NIH Public Access
Author Manuscript
Curr Opin Genet Dev. Author manuscript; available in PMC 2013 June 01.

Published in final edited form as:
Curr Opin Genet Dev. 2012 June ; 22(3): 191–203. doi:10.1016/j.gde.2012.02.006.

Active Human Retrotransposons: Variation and Disease

Dustin C. Hancks and **Haig H. Kazazian Jr.***
McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine

Abstract

Mobile DNAs, also known as transposons or "jumping genes", are widespread in nature and comprise an estimated 45% of the human genome. Transposons are divided into two general classes based on their transposition intermediate (DNA or RNA). Only one subclass, non-LTR retrotransposons, is currently active in humans as indicated by 96 disease-causing insertions. These autonomous Long INterspersed Element-1s (LINE-1s or L1s) are capable of retrotransposing not only a copy of their own RNA but also other RNAs (Alu, SINE-VNTR-Alu (SVA), U6) in *trans* to new genomic locations through an element encoded reverse transcriptase. L1 can also retrotranspose cellular mRNAs, resulting in processed pseudogene formation. Here, we highlight recent reports that update our understanding of human L1 retrotransposition and their role in disease. Finally we discuss studies that provide insights into the past and current activity of these retrotransposons, and shed light on not just when, but where, retrotransposition occurs and its part in genetic variation.

INTRODUCTION

Transposons are present in all eukaryotic genomes sequenced to date, with their abundance and diversity differing across species. There are two different classes of transposons: DNA transposons are DNA sequences which move by a cut-and-paste mechanism via an element-encoded transposase, and retrotransposons, DNA sequences that transpose through an RNA intermediate by a copy-and-paste mechanism.

Retrotransposons are subdivided into those sequences that contain Long Terminal Repeats (LTR) and those that do not (non-LTR). The LTR subclass retrotransposes by a mechanism reminiscent to that used by retroviruses. For brevity, this review focuses on the non-LTR subclass, the only elements thought to be currently active in humans. We refer the reader to excellent reviews on DNA transposons [1] and LTR elements [2–4].

A substantial fraction of the human genome, > 30%, is derived directly or indirectly from LINE-1 retrotransposon activity [5]. Despite the presence of more than 500,000 copies in the human genome, most L1s are inactive due to point mutations, rearrangements, or truncations with only a subset, an estimated 80–100 elements [6], currently active in any individual. Most retrotransposition in the human population is thought to be the consequence of highly active, or "hot", L1 loci that constitute a small minority of total active L1s [6], with many of

© 2012 Elsevier Ltd. All rights reserved.

*Corresponding author: Kazazian@jhmi.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Useful links: http://www.repeatmasker.org/, Repbase (http://www.girinst.org/)[109], http://dbrip.brocku.ca/ [110]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

these being population-specific elements [7,8] or unique to a particular individual, also known as private.

A full-length human L1 is ~6.0 kb in length (Figure 1A), contains an internal promoter located in the 5'-untranslated region (UTR) and two non-overlapping open-reading frames (ORF1 and ORF2), separated by a short inter-ORF spacer (Reviewed in [9]). Both ORFs are required for L1 retrotransposition (Reviewed in [9]). ORF1 encodes a 40 kDa protein (ORF1p) that contains a coiled-coiled domain (CC), a non-canonical RNA recognition motif (RRM) domain [10] and a basic C-terminal domain (CTD).

ORF1p shares little to no homology to known proteins, leaving its origins and its role in retrotransposition unknown. Mouse ORF1p has demonstrated nucleic acid chaperone activity *in vitro* and forms a trimer resembling that of an asymmetric dumbbell [11] which has been confirmed by solution of the crystal structure of the human L1 ORF1p trimer [12]. Now questions, such as how L1 RNA interacts with the ORF1p trimer along with how ORF1p interacts with itself and perhaps other cellular proteins can be modeled. Surprisingly, it appears that L1 RNA may wrap around the ORF1p trimer, rather than be coated by ORF1 protein [12].

ORF2 encodes a 150 kDa protein (ORF2p) that contains endonuclease (EN) [13] and reverse transcriptase (RT) activity [14]. ORF2 has two other conserved domains, the Z domain and cysteine-rich motif (C-domain), of unknown function. The Z domain is just proximal to the RT domain and mutations in this domain can abolish reverse transcriptase activity *in vitro* [15]. Mutations in the C-domain abolish engineered L1 retrotransposition [16] and decrease ORF2p in L1 RNPs [17].

L1s contain their own polyA signal (AATAAA) in a short 3'-UTR, however this polyA signal is commonly read through during transcription generating "chimeric" L1 transcripts that usually contain 3'-flanking sequence along with L1 sequence [16,18,19]. Retrotransposition of these chimeric transcripts will result in 3'-transduction, a process which is estimated to occur for ~15–23% of all genomic L1s [20–22]. SVA elements also commonly transduce 3'-flanking sequence to new genomic locations with ~10% of SVAs in the human genome containing a 3'-transduction event [23,24]. If the sequence downstream of the L1 or SVA contains coding sequence, retrotransposition may result in exon-shuffling [19,24] as in the case of the acyl-malonyl condensing enzyme 1 (*AMAC1*) gene expansion in hominids [24]. In this example, the two-exon *AMAC1* cDNA was retrotransposed as part of an SVA 3'-transduction to three different genomic locations [24]. Often, depending on the age of the insertion or degree of 5'-truncation, the L1 or SVA sequence may be absent at either the source [25] or insertion locus, respectively.

L1 proteins preferentially retrotranspose the RNA from which they were translated over other L1 RNAs by an order of magnitude [26,27], a phenomenon termed *cis* preference. Although, the propensity exists for L1 to mobilize itself, paradoxically, other RNAs, such as SINEs [28] (Alus) [29,30], SVAs [31,32,40], U6 [33–35] and other cellular transcripts [27,36] are commonly retrotransposed by L1. Alu SINE is the most abundant [5], and currently most active retrotransposon in humans with >1 million copies. Alus evolved from 7SL RNA and are primate-specific SINEs (Reviewed in [37]). Alus are 300bp Pol III RNAs that contain an internal A and B box, an internal A-rich region which separates the left monomer from the right monomer, and end in a polyA tail (Reviewed in [38]) (Figure 1B). These elements are thought to localize to the ribosome, via SRP9/14 binding, and it is here where the Alu RNA is hypothesized to interact with the nascent L1 ORF2 protein [29]. A mutation in the Alu SRP 9/14 binding site in the left monomer severely attenuates

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

retrotransposition while a double mutation in the Alu SRP 9/14 binding sites, one in each monomer, abolishes engineered Alu retrotransposition [39].

SVA elements are ~2.0 kb hominid-specific non-coding RNAs [23] derived from different genomic repeats (Figure 1C) and mobilized by L1 *in trans* [32,40]. There are more than 2,700 SVAs in the human genome reference. Canonical SVAs, starting at the 5'-end, contain a variable number of CCCTCT repeats, an Alu-like domain consisting of two antisense Alu fragments, a GC-rich variable number of tandem repeats (VNTR) domain (unit size 36–42 bp and 49–51 bp [31]), and a SINE-R domain which is derived from the envelope (*env*) gene and right LTR of an extinct HERV-K10 (Reviewed in [41]). SVAs contain a canonical polyA signal (AATAAA) and genomic insertions terminate in a polyA tail. Other RNAs, such as U6 small RNAs [33,34], tend to be retrotransposed by L1 via a template-switching mechanism during reverse transcription [35].

Approximately 8000 processed pseudogenes [36], i.e., retrotransposed cellular mRNAs, have been identified in the human genome reference sequence. Processed pseudogene formation is thought to occur infrequently *in vivo* as no retrotransposed cellular RNA insertion has resulted in human disease and pseudogene formation is rare in cultured cells [26,27]. Despite this, it is possible that polymorphic human processed pseudogenes may contribute significantly to phenotype, as an expressed *FGF4* processed pseudogene insertion causes short-legged chondrodysplasia in specific-dog breeds [42] and insertion of a *Cyclophilin A* cDNA into the *TRIM5* gene of the owl monkey confers HIV resistance in that species [43]. L1 and L1-mediated retrotransposition events (Alu, SVA, U6, processed pseudogenes) share similar structural features: 1) polyA tail of variable length, 2) insertion at a DNA sequence resembling the L1 EN consensus cleavage site (5'-TTTT/AA-3'), 3) 5'-truncation, 4) 5'-inversion, and 5) a target-site duplication flanking the insertion (~4–16 bp in length) (Reviewed in [44]).

MOLECULAR MECHANISM OF RETROTRANSPOSITION

General steps in the non-LTR lifecycle have been defined [9] in the context of L1 and other non-autonomous elements, yet the precise details remain obscure. Presumably concomitant with translation, the nascent L1 proteins form a ribonucleoprotein (RNP) complex with the encoding L1 RNA. This RNP is shuttled back into the nucleus, where reverse-transcription occurs at the site of integration, a process termed target-primed reverse transcription (TPRT) [45]. TPRT has been elegantly characterized at the biochemical level, first using the R2 non-LTR retrotransposon from *Bombyx mori* [45,46] and more recently human L1 [47].

For some time it has been difficult to work with the L1 proteins. Recent tagging of the human L1 proteins has made it possible to visualize not only ORF1p but the ever-elusive ORF2p with L1 RNA [17,48]. Mutation of highly conserved amino acids has shed light on residues that are important for retrotransposition in cell culture, but frequently, the precise mechanism remains unknown. These tagging approaches coupled with RNP purification via sucrose gradients [49,50] are providing unprecedented insight into the presence of L1 proteins in the RNP, whether L1 RNA is still bound, and the cellular localization of the RNP components. Tagged L1 proteins should enable identification of proteins and RNAs interacting with L1, and provide a high-resolution, comprehensive map of L1 protein/RNA binding sites.

INSERTIONS AND DISEASE

At the time of submission, we were able to find 96 (Table 2, 25 L1, 60 Alu, 7 SVA, or 4 polyA) retrotransposition events in the literature resulting in single-gene disease (Table 2, Reviewed in [44,51,52]). A new study characterizing mutations in Neurofibromatosis type 1

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

identified 18 retrotransposon insertions (14 Alu, 3 L1, 1 polyA) [53]. Highlighted in this report [53] is that six of the insertions occurred at three-specific insertion sites. Altogether, there are now 6 genomic locations at which two different insertions have independently occurred [53] (Table 2, Reviewed in [51]).

Disease-causing insertions are very useful from a biological standpoint as they can help identify active elements [54] and in the past, they illuminated such phenomena as 3'-transduction [18], and insertion-mediated deletion [55,56], among others. Likewise, preexisting insertions may also be deleterious by mitigating non-allelic homologous recombination (NAHR) [57]. Many reports have described pathogenic, non-pathogenic, and species-specific mobile element NAHR events [58,59].

Of late, disease-causing insertions have revealed an insertion in the *DMD* gene lacking retrotransposon sequence due to severe 5'truncation of the L1 RNA [8]. With the breadth of many non-reference L1s containing 3'-transductions [60] and our previous report of an SVA-mediated orphan 3'-transduction into the *SPTA* gene [31], it is likely many orphan 3'-transductions are misclassified initially or missed altogether. These reports highlight, that the origin of small insertions (~200–1000), perhaps identified through CNV analysis, that contain no retrotransposon sequence, may indeed be L1- or SVA-mediated 3'-transductions with severe 5' truncation during TPRT.

Another example (Table 2, [61]) indicates internal priming by L1 RT [49,62] occurred at an internal L1 RNA A-rich stretch, resulting in a 3'-truncated L1 insertion [63] in the *EYA1* gene [61]. Furthermore, it became clear while annotating the TSDs and L1 EN cleavage sites (consensus bottom strand 5'-YYYY/RR-3', where Y represents pyrimidine and R represents purine) for the disease-causing insertions that there were a subset of 5 that had poor EN cleavage sites (YYYY/YN). Poor endonuclease cleavage sites had been noted previously in experimental systems and *in vivo* [51,64]. Recently, a study modeling L1 retrotransposition at telomeres [65] *in vitro* [66] noted removal of 3'-nucleotides, hereafter referred to as 3'-processing, of the primer in an L1 RNP ORF2p reverse transcriptase assay. This 3'-processing may be due to an exonuclease activity present in the L1 RNPs [66] and may be the same cryptic nuclease activity identified in an *in vitro* TPRT assay associated with L1 ORF2p [47]. In light of these data, the most parsimonious explanation for atypical EN sites for disease-causing insertions (3 Alu, 1 L1, 1 SVA, Table 2) is that these target-sites underwent 3'-processing (1–3 nt) of the bottom strand prior to reverse transcription and that this processing occurred until a stretch of at least 2 Ts were reached on the bottom strand. Finally, we hypothesize that most genomic insertions with atypical L1 EN sites (YYYY/YN) [64] likely represent an actual L1 EN site (YYYY/R) 1–3 nts upstream of the atypical site and 3'-processing of the bottom strand prior to reverse transcription.

Disease-causing insertions also provide insight into mechanisms by which retrotransposons may alter gene expression. In Toda and colleagues' new report [67], the authors demonstrate splicing into the SVA in the 3'UTR of the *fukutin* gene, which is present homozygously in most individuals with Fukuyama muscular dystrophy. This splicing results in an alternative C-terminus of the *fukutin* protein and ultimately mislocalization of the protein. Additional experiments detected SVA alternative splicing in two other examples of genetic disease in cell lines derived from individuals homozygous for a genic SVA insertion. These data along with previous reports [68–70] suggest that homozygosity for an SVA insertion in the sense orientation of a gene will result in loss of gene function.

TIMING OF RETROTRANSPOSITION

Previously, most retrotransposition was thought to occur in the germline. Slowly, over the past several years, evidence has mounted from different experimental systems and from

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analyses of disease-causing insertions, suggesting that most retrotransposition may be occurring in the soma, and in particular early development. The first report of a somatic L1 insertion was located in the *APC* gene in an individual with colon cancer [71]. This insertion was present in the cancer and not in the normal colonic tissue. More recently, analyses of a family with a *de novo* L1 insertion into the *CHM* gene [72] revealed that the mother of the patient was a germline and somatic mosaic for the L1 insertion causing the disease. As disease-causing insertions are overall rare, a significant source of our understanding of L1 retrotransposition, *in vivo*, comes from mice carrying L1 transgenes [73–79]. Retrotransposition from engineered native human, mouse, or synthetic mouse L1s in mice and rats results in frequent somatic retrotransposition [76,77]. Unexpectedly, most insertions from L1 transgenes in mice and rats are somatic and not inherited [76].

Somatic retrotransposition may play a role in tumorigenesis. Nine tumor-specific L1 insertions were identified in 20 lung cancer samples. The six samples with somatic L1 insertions (1–3 new insertions per tumor) also displayed a hypomethylated DNA profile relative to the matched normal tissue [80]. While in the same study, no somatic L1 or Alu insertions were identified in brain tumors. Whether new retrotransposition events are driver or passenger mutations, i.e., whether insertions are causative of cancer or merely occur due to an unstable epigenetic state, is likely being addressed. Further evidence for somatic L1 activity comes from cell culture retrotransposition assays. Engineered L1s readily retrotranspose in human embryonic stem cells [81], induced pluripotent stem cells [82], and different teratocarcinoma cell lines [83]. Likewise, engineered Alu retrotransposition has been documented in human embryonic stem cells [130].

Some of the most intriguing reports, regarding somatic insertions, are those that describe engineered L1 retrotransposition in rat and human neural progenitor cells (NPCs) along with the hippocampus. Using qPCR, it was estimated that there may be ~80 new L1 insertions/NPC in the human brain and that individuals with Rett syndrome, a genetic disorder caused by mutations in Mecp2, have an increased L1 sequence load [88]. Even more surprising is the >20,000 potential somatic L1, Alu, and SVA insertions in human hippocampus and caudate nucleus identified by Baille *et al.* [84] using retrotransposon array capture followed by high-throughput sequencing. This number, >7,000 L1, >13,000, Alu, >1250 SVA insertions is five times the number of non-reference L1s and two times the number of non-reference Alu insertions identified in targeted re-sequencing studies [85–87]. This new work not only corroborates previous reports about L1 retrotransposition in the brain [79,88,89], but is the first report to document Alu and SVA retrotransposition in the human brain. The high validation rate of a subset of somatic L1 (14/14) and Alu (12/15) insertions in the tissue in which the insertions were identified suggests that many are true insertions [84].

The majority of the somatic insertions identified in this study belong to the known active subfamilies (L1-Ta, AluYa5, AluYb8, and other AluYs) with a subset belonging to the older AluS and L1PA2 subfamilies. AluS and PA2 elements are thought to be primarily inactive, because only one disease-causing insertion is associated between the two of them and most PA2s do not contain intact ORFs [90]. Data from cell culture has demonstrated that AluS [39] and L1-Ta elements lacking intact ORFs [26,27] are *trans*-mobilized by active L1s at a modest level and thus could be retrotransposition-competent sequences *in vivo*.

HOST CONTROL

The location and timing of retrotransposition dictates how it is controlled. Similar to other genes, L1s are regulated at various steps in gene action [9], including transcription (*SOX2*, *MECP2*, *RUNX3*, piRNA pathway), post-transcription (translational control of ORF2p), and likely post-translation (*APOBEC3* protein family). One of the primary ways to control

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

retrotransposition is to inhibit retrotransposon expression. Inhibition of retrotransposon expression occurs through extensive DNA methylation at retrotransposon promoter sequences (Reviewed in [91]). In mice, germline-specific Argonaute proteins, specifically the PIWI clade (Reviewed in [92]), along with *DNMT3L* [93] have been implicated in establishing DNA methylation patterns at retrotransposon loci via small RNAs. Male mice homozygous null for either *MILI* [94] or *MIWI2* [95,96] display loss of DNA methylation at retrotransposon loci and show an increase in retrotransposon mRNA expression. These male mice are sterile, with testes being reduced in size. Whether the sterility is due to new retrotransposon insertions or another cause is unknown.

Along these lines, SVA elements are known to be extensively methylated in most human tissues [97] and human embryonic stem cells [98]. Strikingly, SVA DNA in human sperm is primarily hypomethylated while in chimpanzee sperm it remains methylated [98]. One potential explanation for this difference is the presence or absence of a species-specific factor that plays a role in the establishment or maintenance of SVA DNA methylation. SVAs contain numerous CpGs, especially in the VNTR domain, and it has been postulated that SVAs may act as species-specific CpG islands [99]. The nature and number of SVA VNTRs likely differ across individuals as tandem repeats are highly mutable. Thus, it is tempting to speculate that SVAs residing in introns of genes, ~1000 in the human genome reference, may alter the expression of genes in a manner that correlates with their VNTR copy number. Less is known about the chromatin state at retrotransposon loci. No obvious histone modification pattern for L1, Alu, or SVA has yet been reported that parallels DNA methylation. It appears that chromatin surrounding new insertions from engineered L1s in some cell types is rapidly deacetylated [83].

ACTIVE RETROTRANSPOSITION GENERATES GENETIC VARIATION

Previous methods to identify non-reference insertions were modified and/or scaled up using high-throughput sequencing [80,100,101] or other genome-wide approaches [7,60,102,103]. These studies identified many previously unidentified L1 and Alu insertions. Surprisingly, two HERV-K polymorphisms were identified [60]. With a large collection of retrotransposon insertions from targeted-retrotransposon resequencing coupled with whole-genome resequencing efforts (i.e., 1000 genomes project [85,86,104,105]) it becomes possible to update previous estimates of germline retrotransposition frequencies (L1 ~1/100, Alu ~1/20, and SVA ~1/916) [100,103,126,127,128] and the number of insertions within the human population (3,000–10,000 L1 insertions with gene frequencies >.05) [100].

The knowledge of many polymorphic retrotransposon insertions will be powerful for several reasons: 1) In population genetics, retrotransposons are very useful because they exhibit no homoplasy, 2) Shared sequence characteristics of recently retrotransposed elements will help generate hypotheses, such as which nucleotides are most important for activity, 3) They provide markers for disease associations or other phenotypes such as gene expression levels. For the last point, unlike single-nucleotide polymorphisms (SNPs), retrotransposons may not just be linked to the causal variant, but may indeed be the causal variant. There are numerous ways by which L1, Alu, and SVA may disrupt gene expression (Reviewed in ([41,106,107]) with the two primary mechanisms being 1) insertional mutagenesis and 2) aberrant splicing (Table 2).

CONCLUSIONS

The current sequencing technology and new tools, such as the luciferase retrotransposition indicator cassette [108], provide an unprecedented opportunity to explore biological questions regarding retrotransposons and retrotransposition. It is not surprising in the age of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

genomics that active and inactive human retrotransposons show up in most areas of human genetics. At present, the impact of somatic retrotransposition is still unclear. Most somatic insertions will be present in only a small subset of cells. Taken together with the limited number of genes thought to be haploinsufficient in the human genome (~700), most insertions are likely to be benign. Rather, these insertions may act in such a way as to slightly perturb gene expression networks within specific neurons. With improving technology and the vast number of human genomes being sequenced, the catalog of polymorphic retrotransposons should increase dramatically.

THREE BIG PRESSING QUESTIONS

1. Where and how often do somatic insertions occur?
2. What is the impact and/or consequence of retrotransposition in the brain?
3. What is the role, if any, of retrotransposons in cancer?

One final question is, Do retrotransposons matter in man? We think so, but time will tell.

Acknowledgments

We would also like to thank Jean Andorf, Todd Scheetz, Ed Stone, Dana Fuchs-Telem, Eli Sprecher, Veronica Bernard, Licinio Manco, and Leticia Ribeiro for providing information regarding disease-causing insertions. We apologize to colleagues whose work was not expanded or highlighted due to space constraints. Work in the Kazazian Lab is funded by grants from the National Institutes of Health awarded to H.H.K.

REFERENCES

1. Feschotte, Cd; Pritham, EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. Ann Rev Genet. 2007; 41:331–368. [PubMed: 18076328]
2. Jern P, Coffin JM. Effects of retroviruses on host genome function. Annu Rev Genet. 2008; 42:709–732. [PubMed: 18694346]
3. Romanish MT, Cohen CJ, Mager DL. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. Semin Cancer Biol. 2010; 20:246–253. [PubMed: 20685251]
4. Mayer J, Meese E. Human endogenous retroviruses in the primate lineage and their influence on host genomes. Cytogenet Genome Res. 2005; 110:448–456. [PubMed: 16093697]
5. Lander E, Linton L, Birren B, Nusbaum C, Zody M, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]
6. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A. 2003; 100:5280–5285. [PubMed: 12682288]
7. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV. LINE-1 Retrotransposition Activity in Human Genomes. Cell. 2010; 141:1159–1170. [PubMed: 20602998] *This study identified and characterized the most active non-reference L1s to date.
8. Solyom S, Ewing AD, Hancks DC, Takeshima Y, Awano H, Matsuo M, Kazazian HH. Pathogenic orphan transduction created by a non-reference LINE-1 retrotransposon. Human Mutation. 2011 Nov 17.
9. Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 Elements in Structural Variation and Disease. Ann Rev Genom Hum G. 2011; 12:187–215. *Excellent review with greater detail and many more references than provided here.
10. Khazina E, Weichenrieder O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. Proc Natl Acad Sci U S A. 2009; 106:731–736. [PubMed: 19139409] *This study identifies a RRM domain in L1 ORF1p and characterizes this domain in great detail providing a molecular basis for ORF1p's RNA binding properties.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

11. Martin SL. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA Biology. 2010; 7:706–711. [PubMed: 21045547]
12. Khazina E, Truffault V, Büttner R, Schmidt S, Coles M, Weichenrieder O. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. Nat Struct Mol Biol. 2011; 18:1006–1014. [PubMed: 21822284] **This study couples solution structure of the human L1 ORF1 trimer with functional *in vitro* binding and retrotransposition assays.
13. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. 1996; 87:905–916. [PubMed: 8945517]
14. Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science. 1991; 254:1808–1810. [PubMed: 1722352]
15. Clements AP, Singer MF. The human LINE-1 reverse transcriptase: Effect of deletions outside the common reverse transcriptase domain. Nucleic Acids Res. 1998; 26:3528–3535. [PubMed: 9671814]
16. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. High frequency retrotransposition in cultured mammalian cells. Cell. 1996; 87:917–927. [PubMed: 8945518]
17. Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N. Characterization of LINE-1 Ribonucleoprotein Particles. PLoS Genet. 2010; 6 e1001150. *The most detailed study to date of L1 RNPs, provides invaluable insight into L1 biology.
18. Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat Genet. 1994; 7:143–148. [PubMed: 7920631]
19. Moran JV, DeBerardinis RJ, Kazazian HH Jr. Exon shuffling by L1 retrotransposition. Science. 1999; 283:1530–1534. [PubMed: 10066175]
20. Goodier JL, Ostertag EM, Kazazian HH Jr. Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet. 2000; 9:653–657. [PubMed: 10699189]
21. Pickeral OK, Makalowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res. 2000; 10:411–415. [PubMed: 10779482]
22. 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.
23. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA Elements: A Hominid-specific Retroposon Family. J Mol Biol. 2005; 354:994–1007. [PubMed: 16288912]
24. Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci USA. 2006; 103:17608–17613. [PubMed: 17101974]
25. Ejima Y, Yang L. Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. Hum Mol Genet. 2003; 12:1321–1328. [PubMed: 12761047]
26. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH Jr, Boeke JD, Moran JV. Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol. 2001; 21:1429–1439. [PubMed: 11158327]
27. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 2000; 24:363–367. [PubMed: 10742098]
28. Kajikawa M, Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence. Cell. 2002; 111:433–444. [PubMed: 12419252]
29. Boeke JD. LINEs and Alus--the polyA connection. Nat Genet. 1997; 16:6–7. [PubMed: 9140383]
30. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 2003; 35:41–48. [PubMed: 12897783]
31. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. Am J Hum Genet. 2003; 73:1444–1451. [PubMed: 14628287]
32. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum Mol Genet. 2011; 20:3386–3400. [PubMed:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

21636526] *This study describes the development and application of an SVA cell culture retrotransposition assay.

33. Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1. Genomics. 2002; 80:402–406. [PubMed: 12376094]
34. Gilbert N, Lutz S, Morrish TA, Moran JV. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol. 2005; 25:7780–7795. [PubMed: 16107723]
35. Garcia-Perez JL, Doucet AJ, Bucheton A, Moran JV, Gilbert N. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. Genome Res. 2007; 17:602–611. [PubMed: 17416749]
36. Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of Years of Evolution Preserved: A Comprehensive Catalog of the Processed Pseudogenes in the Human Genome. Genome Res. 2003; 13:2541–2558. [PubMed: 14656962]
37. Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. Heredity. 2011; 107:487–495. [PubMed: 21673742]
38. Batzer M, Deininger P. Alu repeats and human genomic diversity. Nat Rev Genet. 2002; 3:370–379. [PubMed: 11988762]
39. Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE. Active Alu retrotransposons in the human genome. Genome Res. 2008; 18:1875–1883. [PubMed: 18836035] *Comprehensive functional analysis of Alus in the human genome and demonstrates which nucleotides are important for Alu retrotransposition.
40. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, Löwer J, Strätling WH, Löwer R, Schumann GG. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 2011 *This study describes the development and application of an SVA cell culture retrotransposition assay.
41. Hancks DC, Kazazian HH Jr. SVA retrotransposons: Evolution and genetic instability. Semin Cancer Biol. 2010; 20:234–245. [PubMed: 20416380]
42. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, Spady TC, Elkahloun A, Cargill M, Jones PG, et al. An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science. 2009; 325:995–998. [PubMed: 19608863] *Interesting study which identifies a processed pseudogene insertion that has a large phenotypic effect in certain dog breeds.
43. Sayah DM, Sokolskaja E, Berthoux L, Luban J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature. 2004; 430:569–573. [PubMed: 15243629]
44. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. Annu Rev Genet. 2001; 35:501–538. [PubMed: 11700292]
45. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993; 72:595–605. [PubMed: 7679954]
46. Christensen SM, Eickbush TH. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol Cell Biol. 2005; 25:6617–6628. [PubMed: 16024797]
47. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. EMBO J. 2002; 21:5899–5910. [PubMed: 12411507]
48. Goodier JL, Mandal PK, Zhang L, Kazazian HH Jr. Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. Hum Mol Genet. 2010; 19:1712–1725. [PubMed: 20147320]
49. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol. 2006; 13:655–660. [PubMed: 16783376]
50. Kulpa DA, Moran JV. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet. 2005; 14:3237–3248. [PubMed: 16183655]
51. Chen JM, Stenson PD, Cooper DN, Ferec C. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet. 2005; 117:411–427. [PubMed: 15983781]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

52. Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res. 2008; 18:343–358. [PubMed: 18256243]
53. Wimmer K, Callens T, Wernstedt A, Messiaen L. The *NF1* Gene Contains Hotspots for L1 Endonuclease-Dependent *De Novo* Insertion. PLoS Genet. 2011; 7 e1002371. **This report describes the most retrotransposition events causing disease in one study to date.
54. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. Isolation of an active human transposable element. Science. 1991; 254:1805–1808. [PubMed: 1662412]
55. Mine M, Chen JM, Brivet M, Desguerre I, Marchant D, de Lonlay P, Bernard A, Ferec C, Abitbol M, Ricquier D, et al. A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. Hum Mutat. 2007; 28:137–142. [PubMed: 17152059]
56. Takasu M, Hayashi R, Maruya E, Ota M, Imura K, Kougo K, Kobayashi C, Saji H, Ishikawa Y, Asai T, et al. Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon. Tissue Antigens. 2007; 70:144–150. [PubMed: 17610419]
57. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab. 1999; 67:183–193. [PubMed: 10381326]
58. Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA. L1 recombination-associated deletions generate human genomic variation. Proc Natl Acad Sci U S A. 2008; 105:19366–19371. [PubMed: 19036926]
59. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. Human Genomic Deletions Mediated by Recombination between Alu Elements. Am J Hum Genet. 2006; 79:41–53. [PubMed: 16773564]
60. Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, Kallicki J, Kaul R, Wilson RK, Eichler EE. A Human Genome Structural Variation Sequencing Resource Reveals Insights into Mutational Mechanisms. Cell. 2010; 143:837–847. [PubMed: 21111241] *This study identifies many novel non-reference human L1s containing 3'-transductions.
61. Morisada N, Rendtorff N, Nozu K, Morishita T, Miyakawa T, Matsumoto T, Hisano S, Iijima K, Tranebjaerg L, Shirahata A, et al. Branchio-oto-renal syndrome caused by partial EYA1 deletion due to LINE-1 insertion. Pediatric Nephrology. 2010; 25:1343–1348. [PubMed: 20130917] *The L1 insertion described herein is the first example of L1 internal priming associated with disease.
62. Srikanta D, Sen SK, Conlin EM, Batzer MA. Internal priming: An opportunistic pathway for L1 and Alu retrotransposition in hominins. Gene. 2009; 448:233–241. [PubMed: 19501635]
63. Ovchinnikov I, Troxel AB, Swergold GD. Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. Genome Res. 2001; 11:2050–2058. [PubMed: 11731495]
64. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet. 2002; 31:159–165. [PubMed: 12006980]
65. Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature. 2007; 446:208–212. [PubMed: 17344853]
66. Kopera HC, Moldovan JB, Morrish TA, Garcia-Perez JL, Moran JV. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. Proc Natl Acad Sci USA. 2011 *This study identifies and characterizes a nuclease activity within L1 RNPs that processes the bottom strand prior to reverse transcription.
67. Taniguchi-Ikeda M, Kobayashi K, Kanagawa M, Yu C-c, Mori K, Oda T, Kuga A, Kurahashi H, Akman HO, DiMauro S, et al. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. Nature. 2011; 478:127–131. [PubMed: 21979053] **Excellent study that describes in great detail an SVA disease-causing mechanism. Furthermore, these authors are able to rescue the phenotype in a cultured cell and mouse model.
68. Damert A, Raiz J, Horn AV, Löwer J, Wang H, Xing J, Batzer MA, Löwer R, Schumann GG. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res. 2009; 19:1992–2008. [PubMed: 19652014]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

69. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 2009; 19:1983–1991. [PubMed: 19635844]

70. Bantysh OB, Buzdin AA. Novel Family of Human Transposable Elements Formed Due to Fusion of the First Exon of Gene MAST2 with Retrotransposon SVA. Biochemistry (Moscow). 2009; 74:1393–1399. [PubMed: 19961423]

71. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer. Cancer Res. 1992; 52:643–645. [PubMed: 1310068]

72. van den Hurk JA, Meij IC, Seleme MC, Kano H, Nikopoulos K, Hoefsloot LH, Sistermans EA, de Wijs IJ, Mukhopadhyay A, Plomp AS, et al. L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet. 2007; 16:1587–1592. [PubMed: 17483097]

73. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH. A mouse model of human L1 retrotransposition. Nat Genet. 2002; 32:655–660. [PubMed: 12415270]

74. Prak ET, Dodson AW, Farkash EA, Kazazian HH Jr. Tracking an embryonic L1 retrotransposition event. Proc Natl Acad Sci U S A. 2003; 100:1832–1837. [PubMed: 12569170]

75. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr. L1 integration in a transgenic mouse model. Genome Res. 2006; 16:240–250. [PubMed: 16365384]

76. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH Jr. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev. 2009; 23:1303–1312. [PubMed: 19487571] *Using L1 transgenic rodent models, this study demonstrated that most retrotransposition events are somatic and not inherited. It also showed L1 RNA carryover from germ cells with insertion in early embryogenesis.

77. An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD. Active retrotransposition by a synthetic L1 element in mice. Proc Natl Acad Sci U S A. 2006; 103:18662–18667. [PubMed: 17124176]

78. An W, Han JS, Schrum CM, Maitra A, Koentgen F, Boeke JD. Conditional activation of a single-copy L1 transgene in mice by Cre. Genesis. 2008; 46:373–383. [PubMed: 18615728]

79. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature. 2005; 435:903–910. [PubMed: 15959507]

80. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE. Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons. Cell. 2010; 141:1253–1261. [PubMed: 20603005] *Identifies many tumor-specific L1 insertions in lung cancer samples displaying a hypomethylated phenotype.

81. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS, Moran JV. LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet. 2007; 16:1569–1577. [PubMed: 17468180]

82. Wissing S, Muñoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-Perez JL, Moran JV, Greene WC. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. Hum Mol Genet. 2011

83. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, et al. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature. 2010; 466:769–773. [PubMed: 20686575] **Elegant and well-controlled study that describes what happens to retrotransposition events after insertion.

84. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan P, Rizzu P, Smith S, Fell M, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011; 479:534–537. [PubMed: 22037309] **Intriguing study that identifies thousands of somatic retrotransposon insertions in human hippocampus samples.

85. Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, Urban AE, Grubert F, Lam HYK, Lee W-P, et al. A Comprehensive Map of Mobile Element Insertion Polymorphisms in Humans. PLoS Genet. 2011; 7 e1002236. *This study develops methods for identification and analysis of non-reference retrotransposon insertions from resequenced genomes.

86. Ewing AD, Kazazian HH Jr. Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res. 2010; 21:985–990. [PubMed: 20980553] *This study

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

develops methods for identification and analysis of non-reference L1 insertions from resequenced genomes.

87. Hormozdiari F, Alkan C, Ventura M, Hajirasouliha I, Malig M, Hach F, Yorukoglu D, Dao P, Bakhshi M, Sahinalp SC, et al. Alu repeat discovery and characterization within human genomes. Genome Res. 2011; 21:840–849. [PubMed: 21131385]
88. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. Nature. 2009; 460:1127–1131. [PubMed: 19657334] *Characterized engineered L1 retrotransposition in human neronal progenitor cells.
89. Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH. L1 retrotransposition in neurons is modulated by MeCP2. Nature. 2010; 468:443–446. [PubMed: 21085180] *Identifies an increased L1 sequence load in the brain of individuals with Rett syndrome.
90. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007; 23:183–191. [PubMed: 17331616]
91. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 1997; 13:335. [PubMed: 9260521]
92. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011; 12:246–258. [PubMed: 21427766]
93. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004; 431:96–99. [PubMed: 15318244]
94. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science. 2007; 316:744–747. [PubMed: 17446352]
95. Carmell MA, Girard A, van de Kant HJ, Bour'his D, Bestor TH, de Rooij DG, Hannon GJ. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev. Cell. 2007; 12:503. [PubMed: 17395546]
96. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 2008; 22:908–917. [PubMed: 18381894]
97. Strichman-Almashanu LZ, Lee RS, Onyango PO, Perlman E, Flam F, Frieman MB, Feinberg AP. A Genome-Wide Screen for Normally Methylated Human CpG Islands That Can Identify Novel Imprinted Genes. Genome Res. 2002; 12:543–554. [PubMed: 11932239]
98. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. Cell. 2011; 146:1029–1041. [PubMed: 21925323]
99. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005; 437:69–87. [PubMed: 16136131]
100. Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 2010; 20:1262–1270. [PubMed: 20488934] *This is the first study to develop a high-throughput sequencing assay to identify non-reference human L1 insertions.
101. Witherspoon DJ, Xing J, Zhang Y, Watkins WS, Batzer MA, Jorde LB. Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. BMC Genomics. 2010; 11:410. [PubMed: 20591181] *This is the first study to develop a high-throughput sequencing assay to identify non-reference human Alu insertions.
102. Huang CR, Schneider AM, Lu Y, Niranjan T, Shen P, Robinson MA, Steranka JP, Valle D, Civin CI, Wang T, et al. Mobile interspersed repeats are major structural variants in the human genome. Cell. 2010; 141:1171–1182. [PubMed: 20602999] *Applied a novel technique involving microarrays to identify non-reference L1s.
103. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, et al. Mobile elements create structural variation: Analysis of a complete human genome.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Genome Res. 2009; 19:1516–1526. [PubMed: 19439515] *First study to compare the mobilomes of two completely sequenced human genomes.

104. The Thousand Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. [PubMed: 20981092]
105. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, et al. Mapping copy number variation by population-scale genome sequencing. Nature. 2011; 470:59–65. [PubMed: 21293372]
106. Goodier JL, Kazazian HH Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell. 2008; 135:23–35. [PubMed: 18854152]
107. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009; 10:691–703. [PubMed: 19763152]
108. Xie Y, Rosser JM, Thompson TL, Boeke JD, An W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. Nucleic Acids Res. 2011; 39:e16. [PubMed: 21071410]
109. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research. 2005; 110:462–467. [PubMed: 16093699]
110. Wang J, Song L, Grover D, Azrak S, Batzer MA, Liang P. dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. Hum Mutat. 2006; 27:323–329. [PubMed: 16511833]
111. Grimaldi G, Skowronski J, Singer MF. Defining the beginning and end of KpnI family segments. EMBO J. 1984; 3:1753. [PubMed: 6090124]
112. Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, Margolet L. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics. 1987; 1:113–125. [PubMed: 3692483]
113. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol. 2001; 21:1973–1985. [PubMed: 11238933]
114. Hohjoh H, Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J. 1996; 15:630–639. [PubMed: 8599946]
115. Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res. 2005; 15:1073–1078. [PubMed: 16024818]
116. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. J Biol Chem. 1994; 269:8466–8476. [PubMed: 8132574]
117. Ono M, Kawakami M, Takezawa T. A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res. 1987; 15:8725–8737. [PubMed: 2825118]
118. Kazazian, HH, Jr. Mobile DNA : finding treasure in junk. Upper Saddle River, New Jersey: FT Press; 2011. p. 216
119. Taşkesen M, Collin GB, Evsikov AV, Güzel Al, Ozqül RK, Marshall JD, Naggert JK. Novel Alu retrotransposon insertion leading to Alström syndrome. Hum Genet. :1–7.
120. Gallus GN, Cardaioli E, Rufa A, Da Pozzo P, Bianchi S, D'Eramo C, Collura M, Tumino M, Pavone L, Federico A. Alu-element insertion in an OPA1 intron sequence associated with autosomal dominant optic atrophy. Mol Vis. 2010; 16:178–183. [PubMed: 20157369]
121. Chen JM, Masson E, Macek M Jr, Raguenes O, Piskackova T, Fercot B, Fila L, Cooper DN, Audrezet MP, Ferec C. Detection of two Alu insertions in the CFTR gene. J Cyst Fibros. 2008; 7:37–43. [PubMed: 17531547]
122. Samuelov L, Fuchs-Telem D, Sarig O, Sprecher E. An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. Br J Dermatol. 2011; 164:1390–1392. [PubMed: 21332462]
123. Akman HO, Davidzon G, Tanji K, MacDermott EJ, Larsen L, Davidson MM, Haller RG, Szczepaniak LS, Lehman TJA, Hirano M, et al. Neutral lipid storage disease with subclinical

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

myopathy due to a retrotransposal insertion in the PNPLA2 gene. Neuromuscular Disorders. 2010; 20:397–402. [PubMed: 20471263]

124. Bernard V, Minnerop M, Burk K, Kreuz F, Gillessen-Kaesbach G, Zuhlke C. Exon deletions and intragenic insertions are not rare in ataxia with oculomotor apraxia 2. Bmc Medical Genetics. 2009; 10

125. Green PM, Bagnall RD, Waseem NH, Giannelli F. Haemophilia A mutations in the UK: results of screening one-third of the population. British Journal of Haematology. 2008; 143:115–128. [PubMed: 18691168]

126. Kazazian HH Jr. An estimated frequency of endogenous insertional mutations in humans. Nat Genet. 1999; 22:130. [PubMed: 10369250]

127. Cordaux R, Hedges DJ, Herke SW, Batzer MA. Estimating the retrotransposition rate of human Alu elements. Gene. 2006; 373:134–137. [PubMed: 16522357]

128. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Careri D, Pinto MT, Kasper CK, Sommer SS. Frequency of recent retrotransposition events in the human factor IX gene. Hum Mutat. 2001; 17:511–519. [PubMed: 11385709]

129. Awano H, Malueka RG, Yagi M, Okizuka Y, Takeshima Y, Matsuo M. Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA. J Hum Genet. 2010; 55:785–790. [PubMed: 20827276]

130. Macia A, Muñoz-Lopez M, Cortes JL, Hastings RK, Morell S, Lucena-Aguilar G, Marchal JA, Badge RM, Garcia-Perez JL. Epigenetic Control of Retrotransposon Expression in Human Embryonic Stem Cells. Molecular and Cellular Biology. 2011; 31:300–316. [PubMed: 21041477]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

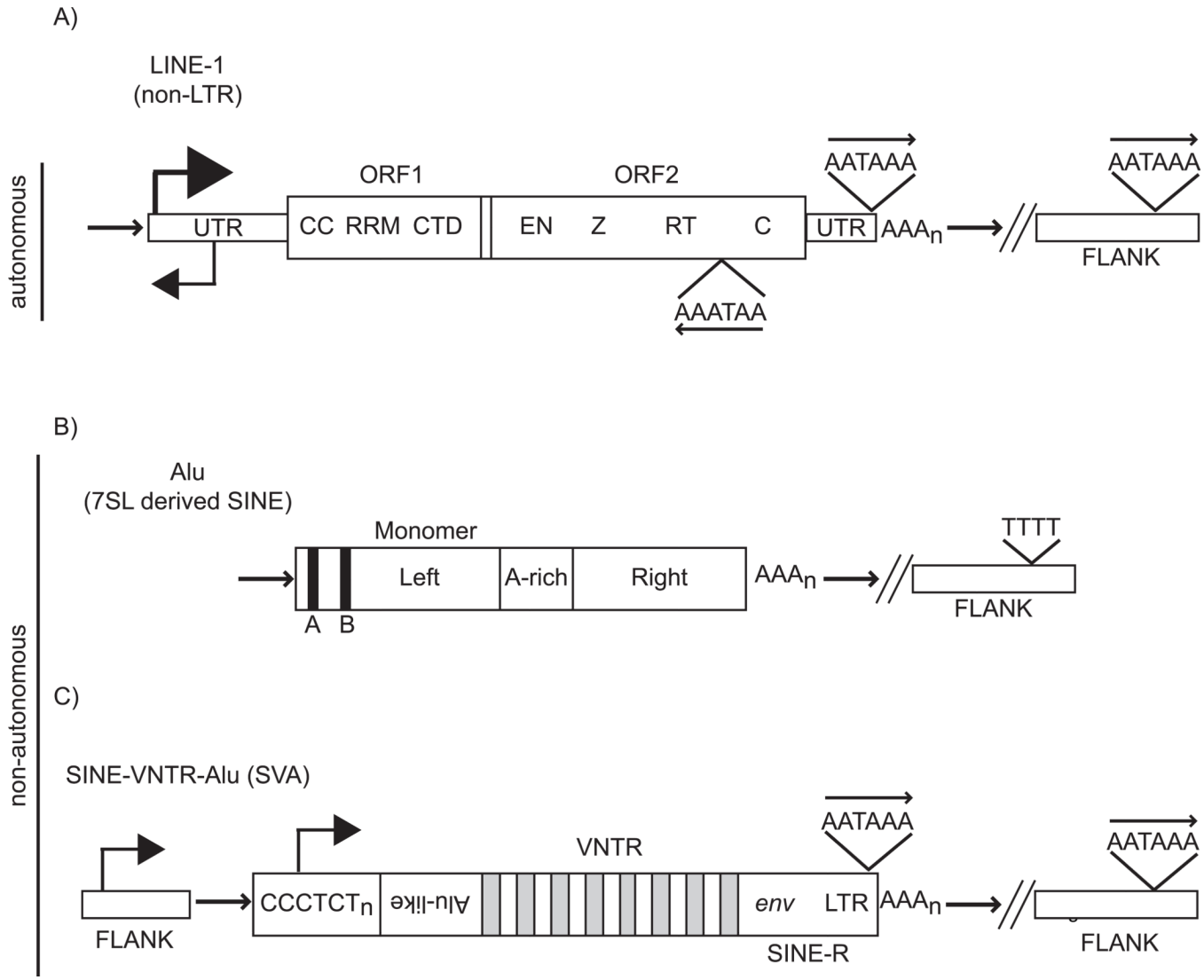

Figure 1.

Active Human Retrotransposons: The three active human retrotransposons are diagrammed with appropriate domains indicated. A) A full-length LINE-1 (L1), 6kb [111,112], in the human genome is shown. L1s have a strong promoter (black bent arrow) located in the 5'-UTR along with a weaker antisense promoter on the bottom strand [113] (smaller black bent arrow). Intact L1s contain two non-overlapping ORFs (ORF1 and ORF2) that encode a 40kDa RNA binding protein (ORF1p) [114] and a 150 kDa protein (ORF2p) with demonstrated DNA endonuclease (EN) [13] and reverse transcriptase (RT) [14] activity. L1 RNAs commonly terminate via a canonical polyA signal (AATAAA) in the 3'-UTR, but do frequently bypass this transcription termination signal for a downstream polyA signal in the 3'-flanking DNA [16,18,20,21]. On the antisense strand, human L1s have a strong polyA signal that in conjunction with the antisense promoter may result in gene-breaking if the insertion is in opposite orientation of the transcriptional unit [115]. L1 genomic insertions terminate in a polyA tail (AAAn) of varying length and are flanked by a TSD (4–16bp in length, black horizontal arrow). UTR = untranslated region, CC = coiled-coiled, RRM = RNA recognition motif, CTD = C-terminal domain, EN = endonuclease, C = cysteine-rich domain, TSD = target-site duplication. B) A full-length Alu, ~300 bp, containing an internal

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

RNA Pol III promoter (A and B box, black boxes) at its 5'-end is shown [38]. Alus were generated from a dimerization event of two 7SL RNA sequences (Left and Right monomer). Alu genomic insertions terminate in a polyA tail and, similar to L1, are flanked by a TSD (black horizontal arrow). Alu transcripts terminate at RNA Pol III terminator sequences (TTTT) located in the downstream flanking sequence. C) A full-length canonical SINE-VNTR-Alu (SVA) consisting of in order from the 5'-end 1) CCCTCT repeat of varying length, 2) sequence sharing homology to two antisense Alu fragments (Alu-like), 3) variable number of GC-rich tandem repeats, unit size 36–42 bp and 49–51bp, and 4) a partial envelope (*env*) and right LTR sequence derived from an extinct HERK-K10 (SINE-R) [31,97,116,117]. SVAs are RNA PolII transcripts, however whether SVAs encode their own promoter is unknown. Transcription of SVA RNAs may occur upstream (black bent arrow) of a genomic SVA or may be initiated throughout the SVA (black bent arrow) [68,69]. Similar to L1, SVA RNAs terminate at a polyA signal (AATAAA) located at the 3'-end of the SINE-R, but may also bypass this signal for a downstream polyA signal [31, 23, 24]. Likewise, SVAs genomic insertions also terminate in a polyA tail (AAAn) and are flanked by a TSD. Note elements are not drawn to scale.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Summary of Human non-LTR Retrotransposon Biology

Element	Copies in human genome (%)*
L1	516,000 (17)
Alu	1100,000 (11)
SVA	2,700 (0.2)
	Known Polymorphic Elements
L1	~1,500[a]
Alu	~6,700[a]
SVA	~150[a]
Total	8,350
	Potentially Active Elements Per Haploid Genome
L1	~40–50[b]
Alu	~852[c]
SVA	~20–50[d]
	Disease-Causing Insertions[e]
L1	25
Alu	60
SVA	7
polyA	4
Total	96
	Somatic Insertions[f]
L1	~8,000
Alu	~14,000
SVA	~1300
Total	>23,000
	Retrotransposition Rates (Live births)[g]
L1	~1/100
Alu	~1/20
SVA	~1/916[h]
Estimated number of retrotransposition events causing single-gene disease	500/yr[i]

*[107]

[a][85–87]

[b][6]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[c] This number represents Alu elements that are consensus in the human genome reference [39].

[d] This estimate is based upon the seven disease-causing SVA insertions which likely come from different source loci with at least one belonging to a multi-member transduction group with more than one full-length element.

[e] Table 2

[f] [84]

[g] [100,102,103, 126–128],

[h] [103]

[i] [118]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Curr Opin Genet Dev. Author manuscript; available in PMC 2013 June 01.

Table 2

Retrotranspositions causing single-gene disease in humans

Insertion	Gene	CHR	Reference	Disease	Subfamily	Size	polyA tail length	Truncation	Transduction (bp)	Strand	Exon/Intron/Mechanism	Target-site duplication (TSD)	L1 endo site (5'-TTTT/AA-3')	Note
Alu	*ABCD1*	X	Kutsche et al. 2002	ALD	AluYb9	98	20	Y/5'TR	N	S	4.7 kb Deletion	No TSD	ATTT/GT	
Alu	*ATP7A*	X	Gu et al. 2007	Menkes Disease	AluYa5a2	282	89	N	N	AS	E	AAAAAGGACAGC	TTTT/AT	
Alu	*BTK*	X	Lester et al. 1997	XLA	AluY	N/A	N/A	N/A	N	AS	E	N/A	N/A	
Alu	*BTK*	X	Conley et al. 2005	XLA	AluY	281	74	N	N	S	E	AGAAATGTATGAGTAAGT	TTCT/AT	Same insertion site Conley et al. SVA
Alu	*CD40LG*	X	Apoil et al. 2007	HIGM	AluYb8	292	8	N	N	AS	E	AAAAATTTTC	TTTT/AT	
Alu	*CLCN5*	X	Claverie-Martin et al. 2003	Dent's Disease	AluYa5	281	50	N	N	S	E	AGAAAATGCTCGAAAGA	TTCT/AT	
Alu	*FVIII*	X	Sukarova et al. 2001	Hemophilia A	AluYb8	290	47	N	N	AS	3 nt Deletion	No TSD	TTTC/AT	
Alu	*FVIII*	X	Ganguly et al. 2003	Hemophilia A	AluYb9	288	37	N	N	AS	I/Splicing	AAAAACCAACAGG	TTTT/AT	Consensus Yb9
Alu	*FVIII*	X	Green et al. 2008 [125]	Hemophilia A	AluYb8	FL	N/A	N	N	AS	E	N/A		
Alu	*FIX*	X	Vidaud et al. 1993	Hemophilia B	AluYa5a2	244	78	Y/5'TR	N	S	E	AAGAATGGCAGATGCGA	TCTT/AA	Same insertion site as Wulff et al. Alu
Alu	*FIX*	X	Wulff et al. 2000	Hemophilia B	AluYa5a2	237	39	Y/5'TR	N	S	E	AAGAATGGCAGATGC	TCTT/AA	Same insertion site as Vidaud et al. Alu
Alu	*FIX*	X	Li et al. 2001	Hemophilia B	AluY	279	40	Y/5'TR	N	AS	E	AAGAAACTGGTCCC	TCTT/AA	
Alu	*GK*	X	Zhang et al. 2000	GKD	AluYc1	241	74	Y/5'TR	N	AS	I	AAAAAATAAG	TTTT/AA	
Alu	*IL2RG*	X	Lester et al. 1997	XSCID	AluYa5	N/A	N/A	N/A	N	AS	I	N/A	N/A	
Alu	*CRB1*	1	den Hollander et al. 1999	RP	AluY	244	70	Y/5'TR	N	AS	E	AAGAGTAAAGATGA	TCTT/GA	
Alu	*SERPINC1*	1	Beauchamp et al. 2000	Type 1 ATP	Alu	6	40	Y/5'TR	N	AS	1.4 kb Deletion	N/A	TTCT/AT	Shortest Alu insertion
Alu	*ALMS1*	2	Taşkesen et al. 2011 [119]	Alström syndrome	AluYa5	257	76	Y/5'TR	N	S	E	AAAAGCCTAGAGAA	TTTT/AA	
Alu	*MSH2*	2	Kloor et al. 2004	HNPCC	AluJ	85	40	Y/5'TR	N	S	E	N/A	N/A	Contains extra 99 nt 3'-of Alu, may be transduction or recombination
Alu	*ZFHX1B*	2	Ishihara et al. 2004	MWS	AluYa5	281	93	N	N	S	E	AAAATTAAAACA	TTTT/AA	
Alu	*BCHE*	3	Muratani et al. 1991	Cholinesterase deficiency	AluYb9	289	38	N	N	S	E	AAAAATATTTTTTCC	TTTT/AA	
Alu	*CASR*	3	Janicic et al. 1995	FHH and NSHPT	AluYa5	280	93	N	N	AS	E	GAAAGCGTGAGCTGC	TTTC/AA	
Alu	*HESX1*	3	Sobrier et al. 2005	Anterior Pituitary Aplasia	AluYb8	288	30	N	N	S	E	AGAAAATGTCTTTAGA	TTCT/AA	
Alu	*OPA1*	3	Gallus et al. 2010 [120]	ADOA	AluYb8	289	25	N	N	AS	I/Splicing	AAAAATTTTAAAAAGTT	TTTT/AC	
Alu	*MLVI2*	5	Economou-Pachnis and Tsichlis 1985	Associated with leukemia	AluYa5	280	26	N	N	AS	I	GAAAATGT	TTTC/AT	
Alu	*APC*	5	Halling et al. 1999	Hereditary desmoid disease	AluYb8	278	40	Y/5'TR	N	S	E	AAGAATAATG	TCTT/AA	Same insertion site as Miki et al. L1
Alu	*APC*	5	Su et al. 2000	FAP	AluYb9	93	60	Y/5'TR	N	AS	I/ Splicing	No TSD	TTTT/AA	1.6 kb intronic deletion
Alu	*MAK*	6	Tucker et al. 2011, Edwin Stone, personal communication	RP	AluYb8	281	57	N	N	AS	E	AAAGAAAAAA	CTTT/AA	Identified by exome resequence

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Insertion	Gene	CHR	Reference	Disease	Subfamily	Size	polyA tail length	Truncation	Transduction (bp)	Strand	Exon/Intron/Mechanism	Target-site duplication (TSD)	L1 endo site (5'-TTTT/AA-3')	Note
			Manco et al. 2006,											
Alu	*NT5C3*	7	Leticia Ribeiro, personal communication	Chronic hemolytic anemia	Alu Ya5	281	36	N	N	S	E	AAGAATGGCAGATGG	TCTT/AA	
Alu	*CFTR*	7	Chen et al. 2008 [121]	Cystic Fibrosis	AluY	46	57	Y/5'TR	N	AS	E	AAGAATCCCACCTATAAT	TCTT/AA	
Alu	*CFTR*	7	Chen et al. 2008 [121]	Cystic Fibrosis	AluYa5	281	56	N	N	S	E	AATAGAAATGATTTTTGTC	TCTC/AT	3'-Processing of (5'-CTC-3')
Alu	*EYA1*	8	Abdelhak et al. 1997	BOR syndrome	AluYa5	n/a	97,31	N/A	N	AS	E	AAAAAATAAATGTGTG	TTTT/AA	PolyA tail shortening between generations
Alu	*LPL*	8	Okubo et al. 2007	LPL deficiency	AluYb9	150	60	Y/5'TR	N	AS	2.2 kb Deletion	No TSD	TTTT/AA	
Alu	*CHD7*	8	Udaka et al. 2007	CHARGE syndrome	AluYa5/8	75	100	Y/5'TR	N	S	10 kb Deletion	No TSD	ATTT/AA	
Alu	*POMT1*	9	Bouchet et al. 2007	Walker Warburg syndrome	AluYa5	290	53	N	N	AS	E	AAAAAGAGATGTACTG	TTTT/AA	
Alu	*FGFR2*	10	Oldridge et al. 1999	Apert syndrome	AluYa5	283	69	N	N	AS	I/Splicing	AGAAAACAAGGGAAGCA	TTCT/AG	
Alu	*FGFR2*	10	Oldridge et al. 1999	Apert syndrome	AluYb8	288	47	N	N	AS	E	AGAATTACCCGCCAAG	TTCT/AT	
Alu	*FGFR2*	10	Bochukova et al. 2009	Apert syndrome	AluYk13	214	12	Y/5'TR	N	AS	E	AAAAGTTACATTCCG	TTTT/GA	
Alu	*FAS*	10	Tighe et al. 2002	ALPS	AluYa5	281	33	N	N	AS	I	AGAATATTCTAAATGTG	TTCT/AA	
Alu	*SERPING1*	11	Stoppa-Lyonnet et al. 1990	HAE	AluYc1	285	42	N	N	S	I	AAAAATACAAAAATTAG	TTTT/AG	
Alu	*HMBS*	11	Mustajoki et al. 1999	AIP	AluYa5	279	39	N	N	AS	E	AAGAATCTTGTCCC	TCTT/GA	
Alu	*GNPTAB*	12	Tappino et al. 2008	ML II	AluYa5	279	17	N	N	AS	E	AAAAACAACAACTGAG	TTTT/GA	
Alu	*BRCA2*	13	Miki et al. 1996	Breast Cancer	AluYc1	281	62	N	N	S	E	AATCACAGGC	GATT/AT	
Alu	*BRCA2*	13	Teugels et al. 2005	Breast Cancer	AluYa5	285	N/A	N	N	S	E	AAGAATCTGAACAT	TTCT/GC	3' Processing 2 nt (5'-CT-3')
Alu	*PMM2*	16	Schollen et al. 2007	CDG-Ia	AluYb8	263	10	Y/5'TR	N	AS	28 kb Deletion	No TSD	TTTT/AA	
Alu	*BRCA1*	17	Teugels et al. 2005	Breast Cancer	AluS	286	N/A	N	N	S	E	GAAAAAGAATCTGCTTT	TTTC/GA	
Alu	*NF1*	17	Wallace et al. 1991	NF1	AluYa5	282	40	N	N	AS	I/Splicing	AAAAAAAAAAACAT	TTTT/AA	First report of *de novo* Alu insertion
Alu	*NF1*	17	Wimmer et al. 2011 [53]	NF1	AluY	280	N/A	N	N	S	I	AAAAAATTCAG	TTTT/AA	Same insertion site as Wimmer et al.[$]
Alu	*NF1*	17	Wimmer et al. 2011 [53]	NF1	AluY	281	N/A	N/A	N	AS	I	N/A		
Alu	*NF1*	17	Wimmer et al. 2011 [53]	NF1	AluYa5	282	60	N	N	S	E	ATAAATAGCCTGGA	TTAT/AA	