120]. Also, base-pairing between Alus embedded in mRNA 3'-UTRs and long non-coding RNAs can be involved in directing Staufen-mediated RNA decay [121].

#### **SVA elements**

The youngest active human retrotransposon is named after the sum of its parts SINE-VNTR-Alu (SVA). SVA elements are ~2 kb hominid-specific non-coding composite elements [24, 122, 123]. The structure of an SVA (Fig. 1c) [124–126], starting from its 5'-end, is 1) a CCCTCT mostly pure repeat ranging from a few copies up to a hundred (also known as the hexamer), 2) an Alu-like domain derived from two Alu antisense fragments, 3) a variable number of very GC-rich tandem repeats (VNTR), 4) a SINE-R domain sharing sequence homology to the *env* gene and right LTR from a HERV-K, and 5) a polyA tail of varying length similar to LINE-1. From element to element within the human genome, these sequences display more structural sequence variation than LINE-1 and Alu [125–127], primarily because of changes in hexamer and VNTR copy number along with 5'- [127, 128] and 3'- transductions [24, 65]. There are approximately 2700 SVA elements in the human genome reference sequence [123] which differs dramatically from LINE-1 and Alu copy number, ~500,000 and ~1,000,000 copies, respectively. Due to its more recent discovery relative to LINE-1 and Alu elements, less is known about SVA biology.

The nature of the SVA transcriptional unit or SVA promoter has not been completely resolved, but SVAs are presumed to be Pol II transcripts due to the poly A tail downstream of a canonical polyA signal (AATAAA). Furthermore, untemplated guanosines [127, 129, 130] have been identified at the 5'-end of full-length insertions, similar to LINE-1, which likely represent reverse-transcription of the 7mG cap. Initiation of SVA transcription can be broadly grouped into 4 classes [126–128]: 1) transcription

initiation from within the hexamer, 2) transcription initiation downstream of the hexamer, 3) transcriptional initiation 5' of the SVA, which can lead to retrotransposition of upstream sequences (e.g. 5'-transduction) (Fig. 2f), and 4) transcription initiation in an upstream exon followed by splicing into SVA which results in a chimeric transcript (Fig. 2h).

It is currently unclear how or where SVA RNA interacts with the LINE-1 proteins. It has been predicted that perhaps the SVA RNA is localized to the ribosome [24, 131] via base-pairing interactions between the SVA Alu-like domain and Alu RNAs. This may be the case for some SVAs, however the recent discovery of a human-specific SVA subfamily generated via splicing from the first exon of the microtubule associated serine/threonine kinase 2 (*MAST2*) gene into the 3'-end of the Alu-like domain [127, 128, 132], suggests that basepairing with Alu at the ribosome is not a requirement. SVAs require ORF2p for retrotransposition [24, 129, 133], whereas the requirement for ORF1p is less clear, in part, because the contribution of endogenous ORF1p for engineered SVA retrotransposition is unknown. Cell-culture retrotransposition assays and deletion analysis indicate that SVAs require the 5'-end (hexamer and Alu-like domain) of the element to retrotranspose [130, 133]. Additional requirements for LINE-1-mediated retrotransposition are currently being investigated using a comparative strategy involving cell culture assays [134, 135] with the newly identified composite VNTR elements like LAVA in gibbons [136–138].

Once an SVA RNA is incorporated into the RNP, integration is hypothesized to occur in a similar manner to LINE-1. SVA insertions share many similarities to LINE-1. Other than typical LINE-1 hallmarks (target-site duplication, insertion at LINE-1 EN site, end in polyA tail), both LINE-1 and SVA insertions occasionally contain 3'-transductions [24, 62, 65, 139]. Some SVA insertions differ from classical LINE-1 insertions by containing 5'-transductions, which are almost non-existent for LINE-1 [6, 82] yet ~10 % of all SVAs contain transduced sequence via upstream transcriptional initiation [127, 128]. Unlike LINE-1 s, SVAs are occasionally 3'-truncated due to premature polyadenylation mediated by polyA signals located in the SINE-R [127]. Another major difference between LINE-1 and SVA insertions, both in the genome [127] and those recovered from cell culture retrotransposition assays [129, 130, 133], is that most SVAs are full-length while less than 5 % of LINE-1 s recovered from cell culture assays are full-length [6, 84]. Indeed, full-length SVAs and LINE-1 do differ in length (SVAs range from ~1 kb to almost 5 kb and LINE-1 = 6 kb), yet many if not most LINE-1 genomic insertions are under 1.5 kb [6]. A notable difference between LINE-1 and SVA is the sequence composition; LINE-1 s are very AT-rich while SVAs are very GC-rich.

### **Retrotransposition of splicesomal RNAs**

In addition to Alu and SVAs, other RNAs encoded by the genome can be retrotransposed. Many small spliceosomal RNAs are commonly integrated by LINE-1 into genomes with U6 being the most frequent (Fig. 1e) [140–144]. Interestingly, U6 retrotransposition events are often characterized as chimeric insertions [140, 144]. Specifically most are fused with a LINE-1 or an Alu element at the 3'-end of the U6 sequence [144]. In addition to evidence from the genome, chimeric U6-LINE-1 insertions have been identified and investigated using cell culture retrotransposition assays [84, 144].

### **Retrotransposition of protein-coding transcripts**

Along with small abundant RNAs, LINE-1 can mobilize protein-coding RNAs [145, 146]. Following retrotransposition, these insertions are referred to as processed pseudogenes (PPs) due to their lack of introns (Fig. 1d). PPs contains all of the hallmarks of LINE-1-mediated retrotransposition (target-site duplications, 5'-truncations, 5'-end inversions, end in a polyA tail of variable length). The reference human genome sequence contains approximately ~8000 processed pseudogenes [147] with the most abundant being sequences encoding ribosomal protein RNAs [148]. Retrotransposition is thought usually to inactivate PPs due to the loss of regulatory elements such as promoter sequences.

### **Mechanisms by which retrotransposons can cause disease**

Retrotransposons can potentially cause disease by a variety of mechanisms [149]. Most of the 124 disease-causing insertions [35, 86, 150–161] reported to date inactivate gene function through insertional mutagenesis or aberrant splicing. Indeed, the precise mechanism of gene inactivation may be more detailed. For example, a LINE-1 insertion into an exon or an intronic insertion that is spliced into may result in a frameshift mutation that will result in nonsense-mediated RNA decay. In contrast, depending on the site of insertion, the retrotransposon may result in an alternative C-terminus of a protein which can in turn alter function of the gene leading to disease. This is best exemplified by the SVA element insertion in the fukutin (*FKTN*) gene which causes fukuyama muscular dystrophy [162, 163]. Here, alternative splicing of the *FKTN* mRNA into the SVA located in the 3'-UTR generates a protein that is mislocalized from the Golgi to the endoplasmic reticulum [163].

Another major mechanism by which LINE-1-mediated insertions result in disease is through target-site deletions (Fig. 2j) [82, 83]. Deletions associated with *de novo* LINE-1-mediated insertions range from a few basepairs [164] up



to a megabase [153]. LINE-1 [165], Alu [166], and SVA [167] associated target-site deletions have also been identified in the human and primate genomes. Thus, these deletions in the short-term may result in disease but may serve as a means by which retrotransposons contribute to genome evolution. Notably, retrotransposon sequences can also generate genetic deletions via non-allelic homologous recombination (NAHR) [168] which is independent of TPRT and DNA breakage mediated by LINE-1 ORF2p. NAHR is most frequently observed for Alu elements presumably due to their high copy-number and results in structural variation which can lead to genetic disease. These deletions may be generated via mispairing of two retrotransposon sequences on the same strand usually on homologous chromosomes, while crossing over between two retrotransposon sequences inverted relative to each other may result in an inversion [149, 169].

Additional hypothesized mechanisms by which new LINE-1, Alu, and SVA element insertions may disrupt gene function relate to epigenetic changes at the site of integration. All three elements are known to be methylated at CpGs. LINE-1 [170, 171] and SVA DNA [172], in the 5'-UTR and VNTR, respectively, are known to be densely methylated in somatic tissue. Interestingly, SVAs were initially identified by one group when carrying out a restriction endonuclease based assay to identify methylated sequences in the human genome [172]. In this study, SVA comprised >70 % of one of the libraries of methylated sequence. Along the epigenetic spectrum, alterations in local histone modifications following LINE-1 insertion have been described in teratocarcinoma cell lines [173]. Specifically, recruitment of a histone-deacetylase enzyme by some unknown mechanism or signal to LINE-1-target sites results in deacetylation of histone tails. Similarly, a new study reports that Sirtuin-6 (SIRT6) can repress LINE-1 [174] by binding the 5'-UTR and ribosylating KRAB-associated protein-1 (KAP1), a major corepressor. This posttranslational modification is important for KAP1 to interact with heterochromatin protein-1 $\alpha$  (HP1 $\alpha$ ). Interestingly, over time (e.g. ageing) SIRT6 is depleted at LINE-1 loci. Although no specific examples have been reported thus far for disease-causing insertions, experimental evidence indicates epigenetic silencing following LINE-1, Alu, or SVA insertion in a gene may result in reduced mRNA expression from a given gene. In contrast, loss of epigenetic mediated repression may lead not only to expression of retrotransposons but also neighboring genes. For example, one report demonstrated that loss of DNA methylation occurs at an intronic LINE-1 insertion near the hepatocyte growth factor receptor (*MET*) gene which leads to expression of a LINE-1-MET fusion transcript encoding a truncated form of this protein known to be oncogenic [175].

### New reports of LINE-1-mediated insertions causing Mendelian disorders

Since our last survey of disease-causing insertions [86], 28 more have been reported in the literature. Disease-causing insertions have been priceless in regards to our understanding of human retrotransposon biology. Genetic disease phenotypes serve as markers to identify *de novo* retrotransposition events. It has been almost 30 years since the first *de novo* retrotransposon insertion was identified in the factor VIII (*F8*) gene of a Hemophilia A patient by Kazazian and colleagues [22]. LINE-1-mediated insertions have been associated with autosomal dominant, autosomal recessive, and X-linked genetic disorders (Table 1). Disease-causing insertions have aided in the recovery of active retrotransposons used in cell-culture retrotransposition assays [41]. Furthermore, these insertions have confirmed and revealed phenomena, such as 3'-transductions [176], observed in cell culture and genomic studies.

Neurofibromatosis Type I is an autosomal dominant disorder caused by mutations in the *NF1* gene. Recently, while characterizing genetic deletions in the neurofibromin 1 (*NF1*) gene, an SVA insertion associated with a 867 kb deletion in one individual and an SVA in a different individual associated with a 1 MB deletion were found [153]. These two insertions represent the largest genomic deletions caused by a *de novo* insertion to date. Using sequence analysis, the authors were also able to identify the source elements for both insertions. One insertion was generated from a full-length SVA located on chromosome 6 belonging to the human-specific subfamily F. The other insertion was generated from an element on chromosome 10 belonging to the human specific SVA\_F1 (*MAST2*) subfamily. The element on chromosome 10 has been associated with other SVA disease-causing insertions and is thought to be the source element for at least 13 genomic SVAs [127, 128]. Interestingly, both insertions were somatic. One patient had the SVA-associated deletion in 93 % of her blood cells (absent in 7 % of her blood cells); while the grandmother of the other patient who passed on the insertion had the SVA in 75 % of her blood cells (absent in 25 % of the blood cells).

Somatic mosaicism has been described for disease-causing insertions as in a LINE-1 retrotransposition event into the choroideremia (rab escort protein 1) (*CHM*) gene [177]. More and more evidence is accumulating that somatic insertions may be more common than previously appreciated and perhaps the norm [178–180]. In addition to disease-causing insertions, insights into somatic mosaicism generated by LINE-1 activity were first gained from two mouse studies: one investigating retrotransposition of engineered LINE-1 in the brain [180] and the other studying LINE-1

inheritance [178]. Next-generation sequencing of cancer genomes and brain samples is starting to uncover a wealth of somatic insertions. The current thought in the field is that many, if not most, LINE-1-mediated insertions are not inherited despite what had been previously inferred based on the abundance of genomic insertions.

Most *de novo* retrotransposition events are likely benign, however coupled with a loss-of-function mutation on the other allele the insertion may result in recessive genetic disease (e.g. compound heterozygosity). A study analyzing the genetic basis for Rotor syndrome [157], an autosomal recessive disorder, uncovered patients homozygous for a near full-length LINE-1 insertion (lacking 24 nt from the 5'-end relative to LINE-1.3) in intron 5 of solute carrier organic anion transporter family member 1B3 (*SLCO1B3*). This insertion results in aberrant mRNA splicing and ultimately loss of *SLCO1B3* protein expression in liver tissues. As Rotor syndrome is a digenic disorder the homozygous LINE-1 insertion alone is not sufficient to cause Rotor syndrome; these patients are also homozygous for a nonsense mutation in the downstream solute carrier organic anion transporter family member 1B1 (*SLCO1B1*) gene. LINE-1<sub>*SLCO1B3*</sub> may represent a population-specific "hot LINE-1" with a gene frequency of ~6 % in Japanese individuals. LINE-1<sub>*SLCO1B3*</sub> contains intact reading frames with ORF1 being 100 % identical to the LINE-1 amino acid consensus and LINE-1<sub>*SLCO1B3*</sub> ORF2 containing three amino acid changes relative to the consensus LINE-1 sequence. A LINE-1 whose sequence is close to the amino acid consensus is typically very active in cell culture retrotransposition assays. For, example LINE-1<sub>*LRE3*</sub> [181], which is one of the most active LINE-1 s isolated to date shares 100 % amino acid identity with the consensus sequence. In addition, a survey of highly active ("hot") LINE-1 s, recently identified a very active element also lacking the first 21 nt (118 % of LINE-1.3) [40]. The first nucleotide of LINE-1<sub>*SLCO1B3*</sub> -a guanine- may actually represent reverse-transcription of the 7mG cap, a phenomenon often observed for very active elements, as most full-length LINE-1 s have a thymine at this same position in the 5'-UTR. Although LINE-1<sub>*SLCO1B3*</sub> may appear 5'-truncated, it is more probable that an alternative transcriptional start site may have been used at the source locus or that the source locus was lacking the first 24 nts. Other instances of LINE-1-mediated insertions associated with recessive diseases are typically due to a founder effect or consanguinity.

LINE-1-mediated retrotransposition events and inactive retrotransposons have long been considered to be agents of genome instability. A new study [155] analyzing a germline chromosome shattering event - a phenomenon referred to as chromothripsis [182, 183] - that can also occur in cancer, which consisted of 7 breakpoints and

rearrangements between two chromosomes, identified a 502 bp 5'-truncated SVA element insertion spanning a breakpoint associated with a 110 kb deletion. This SVA element belongs to a young active subfamily and may be derived from a full-length SVA on chromosome 7. Additional analysis of genomic sequence prior to DNA shattering identified two Alu elements on the same strand flanking the 110 kb sequence that was deleted. Furthermore, two antisense Alus were identified at breakpoints junctions involving an inversion in this chromothripsis event. Unexpectedly, sequence motifs resembling LINE-1 endonuclease cleavage sites were identified at exact breakpoints. A model was developed that integrated Alu-mediated chromosome looping and LINE-1-mediated SVA retrotransposition to account for the genome configuration following chromothripsis [155].

To date, the DNA damage agents causing chromothripsis and the mechanisms driving the rearrangement of chromosomal fragments are poorly understood. [184–186]. Several hypotheses have been generated to explain chromothripsis [184–186] including: 1) replication fork collapse coupled with template-switching and 2) ionizing radiation followed by DNA repair via the non-homologous end-joining pathway. More recently, experimental analysis has demonstrated that partitioning of chromosomes into micronuclei can result in chromothripsis and may explain why chromothriptic rearrangements are restricted to a limited number of chromosomes [187]. In addition, another study has provided evidence that three prime repair exonuclease 1 (TREX1) may cause rearrangements reminiscent of chromothripsis between dicentric chromosomes formed by telomere fusion [188].

Although the new study [155] represents only one instance of active and inactive retrotransposons associated with chromothripsis, it is tempting to speculate that an endonuclease, such as ORF2p expressed in germ cells, in early development, and in some cancer cells may play a role. Finally, although a bit tenuous, it is worthwhile to note that "kataegis," the localized hypermutation frequently observed near DNA breakpoint junctions in chromothripsis [188, 189] and cancers [190], is thought to be caused by the LINE-1 restriction factors [94, 191–194]- the APOBEC3 proteins [195]. APOBEC3A control of LINE-1 typically results in no observed remnants of LINE-1 sequence at the target-site due to uracil DNA glycosylase activity following APOBEC3A deamination of the TPRT intermediate [94]. Perhaps kataegis is a consequence of APOBEC3 defense of the genome against retrotransposons in cancers.

Contemporary LINE-1, Alu, and SVA activity has been documented, in part, due to new insertions causing genetic disease. A long-standing question in the field is

whether processed pseudogene formation (e.g. retrotransposition of cellular mRNAs, retrogenes) is ongoing in humans? Despite the name, retrogenes may serve as a crucible for new genes. Retrogenes have the potential for subfunctionalization or neofunctionalization. For example, 5'-truncation or point mutations could in principle generate a negative regulator of the parent gene. One report noted that processed pseudogenes coupled with 5'-inversion may be a means to generate new genes with novel N-termini [196]. Likewise, it is interesting that many large DNA viruses, such as poxviruses, contain many genes that share homology to host genes, lack introns, and are flanked by A-rich sequences; perhaps, implicating LINE-1 activity in the evolution of viral genomes.

Analysis of the 1000 genomes sequencing and Cancer Genome Atlas data has identified 48 polymorphic processed pseudogenes [197], thus indicating retrotransposition of cellular mRNAs in recent human history. A new study demonstrated ongoing processed pseudogene formation when the investigators identified an insertion of a partly processed TMF1 gene transcript into the chromosome b-245, beta polypeptide (*CYBB*) gene of a chronic granulomatous disease patient [152]. Uniquely, this insertion was very large (~5.8 kb) and represented an RNA that utilized a noncanonical polyA signal [152]. The insertion was flanked by a target-site duplication, inserted at a LINE-1 EN cleavage site, and ended in a 3'- polyA tail. The authors also demonstrated that the mother of the patient displayed somatic mosaicism for the insertion consistent with retrotransposition in early development [152]. Along these lines, retrotransposition of an almost full-length centromere protein W (*CENPW*) RNA, lacking 7 bp relative to the annotated TSS, into exon 8 of *Poc1* centriolar protein A (*Poc1a*) resulted in growth insufficiency and male infertility in mouse [198] (insertion size = 495 bp). This insertion displays all of the hallmarks of LINE-1-mediated retrotransposition (target-site duplication, insertion at EN cleavage site, and 3'-polyA tail).

#### Other diseases

It is well-established that retrotransposition can occasionally result in human genetic disease. Of late there has been a great effort to determine whether these selfish genetic elements may contribute to complex diseases such as cancer, autoimmunity, and neuropsychiatric disorders.

#### *LINE-1 s and cancer*

Genomic instability is a hallmark of cancer [199]. Notably, one of the first disease-causing insertions reported was an LINE-1 insertion into the adenomatous polyposis coli (*APC*) gene of a colon cancer patient described by Nakamura and colleagues [200]. That insertion was

somatic as it was absent in normal colon from the patient. Likewise, a very short somatic LINE-1 insertion (112 bp) was identified from exome data using Transpos-Seq analysis in exon 6 of phosphatase and tensin homolog (*PTEN*) of an endometrial carcinoma [201]. Two new reports further indicate that cancer can be initiated by retrotransposition-mediated gene inactivation. The first example is a full-length LINE-1 insertion located in intron 14 of the tumor-suppressor retinoblastoma 1 (*RB1*) which results in retinoblastoma in the proband and his father [202]. The authors' determined that this insertion was *de novo*, as it was absent from the father's parents and the proband's brother. The insertion causes aberrant RB1 splicing due to its precise integration into the splice acceptor site (target-site duplication (uppercase) tttt/AAATTATCTGTTTC/ag, splice acceptor trinucleotide motif in bold).

The second new report involves the use of population whole-genome sequencing to identify a full-length SVA<sub>E</sub> insertion (2792 bp in length) into intron 8 of the caspase 8 (*CASP8*) gene associated with increased susceptibility to cutaneous basal cell carcinoma (BCC) and breast cancer [203]. It is thought that this SVA insertion accounts for the previously reported germline SNP in *CASP8* linked to BCC. The mechanism by which the antisense SVA insertion results in decreased *CASP8* expression in breast cancer is unclear but it is not thought to be due to aberrant splicing. Extensive genotyping analysis indicated that the same SVA insertion into *CASP8* confers protection against prostate cancer in the same populations. This SVA locus has also been active in recent human history as evidenced by a full-length SVA insertion on chromosome 19 containing a 288 bp 3'-transduction derived from intron 8 *CASP8* sequence.

Although, LINE-1-mediated insertions have been identified in tumor suppressor genes, the overall absence of insertions in these genes has led researchers to focus on the contribution of LINE-1-mediated retrotransposition to cancer progression instead of cancer initiation. Numerous studies by independent labs over the past several years have reported extensive retrotransposition and/or LINE-1 protein expression in a variety of cancer types [201, 204–211]. Our recent studies [205, 207] demonstrate that LINE-1 insertions can occur in somatic gastrointestinal tissues, and that they can be carried forward essentially clonally in the cancers. We postulate that these somatic insertions contribute to the evolvability of cancer and its progression in the presence of limited resources and competition from not only the host but perhaps from other proximal competing cancer cells. Specifically, many somatic insertions may be benign, but following cancer initiation and the onslaught of other types of mutation including deletions, these insertions have the potential to optimize different cellular networks

or if full-length seed new agents of adaptability during cancer progression.

A largely unexplored question is whether LINE-1 proteins play yet undefined roles in cancer [212]. For example, ORF2p may be a source of endonuclease activity contributing to additional genomic rearrangements in these already unstable cells. Furthermore, LINE-1 RT activity may be a means to mend DNA breaks similar to what has been observed for LINE-1 endonuclease independent insertions (Fig. 2k) [213–215]. Although completely speculative, based on LINE-1 endonuclease independent insertions at telomeres in cell culture [216], perhaps on occasion, LINE-1-mediated insertions may aid in telomere elongation in some cancers. In addition, ORF2p expression may have important roles in cancer onset and progression through perturbation of regulatory networks [217–220].

#### **LINE-1 s and autoimmunity**

Autoimmunity is characterized by the immune system attacking “self.” Some autoimmune disorders such as Aicairdes-Goutieres syndrome are caused by mutations in genes, such as *TREX1* or SAM domain and HD domain 1 (*SAMHD1*), known to inhibit LINE-1 activity [221–223]. More recently, it has been demonstrated that a pattern-recognition receptor (PRR) named cyclic GMP-AMP synthase (cGAS) serves as a sensor for cytoplasmic DNA and activates the interferon response in the absence of the DNase *TREX1* [224]. Notably, cell culture and in vitro studies have shown that cGAS can activate the immune response not only by binding double-stranded DNA [225] but also by binding RNA: DNA hybrids [226]. Thus, cGAS or other PRRs may serve as critical cytoplasmic sentinels against retrotransposon replicative intermediates.

Autoantibodies are a hallmark of autoimmune disorders. Antibodies against the RNA binding protein Ro60 are detected in systemic lupus erythematosus and Sjogren’s syndrome. Interestingly, Ro60 RNPs reactive to autoantibodies contained Alu RNAs [227]. Similarly, Alu RNAs have also been implicated in age-related macular degeneration. During disease progression, expression of the microRNA processing enzyme DICER is reduced in retinal pigmented epithelium (RPE) [228]. Surprisingly, knockdown of DICER in human and mouse RPE results in an increase in Alu or B1 and B2 SINE RNA [228]. Knockdown of Alu RNAs using antisense oligonucleotides halts RPE degeneration driven by DICER knockdown in primary RPE culture. It is thought that the loss of DICER and an increase in Alu RNA leads to NLRP3 inflammasome activation resulting in cell death via Caspase-1 activation [229]. Interestingly, nucleoside reverse-transcriptase inhibitors known to inhibit LINE-1 activity [230] can block RPE

degeneration and inflammasome activation in mice injected sub-retinally with a plasmid expressing Alu [231].

The ability of retrotransposon replicative intermediates (e.g. RNA, cDNA) to trigger the innate immune response - activation of apoptotic pathways or interferon signaling- is consistent with a vital role for the immune system in protecting the cell and genome from TEs like LINE-1. It has been speculated that many key innate immunity factors such as APOBEC3 first evolved to control retrotransposition. Although several examples already exist, it is highly likely that additional immunity factors known to inhibit viral replication will be shown to also inhibit LINE-1 activity. In addition, while insertional mutagenesis is thought to be the primary means by which retrotransposons result in human disease, these highlights from the literature indicate that the RNAs themselves may be toxic to host fitness. It remains to be determined whether individuals harboring diseases associated with an increase in retrotransposon RNA have an increase in endogenous LINE-1 mediated retrotransposition.

#### **LINE-1 s and neuronal diversity**

A little over a decade ago, an interesting observation was made by Muotri, Gage, and colleagues regarding which cells are permissive for LINE-1 retrotransposition [180]. Prior to their work, most retrotransposition was thought to occur in the germline as evidenced by the ~500,000 LINE-1 copies in the human genome. Using engineered LINE-1 s, the authors detected LINE-1 retrotransposition in rat neuronal progenitor cells and in the brain of mice carrying an engineered LINE-1 marked with GFP [180]. These data demonstrated that engineered LINE-1 retrotransposition in the brain resulted in somatic mosaicism.

After a few years, the major question of whether endogenous LINE-1 was retrotransposing in the brain in vivo would be answered. Faulkner and colleagues developed a new technique termed Retrotransposon Capture-sequencing (RC-seq) [232]. This method coupled an array targeting the 5’- and 3’- ends of LINE-1, Alu, and SVA elements with high-throughput sequencing to enrich for potentially rare retrotransposition events. RC-seq revealed that LINE-1, Alu, and SVA retrotransposition had occurred somatically in the human hippocampus and caudate nucleus [232].

More recent work involving whole-genome amplification with RC-seq of single hippocampal neurons revealed almost 14 somatic insertions per cell [233]. Another study from the Walsh group on single cells outside the hippocampus has found a much lower incidence of somatic LINE-1 retrotransposition [234]. We in the field are convinced that LINE-1 retrotransposition is occurring in the brain; however its rate is presently a matter of some

controversy. With sensitive methods in place, research over the next 5 years will begin to determine some of the questions research on LINE-1 activity in the brain has generated: [180, 232, 234–238] 1) Are there functional implications for retrotransposition in the brain? 2) Does retrotransposition in the brain contribute to neurological diseases? 3) What is the true rate of retrotransposition in the brain and other somatic tissues?

## Conclusions

LINE-1 and other retrotransposons have moved from mysterious, repetitive sequences in our genome to making appearances in diverse research fields from cancer biology to neuroscience. Future research may reveal that TEs such as LINE-1 are the giant shoulders on which our genome and the cell stands.

## Abbreviations

ADAR: adenosine deaminase acting on RNA; APC: adenomatous polyposis coli; APOBEC3A: apolipoprotein B mRNA editing enzyme catalytic subunit 3A; ASP: antisense promoter; ATM: ataxia telangiectasia mutated; BCC: basal cell carcinoma; C: cysteine-rich; CASP8: caspase 8; CC: coiled coiled; CENPW: centromere protein W; cGAS: cyclic GMP-AMP synthase; CHM: choroideremia; CHO: Chinese Hamster Ovary; CTD: C-terminal domain; CYBB: cytochrome b-245, beta polypeptide; EN: endonuclease; ENi: Endonuclease-independent; ERV: endogenous retroviruses; EVE: endogenous viral elements; FKTN: fukutin; F8: factor VIII; HAL1: half-LINE-1; HP1 $\alpha$ : heterochromatin protein-1 $\alpha$ ; KAP1: KRAB-associated protein-1; LINE-1: L1: Long Interspersed Element-1; LTR: Long-Terminal Repeats; MAST2: microtubule associated serine/threonine kinase 2; NF1: neurofibromin 1; NAHR: non-allelic homologous recombination; PRR: pattern-recognition receptor; PGBD5: PiggyBac transposable element-derived protein 5; Poc1a: Poc1a centriolar protein A; PP: processed pseudogenes; PTEN: phosphatase and tensin homolog; RAG1: recombination activating gene 1; RB1: retinoblastoma 1; RC: retrotranspositionally-competent; RC-seq: retrotransposon capture-sequencing; RNaseH: ribonuclease H; RNP: ribonucleoprotein; RPE: retinal pigmented epithelium; RRM: RNA recognition motif; RT: reverse transcriptase; RUNX3: runt related transcription factor 3; SAMHD1: SAM domain and HD domain 1; SINE: Short Interspersed Element; SVA: SINE-VNTR-Alu; SIRT6: Sirtuin-6; SD: splice donor sites; SLC01B1: solute carrier organic anion transporter family member 1B1; SLC01B3: solute carrier organic anion transporter family member 1B3; TCF/LEF: T-cell factor/lymphoid enhancer factor; TE: transposable element; THAP9: THAP domain containing 9; TPRT: target-primed reverse-transcription; TREX1: three prime repair exonuclease 1; TSD: target-site duplication; VNTR: variable number tandem repeats; YY1: ying yang 1.

## Competing interests

HHK declares that he is a member of the Scientific Advisory Board of Transposagen Biopharmaceuticals, Inc.

## Authors' contributions

DCH and HHK drafted the manuscript. Both authors read and approved the final manuscript.

## Acknowledgments

We would like to thank Dr. Adam Ewing and three anonymous reviewers for helpful comments and suggestions on this manuscript. We apologize to authors whose work was not discussed or described only briefly. Also, we are extremely grateful to those investigators who shared information regarding their reports on disease-causing insertions.

## Funding

D.C.H. is funded by an American Cancer Society Postdoctoral Fellowship 125132-PF-13-371-01-MPC. Work in the Kazazian laboratory is funded by grants from the National Institutes of Health awarded to H.H.K.

Received: 19 January 2016 Accepted: 14 April 2016

Published online: 06 May 2016

## References

1. The Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002;420:520–62.
2. 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.
3. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 2005;437:69–87.
4. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev*. 1999;9:657–63.
5. Smit AF. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev*. 1996;6:743–8.
6. 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, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
7. McClintock B. Controlling Elements and the Gene. *Cold Spring Harb Symp Quant Biol*. 1956;21:197–216.
8. Britten RJ, Davidson EH. Repetitive and Non-Repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty. *Q Rev Biol*. 1971;46: 111–38.
9. Britten RJ, Kohne DE. Repeated Sequences in DNA. *Science*. 1968;161:529–40.
10. Ewing AD. Transposable element detection from whole genome sequence data. *Mobile DNA*. 2015;6:24.
11. Xing J, Witherspoon DJ, Jorde LB. Mobile element biology: new possibilities with high-throughput sequencing. *Trends Genet*. 2013;29:280–9.
12. Aldred SF, Doyle E, Epstein CB, Kaul R, Lajoie BR, Rosenbloom KR, Song L, Trinklein ND, Altschuler RC, Brown JB, Dong X, Hardison RC, Harris RS, Iyer S, Kheradpour P, Kundaje A, Lassmann T, Li Q, Merkel A, Thurman RE, Wu W, Yip KY, Bernstein BE, Birney E, Dunham I, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
13. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008;9:397–405.
14. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*. 2016;351:1083–7.
15. Garfinkel DJ, Boeke JD, Fink GR. Ty element transposition: reverse transcriptase and virus-like particles. *Cell*. 1985;42:507–17.
16. Boeke JD, Garfinkel DJ, Styles CA, Fink GR. Ty elements transpose through an RNA intermediate. *Cell*. 1985;40:491–500.
17. Mitra R, Li X, Kapusta A, Mayhew D, Mitra RD, Feschotte C, Craig NL. Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc Natl Acad Sci*. 2013;110: 234–9.
18. Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res*. 2008;18:717–28.
19. Agrawal A, Eastman QM, Schatz DG. Implications of transposition mediated by V(D) J-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity. *Nature*. 1998;394:744–51.
20. Henssen AG, Henaff E, Jiang E, Eisenberg AR, Carson JR, Villasante CM, Ray M, Still E, Burns M, Gandara J, Feschotte C, Mason CE, Kentsis A. Genomic DNA transposition induced by human PGBD5. *Elife*. 2015;4:488.
21. Majumdar S, Singh A, Rio DC. The Human THAP9 Gene Encodes an Active P-Element DNA Transposase. *Science*. 2013;339:446–8.
22. Kazazian HH, Wong C, Yousoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*. 1988;332:164–6.

23. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. *A de novo* Alu insertion results in neurofibromatosis type 1. *Nature*. 1991;353:864–6.
24. 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.
25. Ewing AD, Kazazian HH. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res*. 2010;20:1262–70.
26. Macfarlane CM, Badge RM. Genome-wide amplification of proviral sequences reveals new polymorphic HERV-K(HML-2) proviruses in humans and chimpanzees that are absent from genome assemblies. *Retrovirology*. 2015;12:161.
27. Marchi E, Kanapin A, Byott M, Magiorkinis G, Belshaw R. Neanderthal and Denisovan retroviruses in modern humans. *Curr Biol*. 2013;23:R994–5.
28. Belshaw R, Dawson ALA, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K(HML2): Implications for Present-Day Activity. *J Virol*. 2005;79:12507–14.
29. Wildschutte JH, Williams ZH, Montesin M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci*. 2016 [Epub ahead of print].
30. Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, Johnson ME, Eichler MY, McPherson JD, Zhao S, Pääbo S, Eichler EE. Lineage-Specific Expansions of Retroviral Insertions within the Genomes of African Great Apes but Not Humans and Orangutans. *PLOS Biol*. 2005;3:e110.
31. Sverdlov ED. Retroviruses and primate evolution. *BioEssays*. 2000;22:161–71.
32. Mayer J, Meese E. Human endogenous retroviruses in the primate lineage and their influence on host genomes. *Cytogenet Genome Res*. 2005;110:448–56.
33. Malik HS, Eickbush TH. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res*. 2001;11:1187–97.
34. Zhuo X, Feschotte C. Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammal Lineages. *PLoS Pathog*. 2015;11:e1005279.
35. Ostertag EM, Kazazian HH. Biology of mammalian L1 retrotransposons. *Annu Rev Genet*. 2001;35:501–38.
36. Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, Margolet L. Origin of the human L1 elements: Proposed progenitor genes deduced from a consensus DNA sequence. *Genomics*. 1987;1:113–25.
37. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. Molecular archeology of L1 insertions in the human genome. *Genome Biol*. 2002;3:research0052.
38. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci*. 2003;100:5280–5.
39. Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH. Many human L1 elements are capable of retrotransposition. *Nat Genet*. 1997;16:37–43.
40. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV. LINE-1 Retrotransposition Activity in Human Genomes. *Cell*. 2010;141:1159–70.
41. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. Isolation of an active human transposable element. *Science*. 1991;254:1805–8.
42. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian Jr HH. High Frequency Retrotransposition in Cultured Mammalian Cells. *Cell*. 1996;87:917–27.
43. 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.
44. Khazina E, Weichenrieder O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc Natl Acad Sci*. 2009;106:731–6.
45. 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.
46. Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol*. 2001;21:467–75.
47. Khazina E, Truffault V, Büttner R, Schmidt S, Coles M, Weichenrieder O. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol*. 2011;18:1006–14.
48. Naufer MN, Callahan KE, Cook PR, Pérez-González CE, Williams MC, Furano AV. L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. *Nucleic Acids Res*. 2016;44:281–93.
49. Callahan KE, Hickman AB, Jones CE, Ghirlando R, Furano AV. Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. *Nucleic Acids Res*. 2012;40:813–27.
50. Cook PR, Jones CE, Furano AV. Phosphorylation of ORF1p is required for L1 retrotransposition. *Proc Natl Acad Sci*. 2015;112:4298–303.
51. Feng Q, Moran JV, Kazazian Jr HH, Boeke JD. Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. *Cell*. 1996;87:905–16.
52. Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science*. 1991;254:1808–10.
53. Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol*. 1990;10:6718–29.
54. 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.
55. 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.
56. Kuwabara T, Hsieh J, Muotri A, Yeo G, Warashina M, Lie DC, Moore L, Nakashima K, Asashima M, Gage FH. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat Neurosci*. 2009;12:1097–105.
57. Harris CR, DeWan A, Zupnick A, Normart R, Gabriel A, Prives C, Levine AJ, Hoh J. p53 responsive elements in human retrotransposons. *Oncogene*. 2009;28:3857–65.
58. 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.
59. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol*. 2001;21:1973–85.
60. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, Saghatelyan A, Gage FH. Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell*. 2015;163:583–93.
61. Grimaldi G, Skowronski J, Singer MF. Defining the beginning and end of KpnI family segments. *EMBO J*. 1984;3:1753–9.
62. Moran JV, DeBerardinis RJ, Kazazian HH. Exon shuffling by L1 retrotransposition. *Science*. 1999;283:1530–4.
63. Goodier JL, Ostertag EM, Kazazian HH. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum Mol Genet*. 2000;9:653–7.
64. Pickeral OK, Makalowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res*. 2000;10:411–5.
65. Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci*. 2006;103:17608–13.
66. Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. Unconventional translocation of mammalian LINE-1 retrotransposons. *Genes Dev*. 2006;20:210–24.
67. Hohjoh H, Singer MF. Ribonuclease and high salt sensitivity of the ribonucleoprotein complex formed by the human LINE-1 retrotransposon. *J Mol Biol*. 1997;271:7–12.
68. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*. 2006;13:655–60.
69. Christensen SM, Eickbush TH. R2 Target-Primed Reverse Transcription: Ordered Cleavage and Polymerization Steps by Protein Subunits Asymmetrically Bound to the Target DNA. *Mol Cell Biol*. 2005;25:6617–28.
70. Doucet AJ, Wilusz JE, Miyoshi T, Liu Y, Moran JV. A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Mol Cell*. 2015;60:728–41.
71. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*. 2003;35:41–8.
72. Mandal PK, Ewing AD, Hancks DC, Kazazian HH. Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. *Hum Mol Genet*. 2013;22:3730–48.
73. 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.
74. Moldovan JB, Moran JV. The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. *PLOS Genet*. 2015;11:e1005121.

75. 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.
76. Dai L, Taylor MS, O'Donnell KA, Boeke JD. Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. *Mol Cell Biol.* 2012;32:4323–36.
77. 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.
78. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 2002;21:5899–910.
79. Xiong YE, Eickbush TH. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell.* 1988;55:235–46.
80. Cost GJ, Golding A, Schlissel MS, Boeke JD. Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* 2001;29:573–7.
81. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc Natl Acad Sci.* 1997;94:1872–7.
82. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. Human L1 Retrotransposition Is Associated with Genetic Instability In Vivo. *Cell.* 2002;110:327–38.
83. Gilbert N, Lutz-Prigge S, Moran JV. Genomic Deletions Created upon LINE-1 Retrotransposition. *Cell.* 2002;110:315–25.
84. Gilbert N, Lutz S, Morrish TA, Moran JV. Multiple Fates of L1 Retrotransposition Intermediates in Cultured Human Cells. *Mol Cell Biol.* 2005;25:7780–95.
85. Kopera HC, Moldovan JB, Morrish TA, Garcia-Perez JL, Moran JV. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. *Proc Natl Acad Sci.* 2011;108:20345–50.
86. Hancks DC, Kazazian Jr HH. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev.* 2012;22:191–203.
87. 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.
88. Ostertag EM, Kazazian HH. Twin Priming: A Proposed Mechanism for the Creation of Inversions in L1 Retrotransposition. *Genome Res.* 2001;11:2059–65.
89. Boissinot S, Entezam A, Furano AV. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol Biol Evol.* 2001;18:926–35.
90. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci.* 2006;103:9590–4.
91. Kuhn A, Ong YM, Cheng C-Y, Wong TY, Quake SR, Burkholder WF. Linkage disequilibrium and signatures of positive selection around LINE-1 retrotransposons in the human genome. *Proc Natl Acad Sci.* 2014;111:8131–6.
92. Bibillo A, Eickbush TH. High processivity of the reverse transcriptase from a non-long terminal repeat retrotransposon. *J Biol Chem.* 2002;277:34836–45.
93. Han JS, Shao S. Circular retrotransposition products generated by a LINE retrotransposon. *Nucleic Acids Res.* 2012;40:10866–77.
94. Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife.* 2014;3:e02008.
95. Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y, Carson CT, Macià A, Moran JV, Gage FH. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci.* 2011;108:20382–7.
96. Houck CM, Rinehart FP, Schmid CW. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol.* 1979;132:289–306.
97. Grimaldi G, Queen C, Singer MF. Interspersed repeated sequences in the African green monkey genome that are homologous to the human Alu family. *Nucleic Acids Res.* 1981;9:553–68.
98. Fuhrman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP. Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res.* 1981;9:6439–56.
99. Weiner AM. An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. *Cell.* 1980;22:209–18.
100. Ullu E, Tschudi C. Alu sequences are processed 7SL RNA genes. *Nature.* 1984;312:171–2.
101. Paoletta G, Lucero MA, Murphy MH, Baralle FE. The Alu family repeat promoter has a tRNA-like bipartite structure. *EMBO J.* 1983;2:691–6.
102. Chu WM, Liu WM, Schmid CW. RNA polymerase III promoter and terminator elements affect Alu RNA expression. *Nucleic Acids Res.* 1995;23:1750–7.
103. Roy AM, West NC, Rao A, Adhikari P, Alemán C, Barnes AP, Deininger PL. Upstream flanking sequences and transcription of SINEs. *J Mol Biol.* 2000;302:17–25.
104. Conti A, Carnevali D, Bollati V, Fustinoni S, Pellegrini M, Dieci G. Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data. *Nucleic Acids Res.* 2015;43:817–35.
105. Sorek R, Ast G, Graur DT. Alu-Containing Exons are Alternatively Spliced. *Genome Res.* 2002;12:1060–7.
106. Quentin Y. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.* 1992;20:487–93.
107. Quentin Y. Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res.* 1992;20:3397–401.
108. Jurka J, Zuckerkandl E. Free left arms as precursor molecules in the evolution of Alu sequences. *J Mol Evol.* 1991;33:49–56.
109. Chang DY, Sasaki-Tozawa N, Green LK, Marais RJ. A trinucleotide repeat-associated increase in the level of Alu RNA-binding protein occurred during the same period as the major Alu amplification that accompanied anthropoid evolution. *Mol Cell Biol.* 1995;15:2109–16.
110. Ahl V, Keller H, Schmidt S, Weichenrieder O. Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation. *Mol Cell.* 2015;60:715–27.
111. 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.
112. Weichenrieder O, wild K, Strub K, Cusack S. Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature.* 2000;408:167–73.
113. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM. LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene.* 2008;419:1–6.
114. Wildschutte JH, Baron A, Diroff NM, Kidd JM. Discovery and characterization of Alu repeat sequences via precise local read assembly. *Nucleic Acids Res.* 2015;43:10292–307.
115. Medstrand P, van de Lagemaat LN, Mager DL. Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. *Genome Res.* 2002;12:1483–95.
116. Smalheiser NR, Torvik VI. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.* 2006;22:532–6.
117. Spengler RM, Oakley CK, Davidson BL. Functional microRNAs and target sites are created by lineage-specific transposition. *Hum Mol Genet.* 2014;23:1783–93.
118. Chen L-L, DeCervo JN, Carmichael GG. Alu element-mediated gene silencing. *EMBO J.* 2008;27:1694–705.
119. Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matisse TC, Buyske S, Gabriel A. Widespread RNA Editing of Embedded Alu Elements in the Human Transcriptome. *Genome Res.* 2004;14:1719–25.
120. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. *PLOS Biol.* 2004;2:e391.
121. Gong C, Maquat LE. IncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3[prime] UTRs via Alu elements. *Nature.* 2011;470:284–8.
122. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, Huang CT, Sandifer E, Hebert K, Barnes EW, Hubley R, Miller W, Smit AFA, Ullmer B, Batzer MA. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. *Science.* 2007;316:238–40.
123. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA Elements: A Hominid-specific Retroposon Family. *J Mol Biol.* 2005;354:994–1007.
124. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. *J Biol Chem.* 1994;269:8466–76.
125. Damert A. Composite non-LTR retrotransposons in hominoid primates. *Mob Genet Elements.* 2015;5:67–71.
126. Hancks DC, Kazazian HH. SVA retrotransposons: Evolution and genetic instability. *Semin Cancer Biol.* 2010;20:234–45.
127. Damert A, Raiz J, Horn AV, Löwer J, Wang H, Xing J, Batzer MA, Löwer R, Schumann GG. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res.* 2009;19:1992–2008.

128. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res.* 2009;19:1983–91.
129. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet.* 2011;20:3386–400.
130. Hancks DC, Mandal PK, Cheung LE, Kazazian HH. The minimal active human SVA retrotransposon requires only the 5'-hexamer and Alu-like domains. *Mol Cell Biol.* 2012;32:4718–26.
131. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends Genet.* 2007;23:183–91.
132. Bantysh OB, Buzdin AA. Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. *Biochemistry Mosc.* 2009;74:1393–9.
133. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, Lower J, Stratling WH, Lower 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.
134. Ianc B, Ochis C, Persch R, Popescu O, Damert A. Hominoid composite non-LTR retrotransposons—variety, assembly, evolution, and structural determinants of mobilization. *Mol Biol Evol.* 2014;31:2847–64.
135. Lupan I, Bulzu P, Popescu O, Damert A. Lineage specific evolution of the VNTR composite retrotransposon central domain and its role in retrotransposition of gibbon LAVA elements. *BMC Genomics.* 2015;16:389.
136. Carbone L, Harris RA, Mootnick AR, Milosavljevic A, Martin DK, Rocchi M, Capozzi O, Archidiacono N, Konkel MK, Walker JA, Batzer MA, de Jong PJ. Centromere remodeling in Hoolock leucconedys (Hylobatidae) by a new transposable element unique to the gibbons. *Genome Biol Evol.* 2012;4: 648–58.
137. Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, Herrero J, Roos C, Aken B, Anaclerio F, Archidiacono N, Baker C, Barrell D, Batzer MA, Beal K, Blancher A, Bohrsen CL, Brameier M, Campbell MS, Capozzi O, Casola C, Chiatante G, Cree A, Damert A, de Jong PJ, Dumas L, Fernandez-Callejo M, Flicek P, Fuchs NV, et al. Gibbon genome and the fast karyotype evolution of small apes. *Nature.* 2014;513:195–201.
138. Hara T, Hirai Y, Baicharoen S, Hayakawa T. A novel composite retrotransposon derived from or generated independently of the SVA (SINE/VNTR/Alu) transposon has undergone proliferation in gibbon genomes. *Genes Genet Syst.* 2012;87:181–90.
139. Moran JV. Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay. *Genetica.* 1999;107:39–51.
140. Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E. A New Family of Chimeric Retrotranscripts Formed by a Full Copy of U6 Small Nuclear RNA Fused to the 3' Terminus of L1. *Genomics.* 2002;80:402–6.
141. Buzdin A, Gogvadze E, Kovalskaya E, Volchkov P, Ustyugova S, Illarionova A, Fushan A, Vinogradova T, Sverdlov E. The human genome contains many types of chimeric retrogenes generated through *in vivo* RNA recombination. *Nucleic Acids Res.* 2003;31:4385–90.
142. Doucet AJ, Droc G, Siol O, Audoux J, Gilbert N. U6 snRNA Pseudogenes: Markers of Retrotransposition Dynamics in Mammals. *Mol Biol Evol.* 2015;32: 1815–32.
143. Hasnaoui M, Doucet AJ, Meziane O, Gilbert N. Ancient repeat sequence derived from U6 snRNA in primate genomes. *Gene.* 2009;448:139–44.
144. 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.
145. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet.* 2000;24:363–7.
146. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol.* 2001;21:1429–39.
147. Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of Years of Evolution Preserved: A Comprehensive Catalog of the Processed Pseudogenes in the Human Genome. *Genome Res.* 2003;13:2541–58.
148. Zhang Z, Harrison P, Gerstein M. Identification and Analysis of Over 2000 Ribosomal Protein Pseudogenes in the Human Genome. *Genome Res.* 2002; 12:1466–82.
149. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 2009;10:691–703.
150. Chen J-M, Stenson PD, Cooper DN, Férec C. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum Genet.* 2005;117:411–27.
151. 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.
152. de Boer M, van Leeuwen K, Geissler J, Weemaes CM, van den Berg TK, Kuijpers TW, Warris A, Roos D. Primary Immunodeficiency Caused by an Exonized Retroposed Gene Copy Inserted in the CYBB Gene. *Hum Mutat.* 2014;35:486–96.
153. Vogt J, Bengesser K, Claes KB, Wimmer K, Mautner V-F, van Minkelen R, Legius E, Brems H, Upadhyaya M, Gel JH, Lazaro C, Rosenbaum T, Bammert S, Messiaen L, Cooper DN, Kehrer-Sawatzki H. SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome Biol.* 2014;15:1–17.
154. van der Klift HM, Tops CM, Hes FJ, Devilee P, Wijnen JT. Insertion of an SVA element, a nonautonomous retrotransposon, in PMS2 intron 7 as a novel cause of lynch syndrome. *Hum Mutat.* 2012;33:1051–5.
155. Nazaryan-Petersen L, Bertelsen B, Bak M, Jonson L, Tommerup N, Hancks DC, Tumer Z. Germline Chromothripsis Driven by L1-Mediated Retrotransposition and Alu/Alu Homologous Recombination. *Hum Mutat.* 2016;37:385–95.
156. Nakamura Y, Murata M, Takagi Y, Kozuka T, Nakata Y, Hasebe R, Takagi A, Kitazawa J-I, Shima M, Kojima T. SVA retrotransposition in exon 6 of the coagulation factor IX gene causing severe hemophilia B. *Int J Hematol.* 2015;102:134–9.
157. Kagawa T, Oka A, Kobayashi Y, Hiasa Y, Kitamura T, Sakugawa H, Adachi Y, Anzai K, Tsuruya K, Arase Y, Hirose S, Shiraiishi K, Shiina T, Sato T, Wang T, Tanaka M, Hayashi H, Kawabe N, Robinson PN, Zemojtel T, Mine T. Recessive inheritance of population-specific intronic LINE-1 insertion causes a rotor syndrome phenotype. *Hum Mutat.* 2015;36:327–32.
158. Qian Y, Mancini-DiNardo D, Judkins T, Cox HC, Daniels C, Holladay J, Ryder M, Coffee B, Bowles KR, Roa BB. Identification of retrotransposon insertion mutations in hereditary cancer. Presented at the 65th annual meeting of the American Society of Human Genetics 2015.
159. Lesmana H, Dyer L, Zhou P, Li X, Denton J, Chonat S, Zhang K, Hopkin RJ, Kalfa TA. Alu-Element Insertion in PKR Gene as a novel cause of severe hereditary nonspherocytic hemolytic anemia. Presented at the 57th annual meeting and exposition of the American Society of Hematology 2015.
160. Masson E, Hammel P, Garceau C, Bénech C, Quémener-Redon S, Chen J-M, Férec C. Characterization of two deletions of the CTCF locus. *Mol Genet Metab.* 2013;109:296–300.
161. Peixoto A, Pinheiro M, Massena L, Santos C, Pinto P, Rocha P, Pinto C, Teixeira MR. Genomic characterization of two large Alu-mediated rearrangements of the BRCA1 gene. *J Hum Genet.* 2013;58:78–83.
162. Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa M, Hamano K, Sakakihara Y, Nonaka I, Nakagome Y, Kanazawa I, Nakamura Y, Tokunaga K, Toda T. An ancient retrotransposon insertion causes Fukuyama-type congenital muscular dystrophy. *Nature.* 1998;394:388–92.
163. Taniguchi-Ikeda M, Kobayashi K, Kanagawa M, Yu C-C, Mori K, Oda T, Kuga A, Kurahashi H, Akman HO, DiMauro S, Kaji R, Yokota T, Takeda S, Toda T. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature.* 2011;478:127–31.
164. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Minami R, Nakamura H, Matsuo M. Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J Clin Invest.* 1993;91:1862–7.
165. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.* 2005;33:4040–52.
166. Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA. Alu Retrotransposition-mediated Deletion. *J Mol Biol.* 2005;348:791–800.
167. Lee J, Ha J, Son S-Y, Han K. Human Genomic Deletions Generated by SVA-Associated Events. *Comp Funct Genom.* 2012;2012:1–7.
168. Deininger PL, Batzer MA. Alu Repeats and Human Disease. *Mol Genet Metab.* 1999;67:183–93.
169. Lee J, Han K, Meyer TJ, Kim H-S, Batzer MA. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE.* 2008;3:e4047.
170. Bestor TH, Bourc'his D. Transposon Silencing and Imprint Establishment in Mammalian Germ Cells. *Cold Spring Harb Symp Quant Biol.* 2004;69:381–8.
171. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science.* 2007;318:761–4.



172. Strichman-Almashanu LZ, Lee RS, Onyango PO, Perlman E, Flam F, Frieman MB, Feinberg AP. A Genome-Wide Screen for Normally Methylated Human CpG Islands That Can Identify Novel Imprinted Genes. *Genome Res.* 2002;12:543–54.
173. Garcia-Perez JL, Morell M, Scheyls 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.
174. Van Meter M, Kashyap M, Rezaadeh S, Geneva AJ, Morello TD, Seluanov A, Gorbunova V. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun.* 2014;5:1–10.
175. Wolff EM, Byun H-M, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer. *PLoS Genet.* 2010;6:e1000917–13.
176. Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat Genet.* 1994;7:143–8.
177. van den Hurk JAUM, Meij IC, del Carmen SM, Kano H, Nikopoulos K, Hoefsloot LH, Sistermans EA, de Wijs IJ, Mukhopadhyay A, Plomp AS, de Jong PTVM, Kazazian HH, Cremers FPM. L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet.* 2007;16:1587–92.
178. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 2009;23:1303–12.
179. Kazazian HH. Mobile DNA transposition in somatic cells. *BMC Biol.* 2011;9:62.
180. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature.* 2005;435:903–10.
181. Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, Roos D, Kazazian HH. Evidence consistent with human L1 retrotransposition in maternal meiosis I. *Am J Hum Genet.* 2002;71:327–36.
182. Maher CA, Wilson RK. Chromothripsis and Human Disease: Piecing Together the Shattering Process. *Cell.* 2012;148:29–32.
183. Kloosterman WP, Cuppen E. Chromothripsis in congenital disorders and cancer: similarities and differences. *Curr Opin Cell Biol.* 2013;25:341–8.
184. Chen J-M, Férec C, Cooper DN. Transient hypermutability, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations. *Mutat Res.* 2011;750:52–9.
185. Forment JV, Kaidi A, Jackson SP. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat Rev Cancer.* 2012;12:663–70.
186. Leibowitz ML, Zhang C-Z, Pellman D. Chromothripsis: A New Mechanism for Rapid Karyotype Evolution. *Annu Rev Genet.* 2015;49:183–211.
187. Zhang CZ, Leibowitz ML, Pellman D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* 2013;27:2513–30.
188. Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell.* 2015;163:1641–54.
189. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell.* 2012;149:979–93.
190. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jager N, Jones DTW, Jones D, Knappskog S, Kool M, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500:415–21.
191. 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.* 2006;103:8780–5.
192. Horn AV, Klawitter S, Held U, Berger A, Vasudevan AAJ, Bock A, Hofmann H, Hanschmann K-MO, Trösemeyer J-H, Flory E, Jabulowsky RA, Han JS, Löwer J, Löwer R, Münk C, Schumann GG. Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. *Nucleic Acids Res.* 2014;42:396–416.
193. 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.
194. Schumann GG. APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem Soc Trans.* 2007;35:637–42.
195. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife.* 2013;2:204–14.
196. Kojima KK, Okada N. mRNA Retrotransposition Coupled with 5' Inversion as a Possible Source of New Genes. *Mol Biol Evol.* 2009;26:1405–20.
197. Ewing AD, Ballinger TJ, Earl D, Broad Institute Genome Sequencing and Analysis Program and Platform, Harris CC, Ding L, Wilson RK, Haussler D. Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. *Genome Biol.* 2013;14:R22.
198. Geister KA, Brinkmeier ML, Cheung LY, Wendt J, Oatley MJ, Burgess DL, Kozloff KM, Cavalcoli JD, Oatley JM, Camper SA. LINE-1 Mediated Insertion into Pocl1 (Protein of Centriole 1 A) Causes Growth Insufficiency and Male Infertility in Mice. *PLoS Genet.* 2015;11:e1005569.
199. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell.* 2011;144:646–74.
200. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. Disruption of the APC Gene by a Retrotranspositional Insertion of L1 Sequence in a Colon Cancer. *Cancer Res.* 1992;52:643–5.
201. Helman E, Lawrence ML, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* 2014;24:1053–63.
202. Rodriguez-Martin C, Cidre F, Fernandez-Teijeiro A, Gomez-Mariano G, de la Vega L, Ramos P, Zaballos A, Monzon S, Alonso J. Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. *J Hum Genet.* 2016. doi:10.1038/jhg2015.173.
203. Stacey SN, Kehr B, Gudmundsson J, Zink F, Jonasdottir A, Gudjonsson SA, Sigurdsson A, Halldorsson BV, Agnarsson BA, Benediktsson KR, Aben KKH, Vermeulen SH, Cremers RG, Panadero A, Helfand BT, Cooper PR, Donovan JL, Hamdy FC, Jinga V, Okamoto I, Jonasson JG, Tryggvadottir L, Johannsdottir H, Kristinsdottir AM, Masson G, Magnusson OT, Iordache PD, Helgason A, Helgason H, Sulem P, et al. Insertion of an SVA-E retrotransposon into the CASP8 gene is associated with protection against prostate cancer. *Hum Mol Genet.* 2016;25:1008–18.
204. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, Menzies A, Roman-Garcia P, Fullam A, Gerstung M, Shlien A, Tarpey PS, Papaemmanuil E, Knappskog S, Van Loo P, Ramakrishna M, Davies HR, Marshall J, Wedge DC, Teague JW, Butler AP, Nik-Zainal S, Alexandrov L, Behjati S, Yates LR, Bolli N, et al. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science.* 2014;345:1251343.
205. Ewing AD, Gacita A, Wood LD, Ma F, Xing D, Kim M-S, Manda SS, Abril G, Pereira G, Makohon-Moore A, Looijenga LHJ, Gillis AJM, Hruban RH, Anders RA, Romans KE, Pandey A, Iacobuzio-Donahue CA, Vogelstein B, Kinzler KW, Kazazian HH, Solyom S. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 2015;25:1536–45.
206. Rodic N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, Kohutek ZA, Huang CR, Ahn D, Mita P, Taylor MS, Barker NJ, Hruban RH, Iacobuzio-Donahue CA, Boeke JD, Burns KH. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med.* 2015; 21:1060–4.
207. Doucet-O'Hare TT, Rodic N, Sharma R, Darbari I, Abril G, Choi JA, Young Ahn J, Cheng Y, Anders RA, Burns KH, Meltzer SJ, Kazazian HH. LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc Natl Acad Sci.* 2015;112:E4894–900.
208. Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, Sinha S, Iannelli F, Radaelli E, Dos Santos A, Rapoud D, Guettier C, Samuel D, Natoli G, Carninci P, Ciccarelli FD, Garcia-Perez JL, Faivre J, Faulkner GJ. Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma. *Cell.* 2014;153:101–11.
209. Solyom S, Ewing AD, Rahmann EP, Doucet T, Nelson HH, Burns MB, Harris RS, Sigmon DF, Casella A, Erlanger B, Wheelan S, Upton KR, Shukla R, Faulkner GJ, Largaespada DA, Kazazian HH. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 2012;22:2328–38.
210. Luca CD, Guadagni F, Sinibaldi-Vallebona P, Sentinelli S, Gallucci M, Hoffmann A, Schumann GG, Spadafora C, Sciamanna I. Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation. *Oncotarget.* 2016;7:4048–61.

211. Gualtieri A, Andreola F, Sciamanna I, Sinibaldi-Vallebona P, Serafino A, Spadafora C. Increased expression and copy number amplification of LINE-1 and SINE B1 retrotransposable elements in murine mammary carcinoma progression. *Oncotarget*. 2013;4:1882–93.
212. Sciamanna I, De Luca C, Spadafora C. The Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer. *Front Chem*. 2016;4:333–10.
213. 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.
214. Srikanta D, Sen SK, Huang CT, Conlin EM, Rhodes RM, Batzer MA. An alternative pathway for Alu retrotransposition suggests a role in DNA double-strand break repair. *Genomics*. 2009;93:205–12.
215. 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.
216. Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature*. 2007;446:208–12.
217. Sciamanna I, Gualtieri A, Cossetti C, Osimo EF, Ferracin M, Macchia G, Arico E, Prosseda G, Vitullo P, Misteli T, Spadafora C. A tumor-promoting mechanism mediated by retrotransposon-encoded reverse transcriptase is active in human transformed cell lines. *Oncotarget*. 2013;4:2271–87.
218. Sciamanna I, Gualtieri A, Piazza PV, Spadafora C. Regulatory roles of LINE-1-encoded reverse transcriptase in cancer onset and progression. *Oncotarget*. 2014;5:8039–51.
219. Oricchio E, Sciamanna I, Beraldi R, Tolstonog GV, Schumann GG, Spadafora C. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene*. 2007;26:4226–33.
220. Sciamanna I, Landriscina M, Pittoggi C, Quirino M, Mearrelli C, Beraldi R, Mattei E, Serafino A, Cassano A, Sinibaldi-Vallebona P, Garaci E, Barone C, Spadafora C. Inhibition of endogenous reverse transcriptase antagonizes human tumor growth. *Oncogene*. 2005;24:3923–31.
221. Hu S, Li J, Xu F, Mei S, Le Duff Y, Yin L, Pang X, Cen S, Jin Q, Liang C, Guo F. SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. *PLoS Genet*. 2015;11:e1005367.
222. Zhao K, Du J, Han X, Goodier JL, Li P, Zhou X, Wei W, Evans SL, Li L, Zhang W, Cheung LE, Wang G, Kazazian Jr HH, Yu X-F. Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1. *Cell Rep*. 2013;4:1108–15.
223. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell*. 2008;134:587–98.
224. Gao D, Li T, Li X-D, Chen X, Li Q-Z, Wight-Carter M, Chen ZJ. Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proc Natl Acad Sci*. 2015;112:E5699–705.
225. Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. *Science*. 2013;339:786–91.
226. Mankan AK, Schmidt T, Chauhan D, Goldeck M, Honing K, Gaidt M, Kubarenko AV, Andreeva L, Hopfner KP, Hornung V. Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. *EMBO J*. 2014;33:2937–46.
227. Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhangale T, Graham RR, Ortmann W, Criswell LA, Yeo GW, Behrens TW. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science*. 2015;350:455–9.
228. Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, Karikó K, Yoo JW, Lee D-K, Hadziahmetovic M, Song Y, Misra S, Chaudhuri G, Buaas FW, Braun RE, Hinton DR, Zhang Q, Grossniklaus HE, Provis JM, Madigan MC, Milam AH, Justice NL, Albuquerque RJC, Blandford AD, Bogdanovich S, Hirano Y, et al. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature*. 2011;471:325–30.
229. Tarallo V, Hirano Y, Gelfand BD, Dridi S, Kerur N, Kim Y, Cho WG, Kaneko H, Fowler BJ, Bogdanovich S, Albuquerque RJC, Hauswirth WW, Chiodo VA, Kugel JF, Goodrich JA, Ponicsan SL, Chaudhuri G, Murphy MP, Dunaief JL, Ambati BK, Ogura Y, Yoo JW, Lee D-K, Provost P, Hinton DR, Núñez G, Baffi JJ, Kleinman ME, Ambati J. DICER1 Loss and Alu RNA Induce Age-Related Macular Degeneration via the NLRP3 Inflammasome and MyD88. *Cell*. 2012;149:847–59.
230. Dai L, Huang Q, Boeke JD. Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochem*. 2011;12:18.
231. Fowler BJ, Gelfand BD, Kim Y, Kerur N, Tarallo V, Hirano Y, Amarnath S, Fowler DH, Radwan M, Young MT, Pittman K, Kubes P, Agarwal HK, Parang K, Hinton DR, Bastos-Carvalho A, Li S, Yasuma T, Mizutani T, Yasuma R, Wright C, Ambati J. Nucleoside reverse transcriptase inhibitors possess intrinsic anti-inflammatory activity. *Science*. 2014;346:1000–3.
232. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carninci P, Jeddeloh JA, Faulkner GJ. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011;479:534–7.
233. Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sánchez-Luque FJ, Bodea GO, Ewing AD, Salvador-Palomeque C, van der Knaap MS, Brennan PM, Vanderver A, Faulkner GJ. Ubiquitous L1 Mosaicism in Hippocampal Neurons. *Cell*. 2015;161:228–39.
234. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, Park PJ, Walsh CA. Single-Neuron Sequencing Analysis of L1 Retrotransposition and Somatic Mutation in the Human Brain. *Cell*. 2012;151:483–96.
235. Richardson SR, Morell S, Faulkner GJ. L1 retrotransposons and somatic mosaicism in the brain. *Annu Rev Genet*. 2014;48:1–27.
236. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. *Nature*. 2009;460:1127–31.
237. 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.
238. Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann HS, Park PJ, Walsh CA. Cell lineage analysis in human brain using endogenous retroelements. *Neuron*. 2015;85:49–59.
239. Skowronski J, Fanning TG, Singer MF. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol*. 1988;8:1385–97.
240. 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.
241. Rangwala SH, Zhang L, Kazazian HH. Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol*. 2009;10:R100.
242. Clements AP, Singer MF. The human LINE-1 reverse transcriptase: effect of deletions outside the common reverse transcriptase domain. *Nucleic Acids Res*. 1998;26:3528–35.
243. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nature Rev Genet*. 2002;3:370–9.
244. Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? *Genome Res*. 2009;19:545–55.
245. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL, Moran JV. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol Spectr*. 2015;3:MDNA3-0061–2014.
246. Szak ST, Pickeral OK, Landsman D, Boeke JD. Identifying related L1 retrotransposons by analyzing 3' transduced sequences. *Genome Biol*. 2003;4:R30.
247. Morisada N, Rendtorff ND, Nozu K, Morishita T, Miyakawa T, Matsumoto T, Hisano S, Iijima K, Tranebjærg L, Shirahata A, Matsuo M, Kusuhara K. Branchio-oto-renal syndrome caused by partial EYA1 deletion due to LINE-1 insertion. *Pediatr Nephrol*. 2010;25:1343–8.
248. Sheen FM, Sherry ST, Risch GM, Robichaux M, Nasidze I, Stoneking M, Batzer MA, Swergold GD. Reading between the LINES: human genomic variation induced by LINE-1 retrotransposition. *Genome Res*. 2000;10:1496–508.
249. Srikanta D, Sen SK, Conlin EM, Batzer MA. Internal priming: an opportunistic pathway for L1 and Alu retrotransposition in hominins. *Gene*. 2009;448:233–41.
250. Belancio VP, Roy-Engel AM, Deininger P. The impact of multiple splice sites in human L1 elements. *Gene*. 2008;411:38–45.
251. Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res*. 2010;38:3909–22.
252. Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet*. 2003;35:363–6.
253. Bao W, Jurka J. Origin and evolution of LINE-1 derived "half-L1" retrotransposons (HAL1). *Gene*. 2010;465:9–16.
254. Voliva CF, Martin SL, Hutchison CA, Edgell MH. Dispersal process associated with the L1 family of interspersed repetitive DNA sequences. *J Mol Biol*. 1984;178:795–813.

255. Kutsche K, Ressler B, Katzera H-G, Orth U, Gillissen-Kaesbach G, Morlot S, Schwinger E, Gal A. Characterization of breakpoint sequences of five rearrangements in *L1CAM* and *ABCD1* (*ALD*) genes. *Hum Mutat.* 2002;19:526–35.
256. Gu Y, Kodama H, Watanabe S, Kikuchi N, Ishitsuka I, Ozawa H, Fujisawa C, Shiga K. The first reported case of Menkes disease caused by an Alu insertion mutation. *Brain Dev-Jpn.* 2007;29:105–8.
257. Lester T, McMahon C, VanRegemorter N, Jones A, Genet S. X-linked immunodeficiency caused by insertion of Alu repeat sequences. *J Med Gen Suppl.* 1997;34:S81.
258. Conley ME, Partain JD, Norland SM, Shurtleff SA, Kazazian HH. Two independent retrotransposon insertions at the same site within the coding region of *BTK*. *Hum Mutat.* 2005;25:324–5.
259. Apoil PA, Kuhlein E, Robert A, Rubie H, Blancher A. HIGM syndrome caused by insertion of an AluYb8 element in exon 1 of the *CD40LG* gene. *Immunogenetics.* 2006;59:17–23.
260. Claverie-Martin F, González-Acosta H, Flores C, Antón-Gamero M, García-Nieto V. *De novo* insertion of an Alu sequence in the coding region of the *CLCN5* gene results in Dent's disease. *Hum Genet.* 2003;113:480–5.
261. Sukarova E, Dimovski AJ, Tchacarova P, Petkov GH, Efremov GD. An Alu insert as the cause of a severe form of hemophilia A. *Acta Haematol.* 2001;106:126–9.
262. Ganguly A, Dunbar T, Chen P, Godmilow L, Ganguly T. Exon skipping caused by an intronic insertion of a young Alu Yb9 element leads to severe hemophilia A. *Hum Genet.* 2003;113:348–52.
263. Green PM, Bagnall RD, Waseem NH, Giannelli F. Haemophilia A mutations in the UK: results of screening one-third of the population. *Br J Haematol.* 2008;143:115–28.
264. Vidaud D, Vidaud M, Bahnak BR, Siguret V, Gispert Sanchez S, Laurian Y, Meyer D, Goossens M, Lavergne JM. Haemophilia B due to a *de novo* insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur J Hum Genet.* 1993;1:30–6.
265. Wulff K, Gazda H, Schröder W, Milewska RR, Herrmann FH. Identification of a novel large *F9* gene mutation—An insertion of an Alu repeated DNA element in exon e of the factor 9 gene. *Hum Mutat.* 2000;15:1–2.
266. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Careri D, Pinto MT, Kasper CK, Sommer SS. Frequency of recent retrotransposition events in the human factor IX gene. *Hum Mutat.* 2001;17:1–9.
267. Zhang Y, Dipple KM, Vilain E, Huang BL, Finlayson G, Therrell BL, Worley K, Deininger P, McCabe ER. AluY insertion (IVS4-52ins316alu) in the glycerol kinase gene from an individual with benign glycerol kinase deficiency. *Hum Mutat.* 2000;15:316–23.
268. den Hollander AI, ten Brink JB, de Kok YJ, van Soest S, van den Born LI, van Driel MA, van de Pol DJ, Payne AM, Bhattacharya SS, Kellner U, Hoyng CB, Westerveld A, Brunner HG, Bleeker-Wagemakers EM, Deutman AF, Hecklenlively JR, Cremers FP, Bergen AA. Mutations in a human homologue of *Drosophila* crumbs cause retinitis pigmentosa (*RP12*). *Nat Genet.* 1999;23:217–21.
269. Beauchamp NJ, Makris M, Preston FE, Peake IR, Daly ME. Major structural defects in the antithrombin gene in four families with type I antithrombin deficiency—partial/complete deletions and rearrangement of the antithrombin gene. *Thromb Haemost.* 2000;83:715–21.
270. Taşkesen M, Collin GB, Evsikov AV, Güzel A, Özgül RK, Marshall JD, Naggert JK. Novel Alu retrotransposon insertion leading to Alström syndrome. *Hum Genet.* 2011;131:407–13.
271. Kloor M, Sutter C, Wentzensen N, Cremer FW, Buckowitz A, Keller M, Knebel Doberitz von M, Gebert J. A large *MSH2* Alu insertion mutation causes HNPCC in a German kindred. *Hum Genet.* 2004;115:432–8.
272. Ishihara N, Yamada K, Yamada Y, Miura K, Kato J, Kuwabara N, Hara Y, Kobayashi Y, Hoshino K, Nomura Y, Mimaki M, Ohya K, Matsushima M, Nitta H, Tanaka K, Ohki T, Ezoe T, Kumagai T, Onuma A, Kuroda T, Yoneda M, Yamanaka T, Saeki M, Segawa M, Saji T, Nagaya M, Wakamatsu N. Clinical and molecular analysis of Mowat-Wilson syndrome associated with *ZFX1B* mutations and deletions at 2q22-q24.1. *J Med Genet.* 2004;41:387–93.
273. Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuayama J, Higashino K. Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. *Proc Natl Acad Sci.* 1991;88:11315–9.
274. Janicic N, Pausova Z, Cole DEC, Hendy GN. Insertion of an Alu sequence in the Ca<sup>2+</sup>-sensing receptor gene in familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Am J Hum Genet.* 1995;56:880–6.
275. Sobrier M-L, Netchine I, Heinrichs C, Thibaud N, Vié-Luton M-P, Van Vliet G, Amselem S. Alu-element insertion in the homeodomain of *HESX1* and aplasia of the anterior pituitary. *Hum Mutat.* 2005;25:503.
276. Gallus GN, Cardaioli E, Rufa A, da Pozzo P, Bianchi S, D'Eramo C, Collura M, Tumino M, Pavone L, Federico A. Alu-element insertion in an *OPA1* intron sequence associated with autosomal dominant optic atrophy. *Mol Vis.* 2010;16:178–83.
277. Economou-Pachnis A, Tschlis PN. Insertion of an Alu SINE in the human homologue of the *Mvi-2* locus. *Nucleic Acids Res.* 1985;13:8379–87.
278. Halling KC, Lazzaro CR, Honchel R, Bufile JA, Powell SM, Arndt CA, Lindor NM. Hereditary desmoid disease in a family with a germline Alu I repeat mutation of the *APC* gene. *Hum Hered.* 1999;49:97–102.
279. Su LK, Steinbach G, Sawyer JC, Hindi M, Ward PA, Lynch PM. Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis. *Hum Genet.* 2000;106:101–7.
280. Tucker BA, Scheetz TE, Mullins RF, DeLuca AP, Hoffmann JM, Johnston RM, Jacobson SG, Sheffield VC, Stone EM. Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene male germ cell-associated kinase (*MAK*) as a cause of retinitis pigmentosa. *Proc Natl Acad Sci.* 2011;108:E569–76.
281. Manco L, Relvas L, Silva-Pinto C, Pereira J, Almeida AB, Ribeiro ML. Molecular characterization of five Portuguese patients with pyrimidine 5'-nucleotidase deficient hemolytic anemia showing three new P5N-I mutations. *Blood Cell Mol Dis.* 2006;91:266–7.
282. Chen J-M, Masson E, Macek Jr M, Raguénès O, Piskackova T, Fercot B, Fila L, Cooper DN, Audrézet M-P, Férec C. Detection of two Alu insertions in the *CFTR* gene. *J Cyst Fibros.* 2008;7:37–43.
283. Abdelhak S, Kalatzis V, Heilig R, Compain S, Samson D, Vincent C, Levi-Acobas F, Cruaud C, Le Merrer M, Mathieu M, König R, Vigneron J, Weissenbach J, Petit C, Weil D. Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of *EYA1*. *Hum Mol Genet.* 1997;6:2247–55.
284. Okubo M, Horinishi A, Saito M, Ebara T, Endo Y, Kaku K, Murase T, Eto M. A novel complex deletion-insertion mutation mediated by Alu repetitive elements leads to lipoprotein lipase deficiency. *Mol Genet Metab.* 2007;92:229–33.
285. Uda T, Okamoto N, Aramaki M, Torii C, Kosaki R, Hosokai N, Hayakawa T, Takahata N, Takahashi T, Kosaki K. An Alu retrotransposition-mediated deletion of *CHD7* in a patient with CHARGE syndrome. *Am J Med Genet.* 2007;143A:721–6.
286. Bouchet C, Vuillaumier-Barrot S, Gonzales M, Boukari S, Bizet CL, Fallet C, Delezoide AL, Moirrot H, Laquerriere A, Encha-Razavi F, Durand G, Seta N. Detection of an Alu insertion in the *POMT1* gene from three French Walker Warburg syndrome families. *Mol Genet Metab.* 2007;90:93–6.
287. Oldridge M, Zackai EH, McDonald-McGinn DM, Iseki S, Morris-Kay GM, Twigg SRF, Johnson D, Wall SA, Jiang W, Theda C, Jabs EW, Wilkie AOM. *De novo* Alu-element insertions in *FGFR2* identify a distinct pathological basis for Apert syndrome. *Am J Hum Genet.* 1999;64:446–61.
288. Bochkukova EG, Roscioli T, Hedges DJ, Taylor IB, Johnson D, David DJ, Deininger PL, Wilkie AOM. Rare mutations of *FGFR2* causing apert syndrome: identification of the first partial gene deletion, and an Alu element insertion from a new subfamily. *Hum Mutat.* 2009;30:204–11.
289. Tighe PJ, Stevens SE, Dempsey S, Le Deist F, Rieux-Laucat F, Edgar JDM. Inactivation of the *FAS* gene by Alu insertion: retrotransposition in an intron causing splicing variation and autoimmune lymphoproliferative syndrome. *Genes Immun.* 2002;3:566–70.
290. Stoppa-Lyonnet D, Carter PE, Meo T, Tosi M. Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. *Proc Natl Acad Sci.* 1990;87:1551–5.
291. Mustajoki S, Ahola H, Mustajoki P, Kauppinen R. Insertion of Alu element responsible for acute intermittent porphyria. *Hum Mutat.* 1999;13:1–8.
292. Tappino B, Regis S, Corsolini F, Filocamo M. An Alu insertion in compound heterozygosity with a microduplication in *GNPTAB* gene underlies Mucopolysaccharidosis II. *Mol Genet Metab.* 2008;93:129–33.
293. Miki Y, Katagiri T, Kasumi F, Yoshimoto T, Nakamura Y. Mutation analysis in the *BRCA2* gene in primary breast cancers. *Nat Genet.* 1996;13:245–7.
294. Teugels E, De Brakeleer S, Goelen L, Lissens W, Sermijn E, De Grève J. *De novo* Alu element insertions targeted to a sequence common to the *BRCA1* and *BRCA2* genes. *Hum Mutation.* 2005;26:284.
295. Schollen E, Keldermans L, Foulquier F, Briones P, Chabas A, Sánchez-Valverde F, Adamowicz M, Pronicka E, Wevers R, Matthijs G. Characterization

- of two unusual truncating *PMM2* mutations in two CDG-Ia patients. *Mol Genet Metab.* 2007;90:408–13.
296. Wimmer K, Callens T, Wernstedt A, Messiaen L. The *NF1* gene contains hotspots for L1 endonuclease-dependent *de novo* Insertion. *PLoS Genet.* 2011;7:e1002371–11.
  297. Meischl C, Roos D. The molecular basis of chronic granulomatous disease. *Seminars Immunopathol.* 1998;19:417–34.
  298. Meischl C, Boer M, Ahlin A, Roos D. A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Eur J Hum Genet.* 2000;8:697–703.
  299. Musova Z, Hedvicakova P, Mohrmann M, Tesarova M, Krepelova A, Zeman J, Sedlacek Z. A novel insertion of a rearranged L1 element in exon 44 of the dystrophin gene: Further evidence for possible bias in retroposon integration. *Biochem Bioph Res Co.* 2006;347:145–9.
  300. Yoshida K, Nakamura A, Yazaki M, Ikeda S-I, Takeda S. Insertional mutation by transposable element, L1, in the *DMD* gene results in X-linked dilated cardiomyopathy. *Hum Mol Genet.* 1998;7:1129–32.
  301. Awano H, Malueka RG, Yagi M, Okizuka Y, Takeshima Y, Matsuo M. Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA. *J Hum Genet.* 2010;55:785–90.
  302. Solyom S, Ewing AD, Hancks DC, Takeshima Y, Awano H, Matsuo M, Kazazian HH. Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. *Hum Mutat.* 2012;33:369–71.
  303. Mukherjee S, Mukhopadhyay A, Banerjee D, Chandak GR, Ray K. Molecular pathology of haemophilia B: identification of five novel mutations including a LINE 1 insertion in Indian patients. *Haemophilia.* 2004;10:259–63.
  304. Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, van Duijnhoven G, Kirschner R, Hemberger M, Bergen AA, Rosenberg T, Pinckers AJ, Fundele R, Rosenthal A, Cremers FP, Ropers HH, Berger W. Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet.* 1998;19:327–32.
  305. Martínez-Garay I, Ballesta MJ, Oltra S, Orellana C, Palomeque A, Moltó MD, Prieto F, Martínez F. Intronic L1 insertion and F268S, novel mutations in *RPS6KA3* (*RSK2*) causing Coffin–Lowry syndrome. *Clin Genet.* 2003;64:491–6.
  306. Samuelov L, Fuchs-Telem D, Sarig O, Sprecher E. An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. *Brit J Dermatol.* 2011;164:1390–2.
  307. Kondo-lida E, Kobayashi K, Watanabe M, Sasaki J, Kumagai T, Koide H, Saito K, Osawa M, Nakamura Y, Toda T. Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). *Hum Mol Genet.* 1999;8:2303–9.
  308. Bernard V, Minnerop M, Bürk K, Kreuz F, Gillessen-Kaesbach G, Zühlke C. Exon deletions and intragenic insertions are not rare in ataxia with oculomotor apraxia 2. *BMC Med Genet.* 2009;10:1–9.
  309. Divoky V, Indrak K, Mrug M, Brabec V, Huisman T, Prchal JT. A novel mechanism of  $\beta$  thalassemia: the insertion of L1 retrotransposable element into  $\beta$  globin IVS II. *Blood.* 1996;88 Suppl 1:148a. abstract.
  310. Miné M, Chen J-M, Brivet M, Desguerre I, Marchant D, de Lonlay P, Bernard A, Férec C, Abitbol M, Ricquier D, Marsac C. A large genomic deletion in the *PDHX* gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Hum Mutat.* 2007;28:137–42.
  311. Rohrer J, Minegishi Y, Richter D, Eguiguren J, Conley ME. Unusual mutations in *BTK*: An insertion, a duplication, an inversion, and four large deletions. *Cl Immunol.* 1999;90:28–37.
  312. Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, Goto S, Matsumoto S, Tabuena MD, Maranon E, Dantes M, Lee LV, Ogasawara K, Tooyama I, Akatsu H, Nishimura M, Tamiya G. Reduced neuron-specific expression of the *TAF1* gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet.* 2007;80:393–406.
  313. Wilund KR, Yi M, Campagna F, Arca M, Zuliani G, Fellin R, Ho Y-K, Garcia JV, Hobbs HH, Cohen JC. Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet.* 2002;11:3019–30.
  314. Hassoun H, Coetzer TL, Vassiliadis JN, Sahr KE, Maalouf GJ, Saad ST, Catanzariti L, Palek J. A novel mobile element inserted in the alpha spectrin gene: spectrin dayton. A truncated alpha spectrin associated with hereditary elliptocytosis. *J Clin Invest.* 1994;94:643–8.
  315. Takasu M, Hayashi R, Maruya E, Ota M, Imura K, Kougo K, Kobayashi C, Saji H, Ishikawa Y, Asai T, Tokunaga K. Deletion of entire *HLA-A* gene accompanied by an insertion of a retrotransposon. *Tissue Antigens.* 2007;70:144–50.
  316. Akman HO, Davidzon G, Tanji K, MacDermott EJ, Larsen L, Davidson MM, Haller RG, Szczepaniak LS, Lehman TJA, Hirano M, DiMauro S. Neutral lipid storage disease with subclinical myopathy due to a retrotransposal insertion in the *PNPLA2* gene. *Neuromuscular Disord.* 2010;20:397–402.
  317. Segal Y, Peissel B, Renieri A, de Marchi M, Ballabio A, Pei Y, Zhou J. LINE-1 elements at the sites of molecular rearrangements in Alport syndrome–diffuse leiomyomatosis. *Am J Hum Genet.* 1999;64:62–9.
  318. Jalanko A, Manninen T, Paltonen L. Deletion of the C-terminal end of aspartylglucosaminidase resulting in a lysosomal accumulation disease: evidence for a unique genomic rearrangement. *Hum Mol Genet.* 1995;4:435–41.
  319. Wang T, Lerer I, Gueta Z, Sagi M, Kadouri L, Peretz T, Abeliovich D. A deletion/insertion mutation in the *BRCA2* gene in a breast cancer family: A possible role of the Alu-polyA tail in the evolution of the deletion. *Gene Chromosome Canc.* 2001;31:91–5.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

