Roles for retrotransposon insertions in human disease

Dustin C. Hancks1* and Haig H. Kazazian Jr.2*

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
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 (PGBDS) [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 unixed 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 for 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 (AAAₙ) 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 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.)
Anatomy of retrotransposon insertions. A variety of structures for retrotransposon insertions [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 Alus [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 (AAATTTA) [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, or 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 insertions. 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′-transcription. H) Exon-trapping. Retrotransposons are dispersed throughout the genome including intronic sequences. 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 [126, 135, 136, 138]. I) 3′-end 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 actually retrotransposed as in the case of ORF1 mVEI 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-targe-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) insertion. En, 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 appears to be absent by annotation (e.g. YYYY/YR), the actual site is merely obscured by this activity [85, 86].

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
| Insertion | Gene | CHR | Reference | Disease | Subfamily | Size (bp) | polyA tail length | Truncation Site | Transduction Mechanism | Strand | Exon/Intron | Mechanism | Target-site duplication (TSD) | L1 EN site | Note |
|-----------|------|-----|-----------|---------|----------|----------|------------------|-----------------|----------------------|--------|-------------|-----------|-----------------------------|-----------|------|
| Alu       | ABCD1| X   | Kutsche et al. 2002 [255] | ALD     | AluYb9   | 98       | 20              | Y5'TR           | N                    | S      |             | 4.7 kb Deletion | No TSD | ATTT/AT |
| Alu       | ATP7A| X   | Gu et al. 2007 [256] | Menkes Disease | AluYaS2 | 282      | 89              | N               | N                    | AS     | E           | AAAGGACAGC | TTTT/AT |           |
| Alu       | BTK  | X   | Lester et al. 1997 [257] | XLA     | AluY     | N/A      | N/A             | N/A             | AS                   | E      | N/A         | N/A       |                |             |
| Alu       | BTK  | X   | Conley et al. 2005 [258] | XLA     | AluY     | 281      | 74              | N               | S                    | E      | AGAAATGTAGTAGTAGT | TTCT/AT    | Same insertion site Conley et al. SVA |
| Alu       | CD40LG| X | Apoil et al. 2007 [259] | HIGM    | AluYb8   | 292      | 8               | N               | N                    | AS     | E           | AAAATTTTC | TTTT/AT |
| Alu       | CLCN5| X   | Claverie-Martin et al. 2003 [260] | Dent's Disease | AluYa5 | 281      | 50              | N               | N                    | S      | E           | AGAAATGCTCGAAGA | TTCT/AT |
| Alu       | CTRC | 1   | Masson et al. 2013 [160] | Chronic pancreatitis | AluYb8 | 31       | 11              | Y5'TR           | N                    | AS     | 53.9 kb Deletion | N/A       | TTCT/AT Deletes entire CTRC and ELA2A genes |
| Alu       | PKLR | 1   | Lesmana et al. 2015 [159] | Severe Hereditary Nonspherocytic Hemolytic Anemia | Yb8     | 288      | 70              | N               | N                    | S      | E           | AAGATCATCAGCGAA | TTCT/KA consanguinity, consensus Yb8 |
| Alu       | FVIII| X   | Sukarova et al. 2001 [261] | Hemophilia A | AluYb8 | 290      | 47              | N               | N                    | AS     | 3 nt Deletion | No TSD | TTTT/AT Consensus Yb9 |
| Alu       | FVIII| X   | Ganguly et al. 2003 [262] | Hemophilia A | AluYb9 | 288      | 37              | N               | N                    | AS     | 0'Splicing | AAAACCAACAGG | TTTT/AT Consensus Yb9 |
| Alu       | FVIII| X   | Green et al. 2008 [263] | Hemophilia A | AluYb8 | FL       | N/A             | N               | N                    | AS     | E           | N/A       |                |             |
| Alu       | FIX  | X   | Vidaud et al. 1993 [264] | Hemophilia B | AluYaS2 | 244      | 78              | Y5'TR           | N                    | S      | E           | AAGAAATGGCAGATGCCA | TTCT/AA Same insertion site as Wulff et al. Alu |
| Alu       | FIX  | X   | Wulff et al. 2000 [265] | Hemophilia B | AluYaS2 | 237      | 39              | Y5'TR           | N                    | S      | E           | AAGAAATGGCAGATGC | TTCT/AA Same insertion site as Vidaud et al. Alu |
| Alu       | FIX  | X   | Li et al. 2001 [266] | Hemophilia B | AluY     | 279      | 40              | Y5'TR           | N                    | AS     | E           | AAGAAACTGTCCTCC | TTCT/AA |
| Alu       | GK   | X   | Zhang et al. 2000 [267] | GKD     | AluYc1   | 241      | 74              | Y5'TR           | N                    | AS     | I           | AAAAAAAAGAGA | TTCT/AA |
| Alu       | IL2RG| X   | Lester et al. 1997 [257] | XSCID   | AluYa5   | N/A      | N/A             | N/A             | N                    | AS     | I           | N/A       | N/A             |             |
| Alu | Genes | References | Type | Position | TSD | AS | E | Notes |
|-----|-------|------------|------|----------|-----|----|---|-------|
| 17  | Alu   | CRB1       | den Hollander et al. 1999 [268] | RP | AluY | 244 | 70 | Y/5TR | N | AS | E | AAGAGTAAAGATGA | TCTT/CA |
| 18  | Alu   | SERPNC1    | Beauchamp et al. 2000 [269] | Type 1 ATP | Alu | 6  | 40 | Y/5TR | N | AS | E | 1.4 kb Deletion | N/A |
| 19  | Alu   | ALMS1      | Thijssen et al. 2012 [270] | Alström syndrome | AluYa5 | 257 | 76 | Y/5TR | N | S | E | AAAACCTAGAGAA | TTTT/AA |
| 20  | Alu   | MSH2       | Koo et al. 2004 [271] | HNPCC | AluJ | 85  | 40 | Y/5TR | N | S | E | N/A | N/A |
| 21  | Alu   | MSH2       | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | N/A | E | N/A | N/A |
| 22  | Alu   | ZFHX1B     | Ishihara et al. 2004 [272] | MWS | AluYa5 | 281 | 93 | N | N | S | E | AAAATAAAAA | TTTT/AA |
| 23  | Alu   | BChE       | Muratani et al. 1991 [273] | Cholinesterase deficiency | AluYb9 | 289 | 38 | N | N | S | E | AAAAATTTTTC | TTTT/AA |
| 24  | Alu   | CASR       | Janicic et al. 1995 [274] | FHH and NSHPT | AluYa5 | 280 | 93 | N | N | S | E | GAAAGCGTGAGCTGC | TTCT/AA |
| 25  | Alu   | HESX1      | Sobrier et al. 2005 [275] | Anterior Pituitary Aplasia | AluYb8 | 288 | 30 | N | N | S | E | AGAAAAATGCTTAAGA | TTCT/AA |
| 26  | Alu   | OPA1       | Gallus et al. 2010 [276] | ADOA | AluYb8 | 289 | 25 | N | N | S | E | AAAATTAAAAAGTT | TTTT/AC |
| 27  | Alu   | MLV2       | Economou-Pachnis and Tischler 1985 [277] | Associated with leukemia | AluYa5 | 280 | 26 | N | N | S | E | GAAAATGT | TTCT/AT |
| 28  | Alu   | APC        | Halling et al. 1999 [278] | Hereditary desmoids disease | AluYb8 | 278 | 40 | Y/5TR | N | S | E | AAGATAATG | TTCT/AA |
| 29  | Alu   | APC        | Su et al. 2000 [279] | FAP | AluYb9 | 93  | 60 | Y/5TR | N | AS | E | AAAATTTTACGAGACAA | TTCT/AA |
| 30  | Alu   | APC        | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | N/A | E | N/A | N/A |
| 31  | Alu   | MAK        | Tucker et al. 2011 [280], Edwin Stone, personal communication | RP | AluYb8 | 281 | 57 | N | N | S | E | AAAAGAAAAA | TTCT/AA |
| 32  | Alu   | NT5C3      | Manco et al. 2006 [281], Leticia Ribeiro, personal communication | Chronic hemolytic anemia | Alu Ya5 | 281 | 36 | N | N | S | E | AAGAATGCGGATATGC | TTCT/AA |
Table 1 Retrotransposition events associated with human disease (Continued)

| #  | Alu | Gene | Year | Ref. | Disorder | Retrotransposon | Integration Site | Direction | Restoring Element | TSD | PolyA tail shortening between generations |
|----|-----|------|------|------|----------|-----------------|-----------------|-----------|------------------|-----|------------------------------------------|
| 33 | Alu | CFTR | 2008 | Chen et al. | Cystic Fibrosis | AluY | 46 | 57 | Y/S' TR | N | AS | E | AAGAATCCACCATATAATT | TCTT/AA |
| 34 | Alu | CFTR | 2008 | Chen et al. | Cystic Fibrosis | AluYa5 | 281 | 56 | N | N | S | E | AATAGAAAATGATTGGTC | TCTC/AT |
| 35 | Alu | EYA1 | 1997 | Abdelhak et al. | BOR syndrome | AluYa5 | N/A | 97,31 | N/A | N | AS | E | AAAAAATAAAATGTTGT | TTTT/AA |
| 36 | Alu | LPL | 2007 | Okubo et al. | LPL deficiency | AluYb9 | 150 | 60 | Y/S' TR | N | AS | 2.2 kb Deletion | No TSD | TTTT/AA |
| 37 | Alu | CHD7 | 2007 | Udaaka et al. | CHARGE syndrome | AluYa5/8 | 75 | 100 | Y/S' TR | N | S | 10 kb Deletion | No TSD | ATT/TT/AA |
| 38 | Alu | POMT1 | 1997 | Bouchet et al. | Walker Warburg syndrome | AluYa5 | 290 | 53 | N | N | AS | E | AAAAAAGATGTACTG | TTTT/AC |
| 39 | Alu | FGFR2 | 2009 | Oldridge et al. | Apert syndrome | AluYa5 | 283 | 69 | N | N | AS | E | GAATACCGCCAAG | TCTT/AT |
| 40 | Alu | FGFR2 | 2007 | Oldridge et al. | Apert syndrome | AluYb8 | 288 | 47 | N | N | AS | E | AAAAGTTACATTCCG | TTTT/AA |
| 41 | Alu | FGFR2 | 2007 | Bochukova et al. | Apert syndrome | AluYk13 | 214 | 12 | Y/S' TR | N | AS | E | AAAAGTTACATTCCG | TTTT/AA |
| 42 | Alu | FAS | 2002 | Tighe et al. | ALPS | AluYa5 | 281 | 33 | N | N | AS | I | AGAATATCTAAATGTG | TCTT/AA |
| 43 | Alu | SERPING1 | 1990 | Stoppa-Lyonnet et al. | HAE | AluYc1 | 285 | 42 | N | N | S | I | AAAAAATACAAAATTAG | TTTT/AG |
| 44 | Alu | HMBS | 1999 | Mustajoki et al. | AlP | AluYa5 | 279 | 39 | N | N | AS | E | AAAATCTGTCCC | TCTT/AA |
| 45 | Alu | ATM | 2015 | Qian et al. | Hereditary Cancer | N/A | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 46 | Alu | GNAI2 | 2008 | Tappino et al. | ML II | AluYa5 | 279 | 17 | N | N | AS | E | AAAAACACAACTGAG | TTTT/AA |
| 47 | Alu | BRCA2 | 1996 | Miki et al. | Breast Cancer | AluYc1 | 281 | 62 | N | N | S | E | AATCACAGGC | GATT/AT |
| 48 | Alu | BRCA2 | 2005 | Teugels et al. | Breast Cancer | AluYa5 | 285 | N/A | N | N | S | E | AAGAAATCTGACAT | TCTT/GC |
| 49 | Alu | BRCA2 | 2015 | Qian et al. | Hereditary Cancer | N/A | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 50 | Alu | BRCA2 | 2015 | Qian et al. | Hereditary Cancer | N/A | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 51 | Alu | BRCA2 | 2015 | Qian et al. | Hereditary Cancer | N/A | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| #  | Insertion Element | Associated Gene | Study Reference | Disease | Insertion Size | TSD | Description | Additional Information |
|----|------------------|-----------------|-----------------|---------|---------------|-----|-------------|-----------------------|
| 52 | Alu | BRCA2 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 53 | Alu | BRCA2 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 54 | Alu | BRCA2 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 55 | Alu | BRCA2 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 56 | Alu | BRCA2 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 57 | Alu | PMM2 | Schollen et al. 2007 [295] | CDG-Ia | AluYb8 | 263 | 10 | Y/5'TR | N | AS | 28 kb Deletion | No TSD | TTTT/AA |
| 58 | Alu | PALB2 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 59 | Alu | BRCA1 | 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 | BRCA1 | Teugels et al. 2005 [294] | Breast Cancer | AluS | 286 | N/A | N | N | S | E | GAAAAAGAATCTGCTTT | TTTC/GA |
| 61 | Alu | BRCA1 | Qian et al. 2015 [158] | Hereditary Cancer | N/A | N/A | N/A | N/A | E | N/A | N/A | Pan-cancer panel testing |
| 62 | Alu | NF1 | Wallace et al. 1991 [23] | NF1 | AluYa5 | 282 | 40 | N | N | AS | I/Splicing | AAAAAAAAACAT | TTTT/AA |
| 63 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluY | 280 | N/A | N | N | S | I | AAAAAATTCCAG | TTTT/AA |
| 64 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluY | 281 | N/A | N | N | AS | I | N/A |
| 65 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluYa5 | 282 | 60 | N | N | S | E | ATAAAAAGCCTGGA | TTAT/AA |
| 66 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluYa5 | 284 | 120 | N | N | AS | E | AAAAAACCTGCT | TTTT/AA |
| 67 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluYa5 | 281 | N/A | N | N | AS | E | AAAAAACCTGCTGATGG | TTAT/AA |
| 68 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluYa5 | 284 | 110 | N | N | AS | E | ATAAAAACCTAAAGA | TTAT/AA |
| 69 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluYa5 | 279 | N/A | N | N | S | E | AAAAAAGAAGACATAT | TTAT/AA |
| 70 | Alu | NF1 | Wimmer et al. 2011 [296] | NF1 | AluYa5 | 264 | 60-85 | Y/5'TR | N | AS | E | AAAAAAGATGCCTACCT | TTAT/AA |
| Table 1 | Retrotransposition events associated with human disease (Continued) |
|---------|---------------------------------------------------------------|
| 71      | Alu  | NF1 | 17      | Wimmer et al. 2011 [296] | NF1 | AluYb8 | 249 | 121 | Y5TR | N | S | E | AAAGCAGTGTC | CTIT/AT |
| 72      | Alu  | NF1 | 17      | Wimmer et al. 2011 [296] | NF1 | AluYb8 | 288 | N/A | N | N | AS | I | AAAAAAGAGAAGACAA | TTTT/AA |
| 73      | Alu  | NF1 | 17      | Wimmer et al. 2011 [296] | NF1 | AluYb8 | 289 | 120 | N | N | AS | E | ACAAATGCTGTTT | TGTT/AA |
| 74      | Alu  | NF1 | 17      | Wimmer et al. 2011 [296] | NF1 | AluYb8 | 288 | 78- | 178 | N | S | E | AAAAAAGAGACATA | TTTT/GT |
| 75      | Alu  | NF1 | 17      | Wimmer et al. 2011 [296] | NF1 | AluYb8 | 288 | 118 | N | N | S | E | AAAAAAGAGACATAT | TTTT/GT |
| 76      | Alu  | NF1 | 17      | Wimmer et al. 2011 [296] | NF1 | AluYb8 | 288 | 78- | 178 | N | N | S | E | AAAAAAGAGACATA | TTTT/GT |
| 77      | L1   | CYBB | X | Meischl et al. 1998 [297], Brouha et al. 2002 [181] | CGD | L1 Ta | 1722 | 101 | Y5TR | Y (280) | S | E | AA | AAAAAACAACAATA | TGTT/GA |
| 78      | L1   | CYBB | X | Meischl et al. 2000 [298] | CGD | L1 Ta | 836  | 69  | Y5TR/INV | N | S | S/Splicing | AGAAATAAATTATAA | TTCT/AA |
| 79      | L1   | CHM  | X | van den Hurk et al. 2003 [177] | Choroideremia | L1 Ta | 6017 | 71  | FL | Y (119/406) | AS | E | AGAAGATCAAATTAG | TTCT/AA |
| 80      | L1   | DMD  | X | Musova et al. 2006 [299] | DMD | L1 Ta | 452  | 41  | Y5TR/INV | N | AS | E | AAATATCTTATATA | ATTT/AA |
| 81      | L1   | DMD  | X | Narita et al. 1993 [164] | DMD | L1 Ta | 608  | 16  | Y5TR | N | AS | E | No TSD | TCTT/AA |
| 82      | L1   | DMD  | X | Holmes et al. 1994 [176] | DMD | L1 Ta | 1400 | 38  | Y5TR/INV | Y (489) | S | E | AAATCATGCTGCTG | ATTT/AA |
| 83      | L1   | DMD  | X | Yoshida et al. 1998 [300] | XLDCM | L1 Ta | 530  | 73  | Y5TR | N | AS | S5′UTR/Loss of mRNA | AAAAAACCTGTTAAA | TTTT/AT |
| 84      | L1   | DMD  | X | E Bakker & G van Omenn, personal communication | DMD | N/A | 878  | N/A | Y5TR | N | S | N/A | N/A | N/A | N/A |
| 85      | L1   | DMD  | X | Awano et al. 2010 [301], Solyom et al. 2011 [302] | DMD | L1 Ta | 212  | 118 | Y5TR | Y (212) | AS | E | GAA | TTCT/AA |
| 86      | L1   | FVIII | X | Kazazian et al. 1988 [22] | Hemophilia A | L1 Ta | 3800 | 54  | Y5TR | N | S | E | AAGACAAAACAAAAC | TTCT/AA |
| 87      | L1   | FVIII | X | Kazazian et al. 1988 [22] | Hemophilia A | L1 preTa | 2300 | 77  | Y5TR/INV | N | AS | E | AATGTGCTTCTTCTTTC | CATT/AA |
| 88      | L1   | FIX  | X | Li et al. 2001 [266] | Hemophilia B | L1 Ta | 463  | 68  | Y5TR | N | S | E | AAAATATGCTGATA | TTTT/AC |
| 89      | L1   | FIX  | X | Mukherjee et al. 2004 [303] | Hemophilia B | L1 Ta | 163  | 125 | Y5TR | N | S | E | GAAAATGGATTG | TTTC/AT |
| #  | L1 | Gene   | Source, Year | Disease/Description | L1 Type | Ta/PreTa | Fl/Fl | N/A | S/CDS | Insertion/Deletion | TSD | 3'-Processing | Allele Frequency | Notes |
|----|----|--------|--------------|--------------------|---------|---------|------|-----|------|-------------------|-----|-------------|-----------------|------|
| 90 | L1 | RP2    | Schwahn et al. 1998 [304] | XLRP | L1 Ta | 6000 | FL | N | S | Loss of mRNA | AAGACTGTAAGGTG | TCTT/AA | Interrupted polyA |
| 91 | L1 | RPS6KA3 | Martinez-Garay et al. 2003 [305] | Coffin-Lowry syndrome | L1 Hs | 2800 | Yes | Y/TR/INV | N | AS | E | AAGAAACCTGATTTT | TCTT/AG |
| 92 | L1 | ABDH5  | Samuelov et al. 2011 [306], Eli Sprecher, personal communication | CDS | N/A | FL | N/A | N | N/A | Splicing | N/A | N/A |
| 93 | L1 | MLH1   | 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 |
| 94 | L1 | MLH1   | 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 |
| 95 | L1 | APC    | Miki et al. 1992 [200] | Colon cancer | L1Ta | 520 | 222 | Y/TR/INV | N | S | E | AAGAATAATG | TCTT/AA | Somatic Insertion/same insertion site as Halling et al. Alu |
| 96 | L1 | EYA1   | Morisada et al. 2010 [247] | BOR syndrome | L1 Hs | 3756 | None | Y/TR | N | AS | 17 kb Deletion | No TSD | TCTC/AG | Internal Priming |
| 97 | L1 | FKTN   | Kondo-Iida et al. 1999 [307] | FCMD | L1Ta | 1200 | 59 | Y/TR | N | S | Splicing | No TSD | TTTT/AA |
| 98 | L1 | SETX   | Bernard et al. 2009 [308], Christine Zühlke, personal communication | AOA2 | L1 Hs | 1300 | 42 | Y/TR/INV | N | S | E | GGAAGAATGTGAATTGCTA | TTCC/AG | 3'-processing 2 nt (5'-CC-3') |
| 99 | L1 | PTEN   | Helman et al. 2014 [201] | endometrial carcinoma | L1 Hs | 90 | 22 | Y/TR | N | S | E | AAAGAATCTTGGATATAG | TCTT/AA | Somatic Insertion |
| 100| L1 | HBB    | Divoky et al. 1996 [309] | β-thalassemia | L1 Ta | 6000 | 107 | FL | N | AS | I | AAAATAAAGCAGA | TTTT/AT |
| 101| L1 | PDHX   | Mine et al. 2007 [310] | PDCD4 deficiency | L1 Hs | 6086 | 67 | FL | N | S | 46 kb Deletion | No TSD | TTTT/AT |
| 102| L1 | SLC25A1 | Kagawa et al. 2015 [157] | Rotor syndrome | L1 Ta-1d | 5089 | 100 | Near FL | N | S | Splicing | AAGAATTATAGTGACCTGTA | TCTT/AC | 0.054 Japanese Allele Frequency, may be "Hot L1" |
| 103| L1 | RB1    | Rodriguez-Martín et al.2016 [202] | Familial Retinoblastoma | L1 Ta-1d | 6044 | 33 | FL | N | S | Splicing | AAATTATCTGTGTTT | ATTT/AA | N/A |
| 104| L1 | NF1    | Wimmer et al. 2011 [296] | NF1 | L1 preTa | 1800 | N/A | Y/TR | N | S | E | AAAAGCAAACGTGTGT | TTTT/AT | Untemplated 3'-T? |
| 105| L1 | NF1    | Wimmer et al. 2011 [296] | NF1 | L1 Ta | 6000 | N/A | FL | N | S | E | AAAATCGAGGG | TTTT/AA | Untemplated 3'-T? |
| 106| L1 | NF1    | Wimmer et al. 2011 [296] | NF1 | N/A | 2200 | N/A | Y/TR/INV | N | AS | Splicing | AAGAAAATGGT | TCTT/AA |
| SVA | Gene | Disease | Reference | Haplotype | Flank | AS | Splicing Site | Additional Information |
|-----|------|---------|-----------|-----------|-------|----|---------------|------------------------|
| 107 | SVA  | BTK     | X         | Rohrer et al. 1999 [311], Conley et al. 2005 [258] | N/A   | 92 | Y/STR        | AGAAATGTATGAGTAA       | TTCT/AT               |
|     |      |         |           |           |       |    |               | Same insertion site as Conley et al. Alu |
| 108 | SVA  | TAF1    | X         | Makino et al. 2007 [312] | F     | 62 | FL            | AAAAAAATGAAATAG        | TCCT/AT               |
|     |      |         |           |           |       |    |               | 3′-Processing 3 nt (5′-CCT-3′) |
| 109 | SVA  | FIX     | X         | Nakamura et al. 2015 [156] | F     | 28 | FL            | AAATGGCACTAGAA         | TCC/CAT               |
|     |      |         |           |           |       |    |               | 3′-Processing 1 nt (5′-C-3′) |
| 110 | SVA  | LDRAP1  | 1         | Wilund et al. 2002 [313] | E     | 57 | FL            | GAAACCTGTTTTTCCT       | TTT/C/AA              |
|     |      |         |           |           |       |    |               | Orphan 3′-transduction |
| 111 | SVA  | SPTA1   | 1         | Hassoun et al. 1994 [314], Ostertag et al. 2003 [24] | E     | 50 | Y/STR/INV    | GAAATTGAAGACTCAGTT    | TTT/C/AA              |
|     |      |         |           |           |       |    |               | Orphan 3′-transduction |
| 112 | SVA  | CASP8   | 2         | Stacey et al. 2016 [203] | E     | 45 | N/A           | AAGAATTGTA            | TCT/C/AT              |
|     |      |         |           |           |       |    |               | Protective against prostate cancer; active locus? |
| 113 | SVA  | A4GIN3  | 3         | Nazaryan et al. 2015 [155] | E     | 25 | None          | GAAACCTGTTTTCTC       | TTT/C/AC              |
|     |      |         |           |           |       |    |               | First report of large scale rearrangement and an insertion. Implicates retrotransposition in germline chromothripsis. |
| 114 | SVA  | HLA-A   | 6         | Takasu et al. 2007 [315] | F1    | 45 | FL            | N/A                   | CTT/T/AG              |
|     |      |         |           |           |       |    |               | Novel SVA subfamily (F1) |
| 115 | SVA  | PMS2    | 7         | van der Klift et al. 2012 [154] | F     | 64 | Y/STR/INV    | AAGAATGCGGATGTA       | TCT/C/AA              |
|     |      |         |           |           |       |    |               | SVA exonization |
| 116 | SVA  | FKTN    | 9         | Kobayashi et al. 1998 [162] | E     | 32 | FL            | AAGAATGCGGATGTA       | TCT/C/AA              |
|     |      |         |           |           |       |    |               | SVA exonization |
| 117 | SVA  | PNPLA2  | 11        | Akman et al. 2010 [316] | E     | 32 | FL            | AAGAATGCGGATGTA       | TCT/C/AA              |
|     |      |         |           |           |       |    |               | SVA exonization |
| 118 | SVA  | SUZ1P   | 17        | Vogt et al. 2014 [153] | F1    | 23 | Y/STR/INV    | AAAACTGAA            | TTT/T/AC              |
|     |      |         |           |           |       |    |               | Largest reported insertion associated deletion (~1 Mb), somatic |
| 119 | SVA  | SUZ1P   | 17        | Vogt et al. 2014 [153] | F     | 40 | Y/STR/INV    | AAAACTGGCA            | TTT/T/AC              |
|     |      |         |           |           |       |    |               | 867 kb deletion, somatic |
| 120 | Processed Pseudogene | CYBB | X | de Boer et al. 2014 [152] | N/A | 100 | FL | No | AAAACTGAA | TTT/T/AC | First reported de novo processed pseudogene (TMF1) |
Table 1 Retrotransposition events associated with human disease (Continued)

| #  | Retrotransposition Event | Gene | Disease | Exon | Nucleotide | TSD | Description | References |
|----|--------------------------|------|---------|------|------------|-----|-------------|------------|
| 121 | pA                       | COL4A6 | X | Segal et al. 1999 [317] | Alport syndrome | N/A | N/A | 70 | 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 | 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 | 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 | AS | E | AAGAAA | TCTTNAA |

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, "","", 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 Chanarin-Dorfman syndrome, CGD Chronic granulomatous disease, DMD Duchenne muscular dystrophy, FAP Familial adenomatous polyposis, FCMD Fukuyama-type congenital muscular dystrophy, FHH and NSHPT Familial hypocholesteric hypercalcaemia and neonatal severe hyperparathyroidism, GHKG Glycerol kinase deficiency, HAE Hereditary form of angioedema, HE and HPP Hereditary elliptocytosis and hereditary pyropoikilocytosis, HIGM Hyper-immunoglobulin M syndrome, HNPPC Hereditary non-polyposis colorectal cancer syndrome, LPL Lipoprotein lipase, MLII Mucolipidosis Type II, MWS Mowat-Wilson syndrome, NF1 Neurofibromatosis Type I, PDCD 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 Alu 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 (TTTTT)) [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 splicesomal 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 contain 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 retrotransponson 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. 2) [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-1α (HP1α). 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 NFI gene. Recently, while characterizing genetic deletions in the neurofibromin 1 (NFI) gene, an SV 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 s in the brain [180] and the other studying LINE-1...
de novo \textsuperscript{′}–the localized with a gene \textsuperscript{′}–SLCO1B3 may appear 5\textsuperscript{[181]}, which is one of the most active \textsuperscript{−}retrotransposition events are likely be-

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 \textsuperscript{[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\textsubscript{SLCO1B3} may represent a population-specific “hot” LINE-1\textsuperscript{′} with a gene frequency of \(\sim 6\%\) in Japanese individuals. LINE-1\textsubscript{SLCO1B3} contains intact reading frames with ORF1 being 100 \% identical to the LINE-1 amino acid consensus and LINE-1\textsubscript{SLCO1B3} 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\textsubscript{DES1} \textsuperscript{[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) \textsuperscript{[40]}. The first nucleotide of LINE-1\textsubscript{SLCO1B3} –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\textsubscript{SLCO1B3} 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 \textsuperscript{[155]} analyzing a germ-line chromosome shattering event - a phenomenon referred to as chromothripsis \textsuperscript{[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 \textsuperscript{[155]}.

To date, the DNA damage agents causing chromothripsis and the mechanisms driving the rearrangement of chromosomal fragments are poorly understood. \textsuperscript{[184–186]}. Several hypotheses have been generated to explain chromothripsis \textsuperscript{[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 \textsuperscript{[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 \textsuperscript{[188]}.

Although the new study \textsuperscript{[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 \textsuperscript{[188, 189]} and cancers \textsuperscript{[190]}, is thought to be caused by the LINE-1 restriction factors \textsuperscript{[94, 191–194]}- the APOBEC3 proteins \textsuperscript{[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 \textsuperscript{[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 cytochrome 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 Transpo-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/AAATTATCTGTTT/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 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; APOBECA3: 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; CENFW: 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; EN: Endonuclease-independent; ERV: endogenous retroviruses; EVE: endogenous viral elements; FKTN: fukutin; F8: factor VIII; HAL1: half-LINE-1; HP1α: heterochromatin protein-1α; KAP1: KRAB-associated protein-1; LINE-1: L1; LTR: Long Terminal Repeats; MAST2: microtubule associated serine/threonine kinase 2; NF1: neurofibromin 1; NAHR: non-allelic homologous recombination; PRR: pattern-recognition receptor; PGBDS: PiggyBac transposable element-derived protein S; Poc1a: Poc1 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 epitheliun; RRMA: RNA recognition motif; RT: reverse transcriptase; RUNX: runt-related transcription factor 3; SAMHD1: SAM domain and HD domain 1; SIN: Short Interspersed Element; SVA: SINE-VNTR-Alu; SIRT6: Sirtuin-6; SD: splice donor sites; SCG181: solute carrier organic anion transporter family member 181; SD: splice donor sites; SCG181: solute carrier organic anion transporter family member 181; TCEA: T-cell factor/lymphoid enhancer factor; TE: transposable element; THAP: THAP domain containing 9; TRPT: target-primer reverse-transcription; TREN: 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;420:202–60.
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, Ldhoyczki J, Levine R, McEwan P, McKenan 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 F, Epstein CB, Kaul R, Lajoie BR, Rosenbloom KR, Song L, Trinklein ND, Altshuler RC, Brown JH, Dong X, Hardison RC, Harris RS, Iyer S, Kheradpour P, Kudnaje A, Lastman T, Li Q, Merkel A, Thurman RE, Wu W, Yip KY, Bernstein BE, Birney E, Dunham J, Pazin MJ, Powell R, 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. Dittrich K, Xiang W, Han J, Lu X, 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 HIT, Smith JD, Pitham EJ, Aarenburger P, Atkinson PW, Craig NL. Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Res. 2008;18:717–28.
19. Agrama A, Eastman QM, Schatz DG. Implications of transposition mediated by VDII-J-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity. Nature. 1998;394:744–51.
20. Heissen AG, Henaff E, Jiang E, Eisenberg AR, Carson JR, Villasante CM, Ray MS, Still E, Burns M, Gandara J, Feschotte C, Mason CE, Kentsis A. Genomic DNA transposition induced by human PGBDS. 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, Youssoufian 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;353:364–6.

24. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. SVA elements are nonautonomous retrotransposons that cause disease in humans. Ann Hum Genet. 2003;67:1444–51.

25. Ewing AD, Kazazian HH. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 2010;20:262–70.

26. Macfarlane OM, Badge RM. Genome-wide amplification of proviral sequences reveals new polymorphic HERV-H(HML-2) proviruses in humans and chimpanzees that are absent from genome assemblies. Retrovirology. 2015;12:161.

27. Marchi E, Kasanip A, Byott M, Magiorkinis G, Belshaw R. Neanderthal and Denisovan retroviruses in modern humans. Curr Biol. 2013;23R894–5.

28. Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M. Genome-wide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K(HML-2): Implications for Present-Day Activity. J Virol. 2005;79:2507–14.

29. Wildschutte JH, Williams ZH, Montesion MJ, Subramanian RP, Kidd JM, Coffin JM. Discovery of unified 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 ME, 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;10:448–56.

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. Zhou X, Freschotte C. Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammalian 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'Malley BJ, Moran JV, Kazazian Jr HH, Boeke JD. 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-Prieger S, Farley AH, Moran JV, Kazazian HH. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci. 2003;100:2820–5.

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 and characterization of a human transposable element. Science. 1999;283:1530–8.

42. Brouha B, Schustak J, Badge RM, Lutz-Prieger S, Farley AH, Moran JV, Kazazian HH. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci. 2003;100:2820–5.

43. Sanders AM, 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.

44. 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.

45. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. Isolation and characterization of a human transposable element. Science. 1999;283:1530–8.

46. Brouha B, Schustak J, Badge RM, Lutz-Prieger S, Farley AH, Moran JV, Kazazian HH. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci. 2003;100:2820–5.

47. Sanders AM, 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.

48. 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.

49. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. Isolation and characterization of a human transposable element. Science. 1999;283:1530–8.
102. Chu WM, Liu WM, Schmid CW. RNA polymerase III promoter and terminator
101. Paolella G, Lucero MA, Murphy MH, Baralle FE. The Alu family repeat
100. Ullu E, Tschudi C. Alu sequences are processed 7SL RNA genes. Nature.
97. Grimaldi G, Queen C, Singer MF. Interspersed repeated sequences in the
96. Houck CM, Rinehart FP, Schmid CW. A ubiquitous family of repeated DNA
95. Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y,
93. Han JS, Shao S. Circular retrotransposition products generated by a LINE
90. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. Fitness cost of LINE-1
89. Boissinot S, Entezam A, Furano AV. Selection against deleterious LINE-1-
88. Ostertag EM, Kazazian HH. Twin Priming: A Proposed Mechanism for the
99. Weiner AM. An abundant cytoplasmic 7S RNA is complementary to the
85. Kopera HC, Moldovan JB, Morrish TA, Garcia-Perez JL, Moran JV. Similarities
84. Gilbert N, Lutz S, Morrish TA, Moran JV. Multiple Fates of L1 Retrotransposition
83. Gilbert N, Lutz-Prigge S, Moran JV. Genomic Deletions Created upon LINE-1
82. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD.
78. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 Retrotransposition is Associated with Genetic Instability In Vivo.
76. Dai L, Taylor MS, O’Donnell KA, Boeke JD. Poly(A) binding protein C1 is 
75. Goodier JL, Cheung LE, Kazazian HH. Mapping the LINE1 ORF1 protein 

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. Conti A, Camevalli 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;43:3817–35. 
80. Sorek R, Ast G, Graur DT. Alu-Containing Exons are Alternatively Spliced. 
Genome Res. 2002;12:1060–7. 
79. Quintero 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. 
78. Quintero 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. 
77. Jurka J, Zuckermandl E. Free left arms as precursor molecules in the 
evolution of Alu sequences. J Mol Evol. 1991;33:49–56. 
76. Chang DY, Sasaki-Tozawa N, Green LX, Mariai 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. 
75. 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. 
74. 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. 
73. 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. 
72. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM. LINE-1 ORF1 protein 
enhances Alu SINE retrotransposition. Gene. 2008;419:1–6. 
71. Wildschutte JH, Baron A, Drolff NM, Kidd JM. Discovery and characterization 
of Alu repeat sequences via precise local read assembly. Nucleic Acids Res. 2015;43:10292–307. 
70. Medstrand P, van de Lagaart MN, Lager DL. Retroelement Distributions in 
the Human Genome: Variations Associated With Age and Proximity to 
Genes. Genome Res. 2002;12:1483–95. 
69. Smallheiser NR, Torvik VI. Alu elements within human mRNAs are probable 
microRNA targets. Trends Genet. 2006;22:532–6. 
68. Spengler RM, Oakley DK, Davidson BL. Functional microRNAs and target 
sites are created by lineage-specific transposition. Hum Mol Genet. 2014;23: 
1783–93. 
67. Chen L-L, DeCrocco JN, Carmichael GG. Alu element-mediated gene 
silencing. EMBO J. 2008;27:1694–705. 
66. Kim DDF, Kim THF, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. 
Widespread RNA Editing of Embedded Alu Elements in the Human 
Transcriptome. Genome Res. 2004;14:1719–25. 
65. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA Editing of Alu 
Containing mRNAs in the Human Transcriptome. PLOS Biol. 2004;2:e91. 
64. Gong C, Maquat LE. IncrRNAs transactivate STA1-mediated mRNA 
decay by duplexing with 3’[prime] UTRs via Alu elements. Nature. 2011;470:284–8. 
63. Han K, Konkel MK, Xing J, Wang H, Lee I, Meyer TJ, Huang CT, Sandifer E, 
Hebert K, Barnes EW, Hubley R, Miller W, Smit APA, Ullmer B, Batzer MA. 
Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science. 2007;316:28–40. 
62. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA 
Elements: A Hominid-specific Retrotransposon Family. J Mol Biol. 2005;354:994–1007. 
61. Shen L, Wu LC, Saniloglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, 
Zhang Y. Structure and genetics of the partially duplicated gene RP- 
lb located immediately upstream of the complement C4A and the C4B genes in 
the HLA class II region. Molecular cloning, exon-intron structure, composite 
retrotransposon, and breakpoint of gene duplication. J Biol Chem. 1994;269:4666–76. 
60. Darnett A. Composite non-LTR retrotransposons in hominid primates. Mob 
Genet Elements. 2015;5:65–71. 
59. Hancks DC, Kazazian HH. SVA retrotransposons: Evolution and genetic 
instability. Semin Cancer Biol. 2010;20:234–45. 
58. Darnett A, Raiz J, Horn AV, Lowier J, Wang H, Xing J, Batzer MA, Lowier R, 
Schumann G. SVA Transcribing SVA retrotransposition 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. Transretroposition of marked SVA elements by human L1 in cultured cells. Hum Mol Genet. 2011;20:23386–400.

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;32:7478–26.

131. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23:183–9.

132. Bantysh OB, Buzdin AA. Novel family of human transposable elements formed due to the fusion of the first exon of gene MAST2 with retrotransposon SVA. Biochemistry Mosc. 2009;74:393–9.

133. Raiz J, Damert O, Chira S, Held U, Klawitter J, Shoma B, Stratling J, Catterall A, Buzdin A, Ustyugova S, Gogvadze E, Volchkov P, Kovalskaya E, Illarionova A, Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E. Ancient repeat sequence in the human genome contains many copies of a novel retrotransposon family. Mol Biol Evol. 2012;24:648–58.

134. Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Miller DC, Batzer MA, de Jong PJ. Centromere remodelling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons. Genome Biol. 2012;13:1466–83.

135. Lupan I, Bulzu P, Popescu O, Damert A. Hominoid composite non-LTR retrotransposons: assembly, evolution and structural determinants of mobilization. Mol Biol Evol. 2014;31:2947–54.

136. Carbone L, Harris RA, Görner SE, Veeramah KR, Lorente-Galdos B, Huddleston J, Miller DC, Batzer MA, de Jong PJ. Centromere remodelling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons. Genome Biol. 2012;13:1466–83.

137. Smith RJH, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Miller DC, Batzer MA, de Jong PJ. Centromere remodelling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons. Genome Biol. 2012;13:1466–83.

138. Moran JV. Human L1 retrotransposition: insights and peculiarities learned from studies of L1 insertion in human genomes. Heredity. 2012;108:38–49.

139. Hara T, Hirai Y, Baicharoen S, Hayakawa T. A novel composite retrotransposon derived from or generated independently of the SVA (SINE/LTR) transposable element unique to the gibbons. Genome Biol. 2012;13:648–58.

140. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 2009;19:1983–91.

141. Hancks DC, Goodier JL, 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;32:7478–26.

142. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23:183–9.

143. Bantysh OB, Buzdin AA. Novel family of human transposable elements formed due to the fusion of the first exon of gene MAST2 with retrotransposon SVA. Biochemistry Mosc. 2009;74:393–9.
172. Strichman-Almasshanu LZ, Lee RS, Oryanggo PO, Perlman E, Flaim F, Friedman 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-Gonzalez J, Morell M, Schjoedt J, Kigusa MR, Cramer CG, Hammer GD, Collins KL. O’Shea KS, Menendez P, Morgan EP. Genomic silencing of engineered LINE-1 retrotransposition events in human embryonic carcinoma cells. Nature. 2010;464:769–73.

174. Van Meter M, Kashyap M, Rezazadeh S, Geneve A, Morell TD, Sekulovac N, Gorbonova V. SIRT6 regulates LINE-1 retrotransposition by ribosylating Kap1 but this repression fails with stress and age. Nat Commun. 2014;5:1–10.

175. Wolff MW, 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, Dombrski BA, Krebs CM, Boehm CD, Kazazian HH. A new retrotranspositional human LINE1 element from the L2E2 focus on chromosome 1q produces a chimeric insertion. Nat Genet. 1994;14:1–8.

177. van den Huij JAJM, Meij MC, Valier SM, Kano H, Nikopoulos K, Hoeflooth LH, Sisternas EA, de Wijt J, Mukhopadhyay A, Plomp AS, de Jong PTVM, Kazazian HH, Cremers FMJ. LINE1 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. LINE1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev. 2009;23:1033–12.

179. Kazazian HH. Mobile DNA Retrotransposition in Somatic Mosaicism. BMC Biol. 2011;9:162.

180. Muotri AR, Chu VT, Marchetto MNC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by LINE1 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 LINE1 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 hyperpermutable, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations. Mutat Res. 2011;705:2–9.

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 karyotype evolution from complex chromosomal rearrangements. Genes Dev. 2013;27:179–200.

188. Maciejewski JI, Ly J, 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 Loosdrecht CM, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Davies HR, Papeammanuel 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, Bignell GR, Bignell CG, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhartt B, Butler AP, Caldas C, Davies HR, Desmet V, Ellis RJ, Eyfjord JE, Folkers KJ, Gaveva M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinska M, Jager N, Jones DTW, Kamb B, Khattra J, Kohutek ZA, Huang CR, Ahn D, Mita P, Taylor MS, Barker NJ, Hruban RH, lacobuzio-Donahue CA, Vogelstein B, Kazazian HH, Solymon S. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. Genome Res. 2015;25:1536–45.

191. Rodic N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, Kihoko TA, Huang CR, Ahn D, Mita P, Taylor MS, Barker NJ, Hruban RH, lacobuzio-Donahue CA, Boeke JD, Burns KH. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nat Med. 2015;21:1060–4.

192. Doucet-O’Hare TT, Rodic N, Sharma R, Darbafi I, Abtig B, 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 cancer. Proc Natl Acad Sci. 2015;112:489–900.

193. Shukla R, Upadhye GK, Gunraj A, Office DJ, Fisher ME, Nguyen T, Brennan PM, Baleeck J, Collino A, Ghisletti S, Sinha S, Iannelli F, Radielli A, Dos Santos A, Rapoud D, Guertler C, Samuel D, Natali G, Caminci P, Ciccarelli FD, Garcia-Perez JL, Faivre J, Faulkner GJ. Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma. Cell. 2014;153:101–11.

194. Solymon S, Ewing AD, Rahmann EP, Doucet T, Nelson HH, Burns MB, Harris RS, Sigmon DF, Casella A, Erlanger B, Wheelan S, Upadhye RR, Shukla R, Faulkner GJ, Largespada DA, Kazazian HH. Extensive somatic LINE1 retrotransposition in colorectal tumors. Genome Res. 2012;22:3238–38.

195. 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;7:4086–11.
Mobile DNA (2016) 7:9

222. Zhao K, Du J, Han X, Goodier JL, Li P, Zhou X, Wei W, Evans SL, Li L, Zhang Y, et al. The Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer. Oncotarget. 2013;4:3093–51.

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:E5699–71.

225. Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP Synthase Is a Cytosolic autoantigen binds endogenous retroelements and regulates inflammatory mechanism mediated by retrotransposon-encoded reverse transcriptase is active in human transformed cell lines. Oncotarget. 2013;4:2271–87.

226. Srikanta D, Sen SK, Huang CT, Conlin EM, Rhodes RM, Batzer MA. An endonuclease-independent insertion mechanism for human retrotransposition. Nucleic Acids Res. 2007;35:3741–51.

227. Hung T, Andreola F, Sciamanna I, Landriscina M, Pittoggi C, Quirino M, Mearelli C, Beraldi R, Oricchio E, Sciamanna I, Beraldi R, Tolstonog GV, Schumann GG, Spadafora C. The Reverse Transcriptase Encoded by Line-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer. Oncotarget. 2013;4:3093–51.

228. Sciamanna I, Gualtieri A, Cotsettii C, Osimo EF, Ferracin M, Macchia G, Arico E, Proseda G, Vitullo P, Misteli T, Spadafora C. A tumor-promoting mechanism mediated by retrotransposon-encoded reverse transcriptase is active in transformed cell lines. Oncotarget. 2013;4:2271–87.

229. Sciamanna I, Gualtieri A, Corsetti C, Oslmo EF, Ferracin M, Macchia G, Arico E, Proseda G, Vitullo P, Misteli T, Spadafora C. A tumour-promoting mechanism mediated by retrotransposon-encoded reverse transcriptase is active in transformed cell lines. Oncotarget. 2013;4:2271–87.

230. Sen SK, Huang CT, Han K, Batzer MA. Endonuclease-independent insertion mechanism for LINE-1 retrotransposition provides an alternative pathway for L1 retrotransposition in the human genome. Nucleic Acids Res. 2007;35:3741–51.

231. Flewler BJ, Gelfand BD, Kim Y, Kerer N, Tarallo V, Hirano Y, Amanath S, Fowler DH, Radwan M, Young MT, Pittman K, Kubcs P, Hinton DR, Kranzig H, Hakala HH, Parang K, Hinton DR, Bastos-Carvalho A, L S, Yasuma T, Mizutani T, Yasuma R, Wright A, Ambati J, Bodea GO, Ewert AD, Salvador-Palomeque c, van der Kraan MS, Brennam PM, Rizzi P, Smith F, Meit T, Talbot RT, Gustinich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carinci P, Jledelioh JA, Faulkner GJ. Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011;479:534–7.

232. Upton KR, Gerhardt DJ, Jesudian JS, Richardson SR, Sanchez-Luque FJ, Bodea GO, Ewert AD, Salvador-Palomeque c, van der Kraan MS, Brennam PM, Vandenee A, Faulkner Gi. Ubiquitous L1 Mosaicin in Hippocampal Neurons. Cell. 2015;161:228–39.

233. Evryo GD, Cai X, Lee E, Hills LL, Elhsoys 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.

234. Richardson SR, Morell S, Faulkner GJ. L1 retrotransposons and somatic mosaicin in the brain. Annu Rev Genet. 2014;48:1–27.

235. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovc M, Morell S, O’Shea KS, Morrisov JV, Gage FH. L1 retrotransposition in human neural progenitor cells. Nature. 2009;460:1127–31.

236. Mucrri AR, Marchetto MCN, Coufal NG, Oelfner R, Yeo G, Nakashima K, Gage FH. L1 retrotransposition in neurons is modulated by MeCP2. Nature. 2010;468:446–53.

237. Evryo GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann H, Park P, Walsh CA. Cell lineage analysis in human brain using endogenous retroelements. Neuron. 2015;85:49–59.

238. Skowronski J, Fanning TG, Singer MF. Unit-length L1 transcripts in human teratocarcinoma cells. Mol Cell Biol. 1988;178:795–8.

239. 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.

240. Rangwala SH, Zhang L, Kazazian HH. Many LINE1 elements contribute to the transcriptome of human somatic cells. Genome Biol. 2009;10:110.

241. 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.

242. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nature Rev Genet. 2002;3:370–9.

243. Corneaux 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.

244. Richardson SR, Doucet AJ, Kopper HC, Moldovan JB, Garcia-Perez JL, Moran JV. The Influence of LINE-1 and SINE Retrotranspositions on Mammalian Genomes. Microbiol Spectr. 2015;3:MDNA3–0061–2014.

245. Szak SZ, Pickler OK, Landsman D, Boeke JD. Identifying related L1 retrotransposons by analyzing 3 transduced sequences. Genome Res. 2003;13:4930.

246. Morisada N, Rendttoff ND, Nozu K, Morishita T, Miyakawa T, Matsumoto T, Hisano S, Iijima K, Tranbejæ AJ, 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.

247. Sheen FM, Sherry ST, Risch GM, Robichaux M, Albuquerque RJC, Hauswirth WW, Chiodo VA, Karikó K, Yoo JW, Lee D-K, 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.

248. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci M, Morell S, O’Shea KS, Morrisov JV, Gage FH. L1 retrotransposition in human neural progenitor cells. Nature. 2009;460:1127–31.

249. Mucrri AR, Marchetto MCN, Coufal NG, Oelfner R, Yeo G, Nakashima K, Gage FH. L1 retrotransposition in neurons is modulated by MeCP2. Nature. 2010;468:446–53.

250. Evryo GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann H, Park P, Walsh CA. Cell lineage analysis in human brain using endogenous retroelements. Neuron. 2015;85:49–59.

251. Skowronski J, Fanning TG, Snger MF. Unit-length L1 transcripts in human teratocarcinoma cells. Mol Cell Biol. 1988;178:795–8.

252. 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.

253. Bao W, Jurka J. Origin and evolution of LINE-1 derived retrotransposons from the L1 family of interspersed repetitive DNA sequences. J Mol Biol. 2003;35:363–6.

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, Reseler B, Katzer H-G, Orth U, Gillessen-Kaesbach G, Morlot S, Schwingler G, 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, Watarabe S, Kikuchi N, Ishitsuka I, Ozawa H, Fujisawa C, Shiga K. The first reported case of Monks disease caused by an Alu insertion mutation. Brain Dev. 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 Genet Suppl. 1997;34:S81.

258. Conley ME, Partain JD, Norland SM, Shurtleff SA, Kazazian HH. Two independent retrotransposition insertions at the same site within the coding region of 87X. Hum Mutat. 2005;25:324–5.

259. Apoll PA, Kuhlein E, Robert A, Rubie H, Blancher A. HGM syndrome caused by insertion of an AluYb8 element in exon 1 of the CD40LG gene. Immunogenetics. 2006;59:17–23.

260. Claverie-Martin F, Gonzalez-Acosta H, Flores C, Antón-Gamero M, García-Sanz M. Insertion of a human-specific Alu subfamily member within the coding region of the CLCN5 gene results in Dent’s disease. Hum Genet. 2003;113:480–5.

261. Sukarova E, Dimitrov AI, Tchacharova P, Perkov 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 hemorrhagicemia A. Hum Genet. 2003;113:348–52.

263. Green PM, Briand JD, Waseem NM, Gannelli F. Haemophilia A mutations in the UK: results of screening one-third of the population. Br J Haematol. 2008;143:115–8.

264. Vidalh D, Vidaud M, Bahnak BR, Sigaret 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. Wulf 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 IX gene. Hum Mutat. 2000;15:1–2.

266. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Caret 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 MW, Vilain E, Huang BL, Finlayson G, Terrell BL, Worley K, Deininger P, McCabe ER. AluY insertion (IVS5-41ins135ea) in the glycerol kinase gene from an individual with benign glycogen 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 der Burg SC. Characterization of five Portuguese patients with pyrimidine 5'-nucleotidase deficient hemolytic anemia showing three new PN5 mutations. Blood Cell Mol Dis. 2006;9:266–7.

269. Chen J-M, Masson E, Macek Jr M, Ragatelli O, Piskavcova T, Fercot B, Fila L, Cooper DD, Audrézet M-F, Pérez C. Detection of two Alu insertions in the CFTR gene. J Cyst Fibros. 2007;6:37–43.

270. Abdelhak S, Kalatzis V, Heilig R, Compan S, Samson D, Vincent C, Lecci Acobas F, Cruaud C, Le Merier M, Mathieu M, König R, Vigneron J, Weissenbach J, Petric C, Well D. Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of ZFY1. Hum Mol Genet. 1997;6:2247–55.

271. Okubo M, Hirohi taka A, Saito M, Ebara T, Endo Y, Kuku 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.

272. Udagawa T, Okamoto N, Aramaki M, Tori 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.

273. Bouchet C, Vulliamar-Barrot S, Gonzales M, Boursi S, Bizec CL, Fallet C, Delezoide AL, Moirot H, Laquerriere A, Encha-Razavi F, Durand G, Seta N. Detection of an Alu insertion in the POMT1 gene from an individual with benign glycerol kinase deficiency. Hum Mutat. 1999;13:217–21.

274. Beauchamp NJ, Makris M, Preston FE, Peake IR, Daly ME, Stoppa Lyonnet D, Carter PE, Meo T, Tosi M. Clusters of intragenic Alu insertions in the human pituitary gland. Proc Natl Acad Sci. 2002;99:10856–9.

275. Mustajoki S, Ahola H, Mustajoki P, Kauppinen R. Insertion of Alu element at the site of a silent deletion in a patient with truncal arthrogryposis. Hum Mutat. 2002;19:230–4.

276. Stoppa Lyonnet D, Carter PE, Meo T, Tosi M. Clusters of intragenic Alu insertions in the human pituitary gland. Proc Natl Acad Sci. 2002;99:10856–9.

277. Higgs DR, Goossens M, Kranenburg O, van Ommen GBJ, Vuillaume-Barrot S, Gonzales M, Boukari S, Bizec CL, Fallet C, Delezoide AL, Moirot H, Laquerriere A, Encha-Razavi F, Durand G, Seta N. Detection of an Alu insertion in the POMT1 gene from an individual with benign glycerol kinase deficiency. Hum Mutat. 1999;13:217–21.

278. Tighe PJ, Stevens SE, Dempsey S, Le Deist F, Rieux-Laucat F, Edgar JD. 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.

279. 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.

280. Takimoto M, Kemble MR, Browning SG. Characterization of five Portuguese patients with pyrimidine 5'-nucleotidase deficient hemolytic anemia showing three new PN5 mutations. Blood Cell Mol Dis. 2006;9:266–7.

281. Chen J-M, Masson E, Macek Jr M, Ragatelli O, Piskavcova T, Fercot B, Fila L, Cooper DD, Audrézet M-F, Pérez C. Detection of two Alu insertions in the CFTR gene. J Cyst Fibros. 2007;6:37–43.

282. Abdelhak S, Kalatzis V, Heilig R, Compan S, Samson D, Vincent C, Lecci Acobas F, Cruaud C, Le Merier M, Mathieu M, König R, Vigneron J, Weissenbach J, Petric C, Well D. Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of ZFY1. Hum Mol Genet. 1997;6:2247–55.

283. Okubo M, Hirohi taka A, Saito M, Ebara T, Endo Y, Kuku 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.

284. Udagawa T, Okamoto N, Aramaki M, Tori 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.
of two unusual truncating PMM2 mutations in two CDG-la patients. Mol Genet Metab. 2007;90:408–13.

296. Wimmer K, Callens T, Werstendt A, Messiaen L. The NFI 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, Hedvikovaova M, 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 Biophys Res Co. 2006;347:145–9.

300. Yoshioka N, Nakamura A, Yazaki M, Ieda S, 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, Malveka RG, Yagi M, Okizuka Y, Takeshina 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, Fering AD, Hancks DC, Takeshina 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, Dorning J, Feil S, Hinznau B, van Duijnhoven G, Kirschner R, Hemmerger M, Bergen AA, Rosenberg T, Pinckers AJ, Funedele 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. Martinez-Baray J, Ballestra MU, Oltra S, Orellana C, Palomoque A, Moltó MD, Prieto F, Martinez F. Intronic L1 insertion and F268S, novel mutations in the gene for X-linked retinitis pigmentosa 2. Ann Genet. 2003;46:491–6.

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-Iida Y, Kobayashi K, Watanabe M, Sasaki J, Kumaigai T, Koida H, Saito K, Osawa M, Nakamura Y, Tod T. Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FMD). Hum Mol Genet. 1999;8:2303–9.

308. Bernard V, Minnereor M, Buirk K, Kreuz F, Gillessen-Kaesbach G, Zühlke C. An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. Brit J Dermatol. 2011;164:1390–2.

309. Divoky V, Indrak K, Mrug M, Brabec V, Huisman T, Prchal J, Divoky V, Indrak K, Mrug M, Brabec V, Huisman T, Prchal J. A novel mechanism of β thalassemia: the insertion of L1 retrotransposable element into β globin V5 IS II. Blood. 1996;88 Suppl 1:148a. abstract.

310. Miné M, Chen J-M, Brivet M, Desguerre I, Marchant D, de Lonlay P, Bernard V, Minnerop M, Bürk K, Kreuz F, Gillessen-Kaesbach G, Zühlke C. A novel mechanism of β thalassemia: the insertion of L1 retrotransposable element into β globin V5 IS II. Blood. 1996;88 Suppl 1:148a. abstract.

311. Awano H, Malveka RG, Yagi M, Okizuka Y, Takeshina Y, Matsuo M. Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA, J Hum Genet. 2010;55:785–90.

312. Solyom S, Fering AD, Hancks DC, Takeshina Y, Awano H, Matsuo M, Kazazian HH. Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. Hum Mutat. 2012;33:369–71.

313. Wang T, Reker J, Gueta Z, Sagi M, Kadouri L, Peretz T, Abellovich 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;31:191–5.