MCB
Journals.ASM.org

The Minimal Active Human SVA Retrotransposon Requires Only the 5′-Hexamer and Alu-Like Domains

Dustin C. Hancks,[a]* Prabhat K. Mandal,[a] Ling E. Cheung,[a] and Haig H. Kazazian, Jr.[a,b]

McKusick-Nathans Institute of Genetic Medicine[a] and Department of Pediatrics,[b] Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

RNA-based duplication mediated by reverse transcriptase (RT), a process termed retrotransposition, is ongoing in humans and is a source of significant inter- and perhaps intraindividual genomic variation. The long interspersed element 1 (LINE-1 or L1) ORF2 protein is the genomic source for RT activity required for mobilization of its own RNA in *cis* and other RNAs, such as SINE/variable-number tandem-repeat (VNTR)/Alu (SVA) elements, in *trans*. SVA elements are ~2-kb hominid-specific noncoding RNAs that have resulted in single-gene disease in humans through insertional mutagenesis or aberrant mRNA splicing. Here, using an SVA retrotransposition cell culture assay in U2OS cells, we investigated SVA domains important in L1-mediated SVA retrotransposition. Partial- and whole-domain deletions revealed that removal of either the Alu-like or SINE-R domain in the context of a full-length SVA has little to no effect, whereas removal of the CT hexamer or the VNTR domain can result in a 75% decrease in activity. Additional experiments demonstrate that the Alu-like fragment alone can retrotranspose at low levels while the addition of the CT hexamer can enhance activity as much as 2-fold compared to that of the full-length SVA. These results suggest that no SVA domain is essential for retrotransposition in U2OS cells and that the 5′ end of SVA (hexamer and Alu-like domain) is sufficient for retrotransposition.

Approximately one-third of the human genome (52) is derived from the direct (*cis*) or indirect (*trans*) reverse transcriptase (RT) activity of the long interspersed element 1 (LINE-1 or L1). A full-length active L1 (6.0 kb) (70) is the only autonomous retrotransposon in humans. It contains two nonoverlapping open reading frames (ORFs) (70), an ~900-bp 5′ untranslated region (UTR), with promoter activity (75), and a short 3′-UTR. ORF1 encodes a 40-kDa RNA binding protein (ORF1p) (40) with chaperone activity *in vitro* (56), whereas ORF2 encodes a 150-kDa protein with demonstrated endonuclease (EN) (25) and RT (57) activities. Both ORF1p and ORF2p are required for retrotransposition (20, 61) of their encoding RNA, a phenomenon termed *cis* preference (20, 23, 61, 82).

Most human L1s are inactive due to 5′ truncations, point mutations, or internal rearrangements; however, a subset, ~80 to 100 in any given individual, remain active (4, 8). These active L1 loci are the source for new L1 insertions and can provide the L1 proteins required to mobilize other RNAs (Alu [7, 19], SVA elements [37, 65, 67], U6 [9, 27, 31], and processed pseudogenes [23, 82]) in *trans*. Retrotransposition is ongoing in the human population (4, 6, 24, 41–43, 46), with almost 100 cases of single-gene disease (5, 11, 39, 83) caused by L1 (47), Alu (80), SVA (50, 65), or poly(A) insertions. L1-mediated insertions may be deleterious by disrupting mRNA expression (32, 76) of a specific gene (~25 examples [39, 78]) or mitigating nonallelic homologous recombination (18, 35, 71). New insertions may also contribute to somatic mosaicism (28, 45, 62, 77), in particular, neuronal diversity (2, 16).

SVA elements are hominid-specific (81) composite nonautonomous retroelements (63, 65, 72, 74, 87). SVAs display the hallmarks of L1-mediated retrotransposition (39, 64, 65) via target-primed reverse transcription (TPRT) (13, 15, 55): (i) 5′ truncation, (ii) 5′ inversions, (iii) insertion at DNA sites resembling the L1 EN consensus site (5′-TTTT/A-3′), (iv) insertion flanked by target site duplications of various lengths (4 to 20 bp), (v) insertion ending in a 3′ poly(A) tail, and (vi) insertions often containing 3′ transductions as a consequence of transcriptional readthrough. Full-length SVA elements (Fig. 1A) vary greatly in size because of repeat variation (17, 81) and the presence or absence of transductions (17, 36, 85), but they primarily consist of four domains, in order from the 5′ end: (i) a CT-rich repeat, with CCCTCT being the most common motif, also referred to as the hexamer (Hex), (ii) a sequence sharing homology to two antisense Alu-like fragments (Alu-like), (iii) a variable number of GC-rich tandem repeats (VNTR), with a unit length of either 36 to 42 or 49 to 51 bp (65), and (iv) an ~490-bp sequence derived from the envelope (*env*) gene and right long terminal repeat (LTR) of an extinct HERV-K10 (SINE-R) (63) that contains a canonical poly(A) signal (AATAAA). SVAs are derived from the ancestral SVA2 element (Fig. 1B) (17, 34, 44). These elements differ from canonical SVAs in that they lack all of the domains except the VNTR and contain a 3′ sequence not found in canonical SVAs (3′-U) (Fig. 1B). The individual canonical SVA domains were most likely acquired through pre-mRNA splicing (38).

Although SVAs lack a defined transcriptional unit (17, 36, 81, 85), firefly luciferase assays indicate that the SVA 5′ end (CT hexamer and Alu-like domain) has some promoter activity (86). However, it remains unclear whether the other individual domains also contain promoter activity. Many different SVA structural variant classes exist in the human genome (17), with some elements belonging to multiple classes. Most SVAs in the human

Received 25 June 2012 Returned for modification 5 August 2012
Accepted 12 September 2012
Published ahead of print 24 September 2012
Address correspondence to Haig H. Kazazian, Jr., kazazian@jhmi.edu.
* Present address: Dustin C. Hancks, Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, USA.
Supplemental material for this article may be found at http://mcb.asm.org/.

doi:10.1128/MCB.00860-12

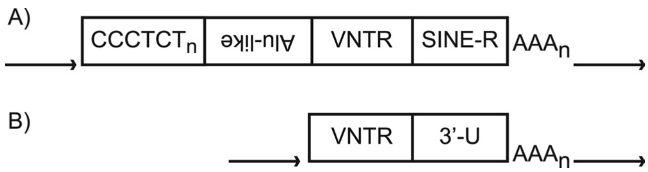


FIG 1 Structures of a canonical SVA element and an ancestral SVA2 element. (A) A full-length SVA is defined by four domains (65, 72, 74, 81) in order from its 5′ end: (i) a CT-rich repeat, with CCCTCT being the predominant repeat unit, (ii) a domain derived from two antisense Alu-like fragments, (iii) a variable number of GC-rich tandem repeats (VNTR) with a unit size of 35 to 50 bp (65), and (iv) sequence derived from the *env* gene and right LTR from an extinct HERV-K10 (SINE-R) (63). (B) SVAs are derived from a sequence, referred to as SVA2 (17, 34, 44), which lacks all of the canonical SVA domains except the VNTR and contains a 3′ sequence not found in canonical SVAs (3′-U). Genomic SVA insertions terminate in a 3′ poly(A) tail and are flanked by target site duplications of various lengths (black horizontal arrows).

genome reference sequence are full length (63%) (17), where full length is defined as containing some portion of the CT hexamer. Some SVAs contain (i) 5′ transductions (3, 17, 36), (ii) 3′ transductions (65, 85), (iii) new 5′ ends acquired via pre-mRNA splicing (3, 17, 36), or (iv) 3′ truncations (17, 81) as a consequence of premature transcriptional termination at noncanonical poly(A) sites in the SINE-R. SVAs are polymorphic in humans (6, 17, 81), and hundreds of insertions are unique to each hominid lineage (12, 54, 60, 79).

Here, we use SVA cell culture retrotransposition assays (37) to define which SVA domains are important for L1-mediated retrotransposition in U2OS cells. U2OS cells are a human osteosarcoma cell line that can support engineered retrotransposition (22, 37) and display the highest levels of SVA retrotransposition activity of the cell lines tested to date. These assays suggest that all domains, to some extent, function in SVA retrotransposition and in certain contexts each domain is dispensable to variable degrees. It appears that removal of the Alu-like or SINE-R domain has minimal effect on retrotransposition activity, whereas removal of the CT hexamer or VNTR severely attenuates retrotransposition. Finally, the SVA 5′ end (a CT hexamer and Alu-like fragment) is sufficient for retrotransposition, and its activity is greater than the sum of its parts, implying synergy between the two domains.

MATERIALS AND METHODS

Plasmids used in this study are listed in Table 1.

Enhanced green fluorescent protein (EGFP) retrotransposition cell culture assays. Approximately 2 × 10^5 U2OS cells were seeded out per well into six-well plates. The following day, the cells in each well were transfected with a total of 2 μg of Maxi/Midi (Qiagen) prepped plasmid DNA using 6 μl of Fugene 6 (Promega) according to the manufacturer's instructions.

In the experiments where one driver plasmid was used, we transfected 1 μg of passenger plasmid DNA and 1 μg of driver plasmid DNA, and in the case of two driver plasmids, we transfected 1 μg passenger plasmid DNA and 0.5 μg of each driver plasmid. One day after transfection, the old medium was replaced with fresh medium. Two days after transfection, fresh medium containing puromycin (Invivogen) at a concentration of 2 μg/ml was added to each well to select for the marked plasmid. Five days after transfection, cells were subject to flow cytometry analysis using a FACSort machine (Becton, Dickinson). The gate for flow cytometry analysis was set on cells transfected with only SVA H2D EGFP (no driver). Data were analyzed with the program CellQuest (Becton, Dickinson) and are presented as the percentage of EGFP-positive cells/puromycin-resistant cells.

Neomycin resistance retrotransposition cell culture assays. Approximately 2 × 10^5 U2OS cells were seeded out per well into six-well plates. The following day, the cells in each well were transfected with a total of 2 μg of Maxi/Midi (Qiagen) prepped plasmid DNA using 6 μl of Fugene 6 (Promega) according to the manufacturer's instructions. In the experiments where one driver plasmid was used, we transfected 1 μg of passenger plasmid DNA and 1 μg of driver plasmid DNA, and in the case of two driver plasmids, we transfected 1 μg passenger plasmid DNA and 0.5 μg of each driver plasmid. One day after transfection the old medium was replaced with fresh medium; 72 h after transfection, fresh medium containing G418 (Invitrogen) at a concentration of 200 μg/ml was added to each well to select for retrotransposition events. Following 12 days of G418 selection, plates were washed, fixed, and stained with Giemsa to visualize colonies.

Characterization of engineered SVA insertions. U2OS G418^R (Neo^r) foci were expanded in cell culture. Genomic DNA was isolated from clonal cell lines using the DNA minikit (Qiagen) according to the manufacturer's instructions. Inverse PCR (iPCR) was carried out to determine the 3′ breakpoint of insertions. First, 3 to 5 μg of genomic DNA was restriction digested at 37°C for at least 2.5 h in a 200-μl volume using SacI (NEB). The restriction enzyme was heat inactivated for 20 min at 65°C. Next, the restricted genomic DNA was ligated overnight at 16°C in a total volume of 500 μl. To isolate potentially ligated DNA fragments, ethanol precipitation was carried out with the precipitated DNA being redissolved in H_2O or Tris-EDTA (TE) in a total volume of 50 μl. A 2-μl sample of the ethanol-precipitated digested DNA was used for each iPCR. Briefly, iPCR was carried out in a 25-μl mixture using GoTaq master mix (Promega) according to the manufacturer's instructions with the primers NeoBridge and Neo3out (see Table S1 in the supplemental material). The PCR from the first round was diluted 40× for a nested round of PCR, and 1 μl of diluted PCR product was used in the nested PCR. The NeoGenoRev and SV40Pro primers (see Table S1) were used in the nested PCR in a total volume of 25 μl using GoTaq. PCRs were analyzed using 1 to 1.5% agarose gel electrophoresis. PCR amplicons of interest were excised, DNA extracted using the gel extraction kit (Qiagen), and subjected to Sanger sequencing. Sanger traces were manually inspected for the restriction site. DNA sequences were aligned to the human genome reference sequence (NCBI GRCh37/hg19) using the BLAST-Like Alignment Tool (BLAT) (48). To determine the 5′ breakpoint, PCR was carried out using primers positioned 5′ of the insertion site and internal SVA reverse primers (see Table S1). The PCR consisted of a 25-μl reaction with ~200 ng of genomic DNA and GoTaq master mix (Promega) and supplemented with betaine at a 1 M final concentration.

RESULTS

SVA retrotransposition in U2OS cells. Here, we investigated SVA biology using two different cell culture assays where a retrotransposition event will result in the generation of either neomycin-resistant foci or EGFP-positive cells. In cell culture, engineered SVA elements, marked with reporter genes (*mneoI* [26, 61] or *mEGFP* [66]) also referred to as retrotransposition indicator cassettes, retrotranspose when cotransfected with plasmids expressing L1 proteins (*trans*-complementation assays). Briefly, the initial configuration of the reporter gene, which is antisense relative to the retrotransposon, remains inactive due to interruption by an intron in the same orientation as the retrotransposon (Fig. 2A). Reporter expression is activated following a round of transcription, splicing, and integration, via target primed-reverse transcription (TPRT), of the element into genomic DNA (Fig. 2B), resulting in either neomycin-resistant foci (Neo^r) (Fig. 2C) or EGFP-positive cells (Fig. 2D). Retrotransposition cell culture assays can be divided into *cis*-complementation (61) and *trans*-complementation (1, 23, 82) assays. The *cis* assay consists of transfecting cells with a plasmid containing an autonomous element, i.e.,

TABLE 1 Plasmids used in this study[a]

Plasmid	Description of plasmid
Driver plasmids	
pcDNA L1-RP (37)	Contains the 6.0-kb FL-L1 RP (49) cloned into pcDNA6 myc-his
pcDNA LRE4	LRE4 was made by replacing the 6.0-kb L1-RP in pcDNA L1-RP with the 6.0-kb LRE4 NotI-AleI fragment from 99 LRE4 *mEGFP* (73)
pcDNA ORF1 (37)	Contains the 5′-UTR and ORF1 coding sequence from L1-RP cloned into pcDNA6 myc-his
pcDNA ORF2 (37)	Contains the 5′-UTR and ORF2 coding sequences from L1.3 (1, 21) cloned into pcDNA6 myc-his
pcDNA ORF2 (RT^-)	Made by swapping a 0.8-kb XbaI-XbaI fragment containing the D702Y mutation (57) from pcDNA L1-RP (D702Y) (37) into pcDNA ORF2
SVA H2D plasmids	
SVA H2D *mneo*I (37)	Contains 4.3-kb KpnI-NotI fragment consisting of SVA H2D and the *mneoI* retrotransposition indicator cassette cloned into pCEP-Pur
SVA H2D *mEGFP* (37)	Contains 4.8-kb KpnI-NotI fragment consisting of SVA H2D and the *mEGFP* retrotransposition indicator cassette cloned into pCEP-Pur
H2DΔAlu-like	SVA H2D contains 0.41-kb deletion from BlpI to PflMI
H2DΔVNTR1	SVA H2D contains 1.13-kb deletion from PflMI to XcmI
H2DΔVNTR2	SVA H2D contains a 1.31-kb deletion from, Tth111l to XcmI
H2DΔVNTR3	SVA H2D contains a 0.76-kb deletion from, XcmI to XcmI
H2DΔSINE-R1	SVA H2D contains a 0.12-kb deletion from PpuMI to AgeI
H2DΔSINE-R2	SVA H2D contains a 0.32-kb deletion from BamHI to AgeI
H2D Hex only	SVA H2D contains 2.01-kb deletion from BlpI to AgeI
H2D Alu-like only	The Alu-like domain was PCR amplified using Phusion (NEB) as a KpnI/AgeI fragment using the following primers positioned from the BlpI to PflMI sites: Alu-likeForKpnI, 5′-TTTTTGGTACCGCTGAGCCAAAGCTGGACTGT-3′; Alu-likeRevAgeI, 5′-TTTTTTACCGGTCCAGACGATGGGCGGCCAGGC-3′
H2D Hex-Alu	SVA H2D contains a 1.60-kb deletion from PflMI to AgeI.
H2DΔHex	Contains a 0.17-kb deletion from KpnI to BlpI. The hexamer was removed by digesting SVA H2D *mEGFP* with KpnI/BlpI followed by ligation with a phosphorylated double-stranded DNA oligonucleotide containing KpnI/BlpI sticky ends (KpnIOligo, 5′-CTTGC-3′; BlpIOligo, 5′-TCAGCAAGGTAC-3′) to restore the KpnI and BlpI sites and make H2DΔHex
SVA H11D plasmids	
pBS H11D	H11D was amplified as 1.7-kb KpnI-NotI PCR product with Phusion (NEB) using H11D_KpnIFor (5′-TTTTTGGTACCAG CAGAAGTGAGAAACCAGGCTCT-3′) and H11D_NotRev (TTTTTGCGGCCGCTTTGGTCTTCAGATGATTGCCAGT-3′) from the bacterial artificial chromosome used for the human genome reference sequence (RP11-465F2; obtained from BACPAC Resources Center [http://bacpac.chori.org/]). This SVA was identified because it differed at only 2 nucleotide positions from the SVA_D Alu-like consensus (81) (99.5% identity) and because it was short in length. Three independent PCRs were combined and sequenced by the Sanger method. SVA H11D differs at two positions in the SINE-R from the human reference genome. Both nucleotide changes are annotated as known SNPs (rs4331123 C → T and rs4554909 T → C). The combined PCR was digested with KpnI and NotI and subcloned into pBluescript KS(−) to make pBS H11D
SVA H11D *mEGFP*	SVA H11D was liberated from pBS SVAH11D as a 1.5-kb KpnI-PpuMI fragment, the EGFP cassette and last 0.1kb of the SINE-R was liberated as a 2.7-kb PpuMI-NotI fragment from pBS SVA2 *mEGFP*, and pCEP-Pur was digested with KpnI-NotI. The three fragments were ligated together to make SVA H11D *mEGFP* Pur
SVA H11D *mneoI*	SVA H11D *mneoI* Pur was cloned similarly to SVA H11D *mEGFP* Pur except using a PpuMI-NotI *mneoI* fragment from pBS SVA2 *mneoI* for a three-way ligation into pCEP-Pur
H11DΔAlu-like	SVA H11D contains a 0.43-kb deletion from NcoI to PflMI
H11DΔVNTR1	SVA H11D contains a 0.56-kb deletion from PflMI to XcmI
H11DΔSINE-R1	SVA H11D contains a 0.12-kb deletion from PpumI to AgeI
H11DΔSINE-R2	SVA H11D contains a 0.32-kb deletion from BamHI to AgeI
H11D Hex-Alu	SVA H11D contains a 1.03-kb deletion from PflMI to AgeI
Other plasmids	
SVA SPTA *mEGFP*	To remove the 0.6-kb flanking DNA cloned with SRE-1, SRE1-*mneoI* was removed from pCEP SREI (37) as a 4.8-kb AleI-NotI fragment and subcloned into pBluescript at EcoRV-NotI sites to make pBS SPTA *mneo*I Δflank. To make SVA SPTA *mEGFP*, SVA SPTA was removed as a 2.5-kb KpnI-PpuMI fragment from pBS SPTA *mneoI* Δflank and swapped into pBS H2D *mEGFP* at KpnI-PpuMI. SVA SPTA now marked with *mEGFP* was removed as a 5.3-kb KpnI-NotI fragment and cloned into pCEP-Pur to make SVA SPTA *mEGFP*
99 RPS *mEGFP* Pur (66)	Contains the full-length L1-RP (49) marked with the *mEGFP* retrotransposition indicator cassette in 99 Pur
99 RPS JM111 *mEGFP* Pur (66)	Contains the full-length L1-RP (49) with amino acid substitutions (R261A/R262A) (61) in ORF1 marked with the *mEGFP* retrotransposition indicator cassette in 99 Pur
ORF1 *mneoI* (80)	Consists of the 5′-UTR and ORF1 coding sequence from L1.3 marked with the *mneoI* retrotransposition indicator cassette cloned into pCEP4 (Invitrogen)

[a] All drivers are cloned into pcDNA6 myc-his (Invitrogen). All elements marked with a retrotransposition indicator cassette are cloned into pCEP4 (Invitrogen) or modified pCEP4 backbones lacking the CMV promoter (99 backbone) or puromycin (pCEP-Pur and 99 Pur) instead of hygromycin resistance. For the SVA deletion constructs, most deletions were made by digesting pBS H2D *mEGFP* or pBS H11D *mEGFP* with both enzymes listed followed by blunting with T4 DNA polymerase (NEB) followed by religation. Each deletion marked with *mEGFP* was then flipped into pCEP-Pur as a KpnI/NotI fragment. The SVAs are cloned as KpnI-AgeI fragments, and the retrotransposition indicator cassettes are AgeI-NotI fragments.

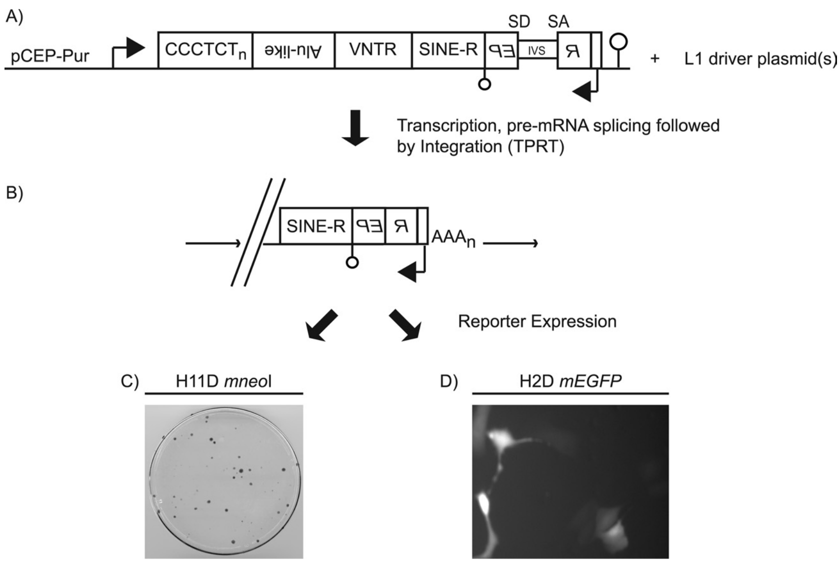


FIG 2 Rationale of SVA *trans*-mobilization assay. (A) An SVA marked with a retrotransposition indicator cassette is shown. Engineered SVAs are cloned into pCEP-Pur, which contains a CMV promoter (bent arrow, top strand) and an SV40 polyadenylation signal (lollipop, top strand). The retrotransposition indicator cassette contains a reporter gene (REP) cloned in opposite orientation relative to the transcriptional orientation of the SVA. The reporter gene contains a promoter (bent arrow, bottom strand) and a poly(A) signal (lollipop, bottom strand); however, due to an intron (IVS) in the same orientation as the SVA the reporter is nonfunctional. Only after a round of transcription with removal of the intron by pre-mRNA splicing and then integration presumably by target-primed reverse transcription (TPRT) (13, 55, 65) will the reporter gene be activated. (B) Cotransfection of an engineered SVA—the passenger—with a plasmid(s) containing active L1 sequence cloned into pcDNA6 will result in an SVA retrotransposition event. New SVA insertions will be of various lengths (diagonal lines) due to occasional 5′ truncations, will be flanked by a target site duplication (black horizontal arrows), lack the intron, and will end in a poly(A) tail. The activated reporter (*mneoI* (26, 61) or EGFP retrotransposition indicator cassette (66) via retrotransposition will result in Neo^r foci (C) or EGFP-positive cells (D). SD, splice donor; SA, splice acceptor.

L1, marked with the retrotransposition indicator cassette. The rationale of the *trans*-complementation assay (Fig. 2) is as follows: a plasmid encoding a DNA sequence being queried for L1-mediated retrotransposition activity (SVA, Alu, and L1 mutant) marked with a retrotransposition indicator cassette (the passenger) is cotransfected along with a L1 plasmid lacking the reporter (the driver) (i.e., full-length L1 [FL-L1] or ORF1 and ORF2 on separate plasmids).

To interrogate retrotransposition permissiveness in U2OS cells, we transfected several marked retrotransposons with the following driver L1 pcDNA6 plasmids: full-length L1 (L1.RP [49]), LRE4 [73]), ORF2 alone, or ORF1 and ORF2 (ORF1/ORF2) on separate plasmids (Fig. 3). For negative controls, passenger plasmids were cotransfected alone (no driver) or with ORF1 and an ORF2 plasmid containing an amino acid substitution (D702Y) that abolishes RT activity (57, 61) [ORF1/ORF2 (RT^-)]. For positive controls, we transfected (i) 99 RPS *mEGFP*, a construct containing L1-RP (*cis* assay) (49, 66), or (ii) 99 RPS JM111 *mEGFP*, a full-length L1, containing two amino acid substitutions in ORF1 known to abolish retrotransposition in *cis* (61). However, 99 RPS JM111 *mEGFP* is retrotransposition competent when *trans* complemented with active L1 proteins (82). To study SVA, we cotransfected in the following SVA constructs: SVA H2D *mEGFP*, SVA H11D *mEGFP*, and SVA SPTA *mEGFP*. SVA H2D has been previously described (37) and is the likely source element for at least 9 human-specific SVA insertions (85). SVA H11D is an SVA that was identified and isolated because the Alu-like domain shared high sequence similarity with the SVA_D Alu-like consensus (81). SVA SPTA is the progenitor to a disease-causing SVA insertion and is modified from the previously described SRE-1 (37, 65) in that it lacks ~600 bp of 5′ flank that was originally isolated with the element.

EGFP retrotransposition cell culture assays were carried out in U2OS cells (see Materials and Methods). Flow cytometry was used to quantify the number of retrotransposition events (EGFP-positive cells) 5 days after transfection (Fig. 3). 99 RPS *EGFP* retrotransposition activity is similar regardless of the driver plasmid with which it was transfected, consistent with *cis* preference and a previous report (82). JM111 activity is detected when transfected with either FL-L1 driver, while JM111 activity is almost background when cotransfected with ORF1 and ORF2 on separate plasmids. The three SVAs (SVA H2D, SVA H11D, and SVA SPTA) retrotranspose at low levels, similar to JM111 retrotransposition, when cotransfected with the FL-L1 drivers, whereas SVA retrotransposition is enhanced when ORF1 and ORF2 are cotransfected on separate plasmids (37). Consistent with a requirement for L1 ORF2 RT activity, *trans* mobilization is almost undetectable when the passenger plasmids (JM111, SVA H2D, SVA H11D, SVA SPTA) are transfected alone (no driver) or with ORF1/ORF2 (RT^-) plasmids. Based on these data, we selected SVA H2D and SVA H11D for further experiments.

SVA insertions in U2OS cells. Next, we carried out SVA neomycin resistance retrotransposition assays (Fig. 4) to obtain Neo^r foci for analysis of engineered SVA genomic insertions. Consistent with the SVA EGFP retrotransposition assays, more retrotransposition is detected when the SVA *mneoI* constructs are cotransfected with ORF1 and ORF2 on separate plasmids (Fig. 4). As a *trans*-mobilization positive control, we cotransfected ORF1 *mneoI* (82) with L1 driver plasmids (Fig. 4). Interestingly, Neo^r

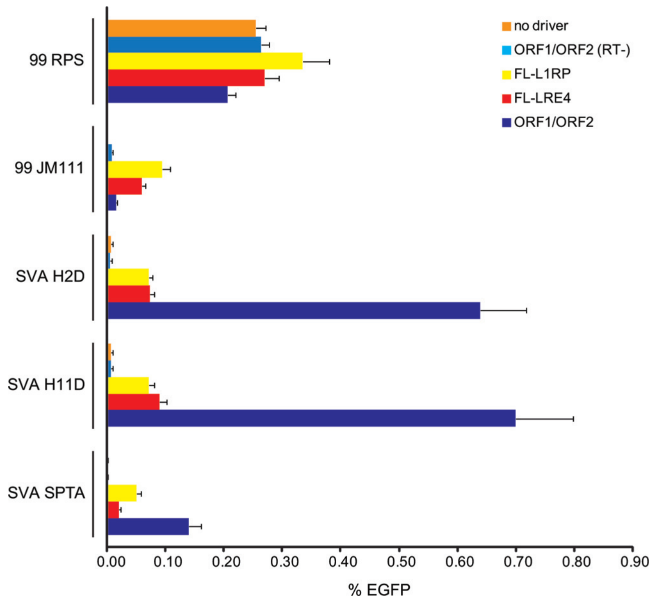


FIG 3 SVA EGFP *trans*-mobilization assays in U2OS cells. Each passenger (vertical axis) was cotransfected either alone (no driver) or with separate ORF1 and ORF2, containing the D702Y missense mutation [ORF1/ORF2 (RT−)], full-length driver L1s (L1-RP or LRE4), or wild-type ORF1 and ORF2 (ORF1/ORF2) on separate plasmids. Flow cytometry was carried out 5 days after transfection. Data are presented as the numbers of EGFP-positive cells per puromycin-resistant cell (% EGFP, horizontal axis) from two separate experiments where each condition was assayed in triplicate for a total of 6 replicates. 99 RPS *EGFP* Pur (66), the full-length L1-RP marked with the *EGFP* cassette cotransfected with each passenger combination, served as a positive control for *cis* mobilization. 99 JM111 *EGFP* Pur (66) served as a positive control for *trans* mobilization. JM111 contains two missense mutations in ORF1 of L1-RP that abolish retrotransposition in *cis* (61); however, this L1 can retrotranspose when complemented with functional L1 proteins supplemented in *trans* (82). Error bars indicate standard errors of the means.

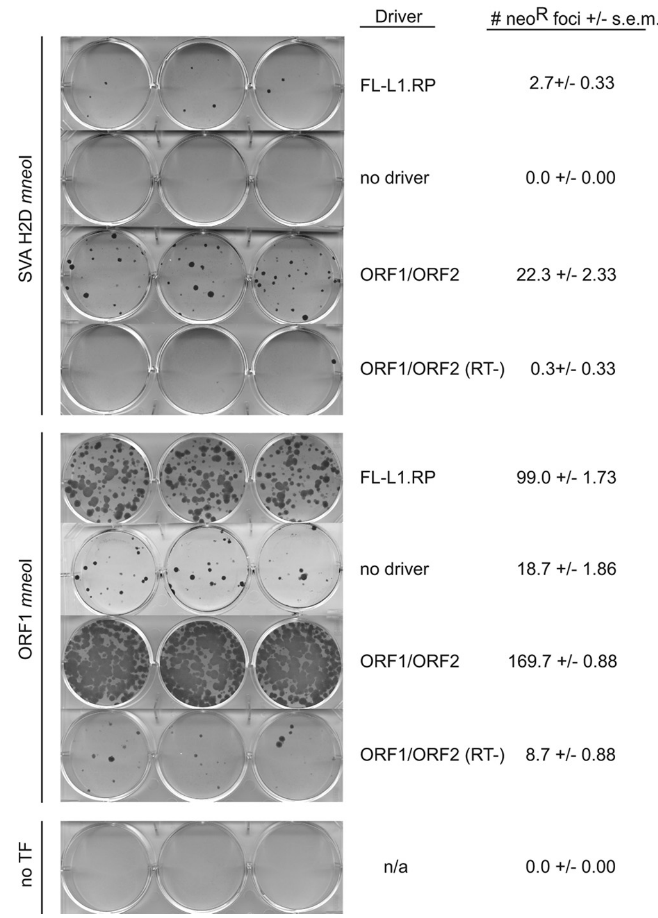


FIG 4 SVA neomycin resistance retrotransposition cell culture assays in U2OS cells. Assays were carried out to establish clonal cell lines for engineered SVA insertion analysis (Table 2). SVA H2D *mneoI* was cotransfected with pcDNA L1-RP (FL-L1.RP) or pcDNA ORF1 and pcDNA ORF2 (ORF1/ORF2) on separate plasmids. ORF1 *mneoI* served as a positive control for *trans* mobilization. The mean number of foci per well ± the standard error of the mean is given. Cotransfection with ORF1/ORF2(RT−) (no driver) and no transfection (no TF) served as negative controls. n/a, not applicable.

foci are observed when ORF1 *mneoI* is cotransfected with ORF1/ORF2 (RT$^-$) or by itself (no driver) (see Discussion).

Using inverse PCR, we recovered the 5′ and 3′ breakpoints for 10 SVA H2D insertions (Table 2), all of which display the hallmarks of L1-mediated retrotransposition. These recovered insertions mimic SVA genomic (65, 74, 81) and recent disease-causing SVA (39) insertions. That is, 6/10 are full length and contain 5′ transductions; ~63% of genomic SVAs are full length, and ~8% contain 5′ transductions (17). The 6 recovered full-length insertions terminate within 5 bp of the cytomegalovirus (CMV) transcriptional start site of the pCEP4 plasmid, and 5/6 contain an untemplated guanosine (G) at the 5′ breakpoint (Table 2). It has been noted that SVA insertions recovered from cell culture terminate at the CMV promoter driving SVA transcription and that full-length insertions, L1 or SVA, may contain 5′ untemplated guanosines (31, 37, 53, 67) presumably due to reverse transcription of the 7mG mRNA cap. Thus, engineered SVA insertions driven by ORF1 and ORF2 on separate plasmids in U2OS cells resemble genomic SVA insertions.

SVA domain deletion analysis. SVAs are characterized by four distinctive domains derived from genomic repeats (72). The role, if any, for each domain in SVA retrotransposition is unclear. SVAs exhibit significant sequence and length variation. This makes it difficult to target specific nucleotides for functional analysis. Therefore, we carried out partial- and whole-domain deletion analysis of two retrotransposition-competent SVAs, SVA H2D and SVA H11D (Fig. 5). To generate domain deletions, we identified restriction enzyme recognition sequences within these SVAs that were close to domain boundaries (Fig. 5A; see Materials and Methods). SVA *mEGFP* constructs, harboring the indicated deletion, were cotransfected into U2OS cells using the ORF1/ORF2 driver combination of ORF1/ORF2 on separate plasmids. Flow cytometry was carried out 5 days later to quantify retrotransposition. The data are presented in Fig. 5 and normalized to FL-length SVA activity (Fig. 5B). Transfection of either SVA H2D or SVA H11D with ORF1/ORF2 (RT$^-$) or alone served as negative controls.

Removal of the CT hexamer (Fig. 5C) resulted in a significant decrease in activity (25% of FL-SVA), while little to no decrease in activity (92 to 123% of FL-SVA) is observed when the Alu-like domain is removed in the context of the full-length SVA (Fig. 5D). Three different deletions were made to explore the role of the VNTR (ΔVNTR1-3) (Fig. 5E to G). The first deletion, ΔVNTR1, removed the SVA VNTR and resulted in a significant decrease in activity (21 to 56% of FL-SVA) for both SVAs (Fig. 5E). The second VNTR deletion, ΔVNTR2, removed the SVA VNTR and a small portion of the Alu-like domain and resulted in a decrease (Fig. 5F) but did not differ from the ΔVNTR1 in activity. The third VNTR deletion (ΔVNTR3) removed a large, 0.8-kb DNA fragment from the VNTR and resulted in an increase in activity (157% of FL-SVA). To investigate the role of the SINE-R domain, a 0.5-kb *env*/right LTR fragment, we made two deletions (ΔSINE-R1 and ΔSINE-R2) (Fig. 5H and I). Removal of the last 0.12 kb of SVA (ΔSINE-R1) does not alter activity (99 to 111% of FL-SVA), whereas removal of the last 0.35 kb results in a 38% decrease in activity in H2D and a slight gain, 23%, in H11D activity.

Removal of the hexamer resulted in a significant activity drop, implicating a role for the hexamer. To examine this further, we tested a hexamer-alone fragment (Fig. 5J). The hexamer alone did not retrotranspose above background (2% FL-SVA). The Alu-like domain has been hypothesized to play a role (59, 65); therefore, we tested an Alu-like fragment alone for retrotransposition activity (Fig. 5K). The Alu-like fragment is retrotransposition competent but is significantly less so than the full-length SVA (39%). To examine whether addition of the hexamer to the Alu-like fragment enhanced retrotransposition, we generated constructs consisting of the hexamer and Alu-like fragment (Hex-Alu) (Fig. 5L). The Hex-Alu fragment is sufficient for retrotransposition and is 1.2 to 1.9× as active as the full-length SVA counterpart.

DISCUSSION

Engineered SVA retrotransposition in U2OS cells. The work presented here indicates that U2OS cells may be useful in exploring SVA element and perhaps L1 retrotransposition. Engineered SVAs retrotranspose at low levels when driven by full-length L1s (Fig. 3). However, delivery of the L1 ORFs on separate plasmids enhances activity greater than 6-fold, providing a better signal-to-noise ratio and enabling SVA functional analysis. Briefly, as a *trans*-mobilization control we transfected ORF1 *mneoI* with a variety of drivers, including ORF1/ORF2 (RT^-) and by itself (no driver). We observed ~10 retrotransposition events per well when ORF1 *mneoI* was cotransfected with ORF1/ORF2 (RT^-) and twice the number of events (~20 events) when it was transfected alone. These events suggest that endogenous ORF2 may be expressed in this cell line and that the ORF2 mutant (RT^-) may compete with endogenous ORF2 for an RNA template. Likewise, the absence of foci observed in untransfected controls and only an occasional event found for the less active SVA, with RT^- or no driver, indicate that these are retrotransposition events. To date, most L1 and L1 *trans*-mobilization assays have been carried out in HeLa cells (1, 31, 61, 69, 82), which exhibit low levels of endogenous L1 activity; this was, in part, one of the original reasons why the assays were carried out in those cells (61). Rare retrotransposition events have been observed in the absence of a driver, and this has been utilized most frequently for Alu assays in HeLa cells (14).

Retrotransposition in the absence of a driver L1 may be useful for future analyses. Of great interest is whether endogenous L1 activity is significantly upregulated in certain tissues (hippocampus) (2, 62) or in diseases like cancer (43, 51). The primary way to detect this L1 activity has been targeted high-throughput sequencing for new insertions (2, 24, 43, 84). It has been noted that *trans*-mobilization assays provide an extremely sensitive means to test for ORF2p activity (1) as ORF2p is very difficult to detect (33). Perhaps ORF1 *mneoI*, which has been useful in studying mecha-

TABLE 2 U2OS SVA H2D *mneoI* recovered insertions[a]

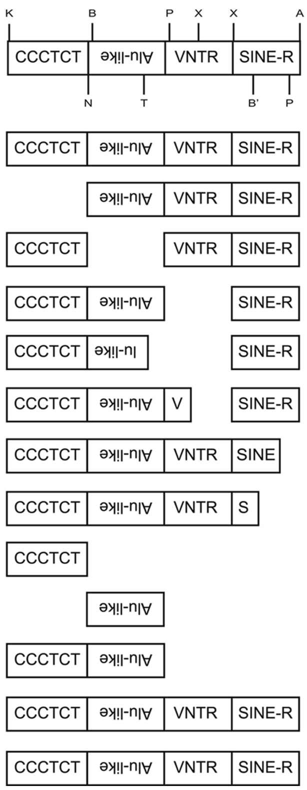
No.	Clone	Breakpoint[b]	Strand	Gene	TSD[c]	EN site	Poly(A) length (nt)	Structure (domain)/size (kb)[d]	Untemplated 5′ nt	TSS
1	1	chr1:115,782,912			AAAAGTGCATGGTT	TTTT/AT	~58	FL/3.6	G	CMV (AGATCT . . .)
2	3	chr14:55,150,391	Antisense	SAMD4A	AAAAAGACTCCTGTGCC	TTTT/AA	~22	FL/3.6	G	CMV (AGATCT . . .)
3	4	chr11:9,115,342	Sense	FLJ46111	AGAAAGGCCTCA	TTCT/GA	~33	5′-TR (VNTR)/2.5		Unknown
4	5	chr16:20,914,555	Sense	LYRM1	GAAAAATCCG	TTTC/AA	~72	5′-TR (VNTR)/3.0		Unknown
5	6	chr20:22,763,381			GAAAAACAAAAGCC	TTTC/AA	~34	FL/3.6	G	CMV (AGATCT . . .)
6	7	chr1:8,360,967			AAACACCATGAGG	GTTT/AA	~46	FL/3.6	G	CMV (AGATCT . . .)
7	10	chr9:132,538,944			AAAGGAGCGTGTGC	CTTT/AC	~23	5′-INV(20 nt), 761-nt deletion (VNTR)/2.8		CMV (TCTCT . . .)
8	11	chr1:48,497,863			AAAAATTTACC	TTTT/AA	~154	5′-TR (VNTR)/2.6	CC	Unknown
9	12	chr12:121,914,401	Antisense	KDM2B	AAAACTAACTGAAACAAA	TTTT/AA	~81	FL/3.6		Unknown
10	13	chr6:37,611,066	Sense	MDGA1	AAAAAAAAA	TTTT/GA	~93	FL/3.6	G	CMV (AGATCT . . .)

[a] SVA neomycin resistance retrotransposition assays were carried out as described in Materials and Methods. SVA H2D was cotransfected with ORF1 and ORF2 on separate plasmids. Insertion breakpoints were recovered using inverse PCR (3′ breakpoint) and PCR using a primer positioned in the 5′ flank and an internal SVA primer (5′ breakpoint). TSD, target site duplication; EN, endonuclease; nt, nucleotide(s); TSS, transcriptional start site; FL, full length; TR, truncated; CMV, cytomegalovirus; VNTR, variable number of tandem repeats.

[b] According to (GRCh37/hg19).

[c] The underlined sequence represents microhomology at the site of inversion presumably used as a primer in the second priming reaction.

[d] Size indicates the entire insertion [SVA +spliced *mneoI* + poly(A) tail]. The size of the spliced *mneoI* cassette to the SV40 poly(A) signal (AATAAA) is 1.3 kb.Beginning with the first nt (+1), the CMV TSS is 5′-TCAGATCTCT. . .-3′.



	H2D			H11D		
	% FL-H2D	% EGFP	n	% FL-H11D	% EGFP	n
B) FL-SVA	100%	0.52 +/- 0.02	6	100%	0.78 +/- 0.02	3
C) Δ Hex	25%	0.13 +/- 0.02	6	n.d.	- +/- -	-
D) Δ Alu-like	123%	0.64 +/- 0.04	6	92%	0.72 +/- 0.03	3
E) Δ VNTR1	21%	0.11 +/- 0.02	6	56%	0.43 +/- 0.03	3
F) Δ VNTR2	27%	0.14 +/- 0.01	6	n.d.	- +/- -	-
G) Δ VNTR3	157%	0.82 +/- 0.04	6	n.d.	- +/- -	-
H) Δ SINE-R1	99%	0.51 +/- 0.07	6	111%	0.87 +/- 0.03	3
I) Δ SINE-R2	62%	0.32 +/- 0.07	6	123%	0.96 +/- 0.03	3
J) Hex only	2%	0.01 +/- 0.01	6	n.d.	- +/- -	-
K) Alu-like only	39%	0.20 +/- 0.04	6	n.d.	- +/- -	-
L) Hex-Alu Only	190%	0.99 +/- 0.03	6	118%	0.92 +/- 0.01	3
M) FL-SVA/RT-	3%	0.01 +/- 0.00	6	2%	0.02 +/- 0.00	3
N) FL-SVA/ no driver	6%	0.03 +/- 0.02	6	2%	0.01 +/- 0.00	3

FIG 5 SVA domain deletion analysis in U2OS cells. (A) A schematic of an SVA element with the relative positions of restriction enzyme recognition sites used for generating deletions is shown. K, KpnI; B, BlpI; N, NcoI; T, Tth111I; P, PflMI; X, XcmI; B′, BamHI; P, PpuMI; A, AgeI. (B to N) FL-SVAs (B) (SVA H2D *mEGFP* and SVA H11D *mEGFP*) or SVAs carrying a deletion (C to L) marked with the *mEGFP* retrotransposition indicator cassette were cotransfected into U2OS cells with ORF1/ORF2 as a driver, and retrotransposition was quantified using flow cytometry 5 days after transfection. Retrotransposition activity was scored as the number of EGFP-positive cells/number of transfected cells (% EGFP) ± standard error of the mean. FL-SVA activity was set to 100%. n, number of replicates. n.d., no data. Cotransfection of each SVA with ORF1/ORF2(RT$^-$) (M) or no driver (N) served as a negative control.

nisms of L1 retrotransposition (1, 27, 82), or other passengers (Alu, SVA) may be useful in detecting endogenous ORF2 activity in these tissues of interest, given that cell lines are available or can be derived.

Role of the individual domains. The SVA deletion analysis revealed that all domains are dispensable to a degree. The dispensability of the SINE-R sequence (ΔSINE-R1 [Fig. 5H], ΔSINE-R2 [Fig. 5I], Hex-Alu-like fragment [Fig. 5L]) in these assays is supported by (i) "natural" SINE-R deletions, i.e., 3′-truncated SVAs (17, 81), and (ii) the gibbon-specific LINE-Alu-VNTR-Alu (LAVA) retrotransposon (10), an "SVA-like element," which is an SVA where the SVA SINE-R domain was replaced with sequence derived from an evolutionarily old Alu (AluSz6) and LINE (L1ME5). We hypothesize that the SINE-R may be inhibitory, as most of the SVA deletions lacking SINE-R sequence (Fig. 5H, I, and L) are more active than their full-length counterparts. If so, this SINE-R inhibition might in part explain the recent amplification of LAVA elements in gibbons. Because the SINE-R is derived from an ancient endogenous retrovirus (HERV-K), this domain might be targeted for silencing at some level by epigenetic modification, perhaps by KRAB-ZNF and KAP1 proteins (58, 68). Although these data do not rule out SINE-R function(s) in SVA transcription, they are consistent with a model where the SINE-R has been maintained because it provides the nearest canonical poly(A) signal.

The role of the VNTR is unclear. However, by definition all SVAs contain some number of tandem repeats. The extreme GC-richness of the tandem repeats indicates that this domain is likely very structured. The low copy number of the ancestral SVA2 elements (17, 34, 44), which also harbor a VNTR but lack the other canonical SVA domains, signifies that this domain at least by itself may have little to do with retrotransposition activity. We observed an increase in retrotransposition activity following a large deletion in the VNTR (Fig. 5G), whereas a large decrease in activity was observed when the entire VNTR was removed in the presence of the other domains (Fig. 5E and F). Preliminary RNA analysis suggests that removal of the entire VNTR results in a decrease in steady-state RNA levels (D. C. Hancks, P. K. Mandal, and H. H. Kazazian, Jr., unpublished data). Therefore, despite younger elements having a larger VNTR domain (81), VNTR size has probably little to do with activity and more to do with the nature of tandem repeats (30), that is, expansion/contraction due to nonallelic homologous recombination.

It has been postulated that the Alu-like domain plays a role in SVA retrotransposition (59, 65) and that perhaps the addition of the hexamer to the Alu-like domain may have changed the properties of the Alu-like domain (81). SVAs lacking the Alu-like domain did not differ from the full-length counterpart in activity, and the Hex-Alu fragment exhibited a 5-fold increase relative to the Alu-like domain alone. Both results are consistent with the notion that the hexamer enhanced or altered the Alu-like domain in some way. Likewise, an SVA deletion lacking the 5′ end (Hex-

Alu-like domain) was recently demonstrated to result in an ~50% reduction in SVA activity (67). Currently, the role of the hexamers in SVA retrotransposition is unclear. No significant rescue of SVA retrotransposition activity has been observed by adding back up to 20 bp of the CT hexamer for H2D or 35 bp for H11D (D. C. Hancks and H. H. Kazazian, Jr., unpublished data). The biology of the hexamer (i.e., length variation, along with variations in purity, and indels) makes it even more difficult to predict its function. Some possibilities are, but are not limited to, the following: (i) some factor exists that binds the pyrimidine-rich hexamers, (ii) the presence of the hexamer positions some factor to interact with other factors important for retrotransposition, or (iii) the hexamers position the Alu-like sequence in such a way as to interact with an unknown factor.

This study furthers our understanding of SVA biology and validates U2OS cells as a useful cell line to study SVA biology. Despite SVAs being relatively active in this cell line, their activity does not correspond with how "hot" SVAs should be based on the ratio of SVA disease-causing insertions to disease-causing insertions due to L1 or Alu (39, 65). We posit that SVAs may be more active in other cell lines. Recently, a paradigm shift has occurred suggesting that most retrotransposition occurs somatically in early development (28, 29, 45, 77). Perhaps a cell line modeling early development, i.e., embryonic stem cells, will be particularly useful in studying SVA elements.

ACKNOWLEDGMENTS

We are very grateful to Kathleen Burns at Johns Hopkins for allowing us to use her FACSort machine. We thank Adam Ewing and Jose Garcia-Perez for critical reading of the manuscript. We also thank David Sigmon and Tara Doucet for technical assistance with this project and the rest of the Kazazian lab for helpful comments. We also thank the staff of the DNA sequencing cores at Johns Hopkins. Finally, we are thankful for the helpful comments and suggestions provided by two anonymous reviewers.

The Kazazian lab is funded by grants from the NIH awarded to H.H.K.

REFERENCES

1. **Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV.** 2006. Unconventional translation of mammalian LINE-1 retrotransposons. Genes Dev. **20**:210–224.
2. **Baillie JK, et al.** 2011. Somatic retrotransposition alters the genetic landscape of the human brain. Nature **479**:534–537.
3. **Bantysh OB, Buzdin AA.** 2009. Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. Biochemistry **74**:1393–1399.
4. **Beck CR, et al.** 2010. LINE-1 retrotransposition activity in human genomes. Cell **141**:1159–1170.
5. **Belancio VP, Hedges DJ, Deininger P.** 2008. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res. **18**:343–358.
6. **Bennett E, Coleman L, Tsui C, Pittard W, Devine S.** 2004. Natural genetic variation caused by transposable elements in humans. Genetics **168**:933–951.
7. **Boeke JD.** 1997. LINEs and Alus—the polyA connection. Nat. Genet. **16**:6–7.
8. **Brouha B, et al.** 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. U. S. A. **100**:5280–5285.
9. **Buzdin A, et al.** 2002. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of L1. Genomics **80**:402–406.
10. **Carbone L, et al.** 2012. Centromere remodeling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons. Genome Biol. Evol. **4**:648–658.
11. **Chen JM, Stenson PD, Cooper DN, Ferec C.** 2005. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum. Genet. **117**:411–427.
12. **Chimpanzee Sequencing and Analysis Consortium.** 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437**:69–87.
13. **Christensen SM, Eickbush TH.** 2005. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol. Cell. Biol. **25**:6617–6628.
14. **Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL.** 2009. Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? Genome Res. **19**:545–555.
15. **Cost GJ, Feng Q, Jacquier A, Boeke JD.** 2002. Human L1 element target-primed reverse transcription in vitro. EMBO J. **21**:5899–5910.
16. **Coufal NG, et al.** 2009. L1 retrotransposition in human neural progenitor cells. Nature **460**:1127–1131.
17. **Damert A, et al.** 2009. 5′-transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res. **19**:1992–2008.
18. **Deininger PL, Batzer MA.** 1999. Alu repeats and human disease. Mol. Genet. Metab. **67**:183–193.
19. **Dewannieux M, Esnault C, Heidmann T.** 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. **35**:41–48.
20. **Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH, Jr.** 1991. Isolation of an active human transposable element. Science **254**: 1805–1808.
21. **Dombroski BA, Scott AF, Kazazian HH, Jr.** 1993. Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. Proc. Natl. Acad. Sci. U. S. A. **90**: 6513–6517.
22. **Doucet AJ J, et al.** 2010. Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet. **6**:e1001150. doi:10.1371/journal.pgen.1001150.
23. **Esnault C, Maestre J, Heidmann T.** 2000. Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. **24**:363–367.
24. **Ewing AD, Kazazian HH, Jr.** 2010. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. **20**:1262–1270.
25. **Feng Q, Moran JV, Kazazian HH, Jr, Boeke JD.** 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87**:905–916.
26. **Freeman JD, Goodchild NL, Mager DL.** 1994. A modified indicator gene for selection of retrotransposition events in mammalian cells. Biotechniques **17**:46, 48–49, 52.
27. **Garcia-Perez JL, Doucet AJ, Bucheton A, Moran JV, Gilbert N.** 2007. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. Genome Res. **17**:602–611.
28. **Garcia-Perez JL, et al.** 2007. LINE-1 retrotransposition in human embryonic stem cells. Hum. Mol. Genet. **16**:1569–1577.
29. **Garcia-Perez JL, et al.** 2010. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature **466**: 769–773.
30. **Gemayel R, Vinces MD, Legendre M, Verstrepen KJ.** 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. **44**:445–477.
31. **Gilbert N, Lutz S, Morrish TA, Moran JV.** 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol. Cell. Biol. **25**:7780–7795.
32. **Goodier JL, Kazazian HH, Jr.** 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell **135**:23–35.
33. **Goodier JL, Ostertag EM, Engleka KA, Seleme MC, Kazazian HH, Jr.** 2004. A potential role for the nucleolus in L1 retrotransposition. Hum. Mol. Genet. **13**:1041–1048.
34. **Han K, et al.** 2007. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science **316**:238–240.
35. **Han K, et al.** 2008. L1 recombination-associated deletions generate human genomic variation. Proc. Natl. Acad. Sci. U. S. A. **105**:19366–19371.
36. **Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH, Jr.** 2009. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. **19**:1983–1991.
37. **Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH. Jr.** 2011 Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum. Mol. Genet. **20**:3386–3400.
38. **Hancks DC, Kazazian HH, Jr.** 2010. SVA retrotransposons: evolution and genetic instability. Semin. Cancer Biol. **20**:234–245.
39. **Hancks DC, Kazazian HH, Jr.** 2012. Active human retrotransposons: variation and disease. Curr. Opin. Genet. Dev. **22**:191–203.

40. **Holmes SE, Singer MF, Swergold GD.** 1992. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J. Biol. Chem. **267**: 19765–19768.
41. **Hormozdiari F, et al.** 2011. Alu repeat discovery and characterization within human genomes. Genome Res. **21**:840–849.
42. **Huang CR, et al.** 2010. Mobile interspersed repeats are major structural variants in the human genome. Cell **141**:1171–1182.
43. **Iskow RC, et al.** 2010. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell **141**:1253–1261.
44. **Jurka J, et al.** 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. **110**:462–467.
45. **Kano H, et al.** 2009. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev. **23**:1303–1312.
46. **Karakoc E, et al.** 2012. Detection of structural variants and indels within exome data. Nat. Methods **9**:176–178.
47. **Kazazian HH, Jr et al.** 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature **332**:164–166.
48. **Kent WJ.** 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. **12**:656–664.
49. **Kimberland ML, et al.** 1999. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum. Mol. Genet. **8**:1557–1560.
50. **Kobayashi K, et al.** 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature **394**:388–392.
51. **Konkel MK, Batzer MA.** 2010. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. Semin. Cancer Biol. **20**:211–221.
52. **Lander E, et al.** 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.
53. **Lavie L, Maldener E, Brouha B, Meese EU, Mayer J.** 2004. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. Genome Res. **14**:2253–2260.
54. **Locke DP, et al.** 2011. Comparative and demographic analysis of orangutan genomes. Nature **469**:529–533.
55. **Luan DD, Korman MH, Jakubczak JL, Eickbush TH.** 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72**:595–605.
56. **Martin SL, Bushman FD.** 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol. Cell. Biol. **21**:467–475.
57. **Mathias SL, Scott AF, Kazazian HH, Jr, Boeke JD, Gabriel A.** 1991. Reverse transcriptase encoded by a human transposable element. Science **254**:1808–1810.
58. **Matsui T, et al.** 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature **464**:927–931.
59. **Mills RE, Bennett EA, Iskow RC, Devine SE.** 2007. Which transposable elements are active in the human genome? Trends Genet. **23**:183–191.
60. **Mills RE, et al.** 2006. Recently mobilized transposons in the human and chimpanzee genomes. Am. J. Hum. Genet. **78**:671–679.
61. **Moran JV, et al.** 1996. High frequency retrotransposition in cultured mammalian cells. Cell **87**:917–927.
62. **Muotri AR, et al.** 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature **435**:903–910.
63. **Ono M, Kawakami M, Takezawa T.** 1987. A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res. **15**:8725–8737.
64. **Ostertag E, Kazazian H, Jr.** 2001. Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. **35**:501–538.
65. **Ostertag EM, Goodier JL, Zhang Y, Kazazian HH, Jr.** 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am. J. Hum. Genet. **73**:1444–1451.
66. **Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV, Kazazian HH, Jr.** 2000. Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res. **28**:1418–1423.
67. **Raiz J, et al.** 2012. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. **40**:1666–1683.
68. **Rowe HM, et al.** 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature **463**:237–240.
69. **Sassaman DM, et al.** 1997. Many human L1 elements are capable of retrotransposition. Nat. Genet. **16**:37–43.
70. **Scott AF, et al.** 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1**:113–125.
71. **Sen SK, et al.** 2006. Human genomic deletions mediated by recombination between Alu elements. Am. J. Hum. Genet. **79**:41–53.
72. **Shen L, et al.** 1994. 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. **269**:8466–8476.
73. **Solyom S, et al.** 2012. Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. Hum. Mutat. **33**:369–371.
74. **Strichman-Almashanu LZ, et al.** 2002. A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. Genome Res. **12**:543–554.
75. **Swergold GD.** 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. **10**:6718–6729.
76. **Taniguchi-Ikeda M, et al.** 2011. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. Nature **478**: 127–131.
77. **van den Hurk JA, et al.** 2007. L1 retrotransposition can occur early in human embryonic development. Hum. Mol. Genet. **16**:1587–1592.
78. **van der Klift HM, Tops CM, Hes FJ, Devilee P, Wijnen JT.** 2012. Insertion of an SVA element, a nonautonomous retrotransposon, in PMS2 intron 7 as a novel cause of Lynch syndrome. Hum. Mutat. **33**: 1051–1055.
79. **Ventura M, et al.** 2011. Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee. Genome Res. **21**:1640–1649.
80. **Wallace MR, et al.** 1991. A de novo Alu insertion results in neurofibromatosis type 1. Nature **353**:864–866.
81. **Wang H, et al.** 2005. SVA elements: a hominid-specific retroposon family. J. Mol. Biol. **354**:994–1007.
82. **Wei W, et al.** 2001. Human L1 retrotransposition: *cis* preference versus *trans* complementation. Mol. Cell. Biol. **21**:1429–1439.
83. **Wimmer K, Callens T, Wernstedt A, Messiaen L.** 2011. The *NF1* gene contains hotspots for L1 endonuclease-dependent *de novo* insertion. PLoS Genet. 7:e1002371. doi:10.1371/journal.pgen.1002371.
84. **Witherspoon D, et al.** 2010. Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. BMC Genomics **11**:410. doi: 10.1186/1471-2164-11-410.
85. **Xing J, et al.** 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc. Natl. Acad. Sci. U. S. A. **103**: 17608–17613.
86. **Zabolotneva AA, et al.** 2012. Transcriptional regulation of human-specific SVAF1 retrotransposons by cis-regulatory MAST2 sequences. Gene **505**:128–136.
87. **Zhu ZB, Hsieh SL, Bentley DR, Campbell RD, Volanakis JE.** 1992. A variable number of tandem repeats locus within the human complement C2 gene is associated with a retroposon derived from a human endogenous retrovirus. J. Exp. Med. **175**:1783–1787.