Exon-trapping mediated by the human retrotransposon SVA

Dustin C. Hancks,[1,3] Adam D. Ewing,[1,3] Jesse E. Chen,[1] Katsushi Tokunaga,[2] and Haig H. Kazazian Jr.[1,4]

[1]*Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA;* [2]*Department of Human Genetics, Graduate School of Medicine, University of Tokyo, 113-0033 Tokyo, Japan*

Although most human retrotransposons are inactive, both inactive and active retrotransposons drive genome evolution and may influence transcription through various mechanisms. In humans, three retrotransposon families are still active, but one of these, SVA, remains mysterious. Here we report the identification of a new subfamily of SVA, which apparently formed after an alternative splicing event where the first exon of the *MAST2* gene spliced into an intronic SVA and subsequently retrotransposed. Additional examples of SVA retrotransposing upstream exons due to splicing into SVA were also identified in other primate genomes. After molecular and computational experiments, we found a number of functional 3′ splice sites within many different transcribed SVAs across the human and chimpanzee genomes. Using a minigene splicing construct containing an SVA, we observed splicing in cell culture, along with SVA exonization events that introduced premature termination codons (PTCs). These data imply that an SVA residing within an intron in the same orientation as the gene may alter normal gene transcription either by gene-trapping or by introducing PTCs through exonization, possibly creating differences within and across species.

[Supplemental material is available online at http://www.genome.org. The 5′ RACE sequence data from this study have been submitted to dbEST (http://www.ncbi.nlm.nih.gov/dbEST) under accession nos. GR564526–GR564716.]

Most eukaryotic genomes harbor retrotransposable elements (Malik et al. 1999). About 35% of the human genome is derived from retrotransposed sequences such as LINE-1, *Alu*, SVA, endogenous retroviruses, and processed pseudogenes (Lander et al. 2001). Although most human retroelement copies are no longer mobile, both active and inactive human elements have been shown to drive genome evolution and influence gene expression (Moran et al. 1999; Han et al. 2004; for reviews, see Belancio et al. 2008; Goodier and Kazazian 2008).

SVA RNAs are hominid-specific noncoding RNAs, which vary in size from 700–4000 bp, and are likely mobilized by the human LINE-1 in *trans* (Ono et al. 1987; Shen et al. 1994; Ostertag et al. 2003; Wang et al. 2005), similar to the human *Alu* (Dewannieux et al. 2003). There are roughly 2700 SVA copies in the human genome (Wang et al. 2005). A canonical full-length SVA (Fig. 1A) contains a number of sequence features proceeding from its 5′ end: (1) a CCCTCT hexameric repeat, ranging in repeat number from a few to as many as 71; (2) a sequence that shares homology with two antisense *Alu* fragments; (3) a variable number of tandem repeat sequence (VNTR); and (4) a sequence derived from the ENV gene and right LTR of an extinct HERV-K, hereafter referred to as SINE-R. SVAs typically terminate at their own polyA signal, with genomic insertions usually containing a number of adenines at the 3′ end. A target site duplication common to other LINE-1-driven retroelements (6–20 bp) flanks the inserted element (Ostertag et al. 2003).

Little is known about the biology of SVA apart from its structure. SVAs are currently active in the human genome, as indicated by the identification of de novo SVA insertions associated with disease (Hassoun et al. 1994; Rohrer et al. 1999; Makino et al. 2007). SVA disease insertions are associated with exon-skipping (Hassoun et al. 1994; Rohrer et al. 1999), deletion of genomic DNA (Takasu et al. 2007), and reduced or absent mRNA expression (Kobayashi et al. 1998; Wilund et al. 2002; Makino et al. 2007). In a manner similar to L1, SVAs have been shown to transduce 3′ flanking sequences to new genomic locations (Moran et al. 1999; Ostertag et al. 2003; Xing et al. 2006). SVAs are thought to be highly active due to the ratio of disease insertions to genomic copies. Furthermore, the high levels of insertion polymorphism of the human-specific subfamilies, E and F (Bennett et al. 2004; Wang et al. 2005), support the notion that SVAs are evolutionarily young and relatively active in the human population.

The mechanism of SVA transcription and the location of its promoter are unknown and are critical to our understanding of SVA retrotransposition. To date, experiments to characterize the SVA promoter have led to ambiguous results. Recently, we reexamined an SVA insertion on CH6 associated with a genomic deletion including the entire *HLA-A* gene that resulted in leukemia in three Japanese individuals (Takasu et al. 2007). The *HLA-A* insertion led to several interesting findings, including the identification of a new SVA subfamily formed by alternative splicing from the first exon of the *MAST2* gene on CH1 into an SVA, and the identification of a SVA master element on CH10. We have also identified numerous functional 3′ splice sites (SSs) within SVA while analyzing human and chimp SVA 5′ transcriptional start sites (TSSs), and three more examples where splicing into the SVA followed by retrotransposition led to exon shuffling. Using a minigene construct containing an intronic SVA, we are able to show that splicing into the SVA is not rare and that this splicing results in both exonization and SVA gene-trapping. These data suggest that splicing into SVA elements enables their expression and can allow for adaptive evolution at the cost of altering the transcriptome in both humans and great apes.

[3]**These authors contributed equally to this work.**
[4]**Corresponding author.**
E-mail kazazian@mail.med.upenn.edu; fax (215) 573-7760.

Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.093153.109.

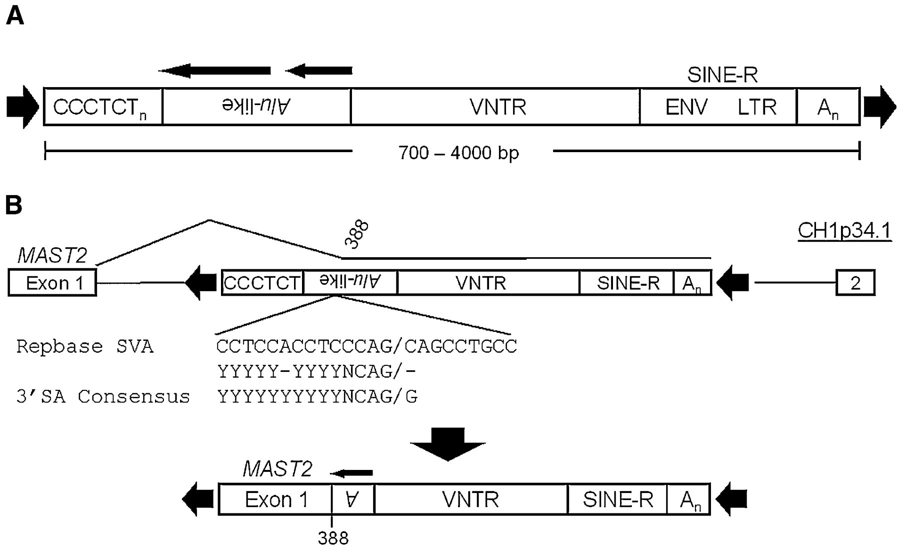

Figure 1. SVA and alternative splicing at the *MAST2* locus. (*A*) The canonical SVA is displayed, consisting of some number of CCCTCT hexameric repeats followed by the *Alu*-like region, a variable number of tandem repeats, and then the SINE-R region, followed by a polyA signal with the entire SVA flanked by target-site duplications (black arrowheads). Two black arrows over the *Alu*-like region indicate sequence homology in the SVA to two ancestral *Alu* elements. (*B*) Shown is an ancestral CH1 *MAST2* locus, sometime after the human–chimp split, containing a full-length SVA within intron 1 (*top*). An alternative splicing event into the *Alu*-like region (black line) likely occurred resulting in the original SVA$_{F1}$ founder insertion (*bottom*) flanked by target-site duplications (black arrowheads) in the human genome. Note the loss of the entire CCCTCT hexamer and most of the *Alu*-like region after the alternative splicing event. The SVA splice site relative to SVA$_{Rep}$ is aligned to the known splice site sequence (*middle*).

Results

Identification of *MAST2*-SVA

In 2007, Takasu et al. described a 14-kb deletion that included the entire *HLA-A* locus in three unrelated families, leading to leukemia in one individual from each of the families. Analysis of the deletion site identified an SVA insertion, hereafter referred to as SVA$_{HLA\text{-}A}$. Using the SVA$_{HLA\text{-}A}$ DNA sequence, we located an SVA insertion on chromosome 3p21.31 as the likely progenitor of the SVA$_{\mathrm{HLA\text{-}A}}$ insertion (Takasu et al. 2007). Further analysis of SVA$_{HLA\text{-}A}$ and its progenitor revealed several interesting details.

First, when using BLAST (Altschul et al. 1990) to align the SVA$_{HLA\text{-}A}$ sequence to the reference genome, many hits were obtained, most being SVAs. However, a few hits had a unique sequence upstream of the SVA. The unique sequence juxtaposed to the SVA$_{HLA\text{-}A}$ query sequence mapped with 99% identity (210/211 nucleotides [nt]) to the 5′ UTR and the first exon of the *MAST2* (M2) gene on chromosome 1. Further analysis showed that 262 nt of the SVA$_{HLA\text{-}A}$ sequence mapped to *MAST2*; however, both the SVA$_{HLA\text{-}A}$ and CH3 insertions had a 40-bp deletion of nucleotides 210–249 relative to the *MAST2* 5′ UTR. Bioinformatics analysis identified 73 SVAs (Supplemental Table 1) containing some fraction of the *MAST2* 5′ UTR and first exon in the human genome. Subsequent analyses concluded that 3′ SVA sequences containing *MAST2*-derived 5′ ends cluster together phylogenetically in a clade, consistent with this subgroup being derived from a founder event. Moreover, sequence analysis grouped the *MAST2*-SVAs with the youngest human-specific SVA subfamily (F), a result consistent with the absence of SVAs containing *MAST2* 5′ transductions in the chimpanzee reference sequence. Hereafter, the *MAST2*-SVA subfamily will be referred to as SVA$_{F1}$.

The number of nucleotides derived from *MAST2* directly upstream of SVA varied from 35–382 (Supplemental Table 1). Given that the *MAST2* 5′ UTR and first exon combined is 460 nt, no SVA$_{F1}$s present in the human reference genome contains the entire 5′ UTR-first exon.

The *MAST2* sequence abutting the genomic copies of SVA in SVA$_{F1}$ elements terminated directly at the 5′ SS of the *MAST2* first exon. However, there is no SVA present in the reference *MAST2* intron sequence. It is likely that an SVA with an allele frequency <1 resided in intron 1 of *MAST2* (Fig. 1B) in the individual in which the first M2-SVA splicing and subsequent retrotransposition event occurred. We generated a consensus sequence from the SVA$_{F1}$ in the human genome and aligned it to the SVA present in Repbase (Jurka 2000; Jurka et al. 2005), henceforth called SVA$_{Rep}$, to determine whether the site where *MAST2* and SVA intersect would have provided a suitable 3′ SS in a consensus SVA$_{F1}$. The 3′ SS consensus sequence is YYYYYYYYYYN CAG/G, where Y represents a pyrimidine and N is any nucleotide (Wang and Cooper 2007). The SVA portion of SVA$_{F1}$ aligns to SVA$_{Rep}$ beginning at position 388 of SVA$_{Rep}$ which is located in the 3′ region of the *Alu*-like fragment, 35 bp upstream of the VNTR. The sequence upstream of position 388 in SVA$_{Rep}$ is CCTCCACCTCC CAG (YYYYY-YYYYNCAG), a close match to the 3′ SS consensus sequence.

An SVA master element on CH10

The SVA on CH3, the progenitor to SVA$_{HLA\text{-}A}$ (Fig. 2A), lacks a target-site duplication (TSD) directly flanking the SVA$_{F1}$ (Fig. 2B). Given that retrotransposons are able to transduce sequences 5′ and 3′ of their location in the genome (Moran et al. 1999; Xing et al. 2006; Goodier and Kazazian 2008), we searched for a TSD further upstream and downstream from the CH3 SVA$_{F1}$. We identified a 15-nt TSD, with the 5′ duplication directly in front of a truncated *AluSc* (153 nt) and the 3′ duplication following a polyA tail downstream from a non-RepeatMasker annotated sequence. The entire insertion between the *AluSc* and the terminal 3′ polyA tail nearest the 3′ TSD consisted of (1) the *AluSc*, followed by (2) a SVA$_{F1}$, (3) an *AluSp*, and then (4) a 3′ transduction of 82 nt 3′ to the *AluSp*. When using BLAT (Kent 2002) to identify the source locus for the CH3 3′ transduction, 13 hits in addition to the CH3 query sequence were obtained (Supplemental Table 2). The source locus was identified on CH10 (Fig. 2C) due to the absence of a polyA tail 3′ of the transduced sequence present on CH3. Interestingly, the SVA on CH10 was flanked by a 5′ *AluSc* (320 nt) and a 3′ *AluSp* (299 nt). Overall, 13 SVA$_{F1}$ insertions contained the 3′ transduction from chromosome 10 (Supplemental Table 2), and all 13 had the *AluSp* and were variably truncated with three containing the *AluSc*, four containing some portion of the *MAST2* sequence and no *AluSc*, five truncated in the VNTR, and one truncated in the SVA polyA tail. Furthermore, one of the SVAs, which

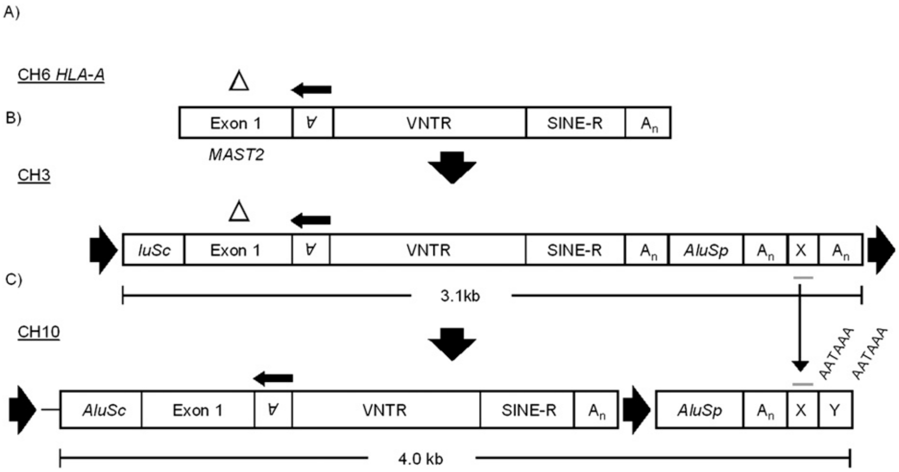

Figure 2. Identification of the SVA master locus on CH10. (*A*) The SVA insertion on CH6. The 40-bp deletion in the *MAST2* sequence (△) allowed the identification of the (*B*) CH6 progenitor element to be identified on CH3. The SVA insertion on CH3 contains both 5′ and 3′ tranductions. At the 5′ end, the SVA contains a truncated *AluSc* while at the 3′ end, an *AluSp* along with additional sequence, indicated by X, followed by a polyA tail, with the entire insertion flanked by target-site duplications. The 3′ transduction (X, red line) on CH3 allowed the identification of (*C*) the master element on CH10 along with 12 additional elements derived from the CH10 locus. The SVA on CH10 contains starting at it's 5′ end: 185 bp transduction derived from CH9, a full-length *AluSc*, the *MAST2*-SVA, an *AluSp*, polyA tail, and then a unique sequence, which was 3′ transduced, and which allowed the identification of CH10 as the source locus. A target-site duplication flanks the insertion, which inserted on CH10. Downstream polyA signals are displayed over the X and Y transduction sequences.

CH10 was the source locus for, had a 160-nt 3′ transduction. This element represents a transcript from the CH10 locus that bypassed the polyA signal at which the other 12 elements terminated. The sequence directly after the *AluSp*, the original source for these transductions, contains two canonical polyA signals, AATAAA (Colgan and Manley 1997), which are 15–20 nt upstream of the polyA tails of the SVAs derived from the CH10 locus (Fig. 2C).

To distinguish whether the SVA on CH10 inserted alone, or with the *AluSc* and/or *AluSp*, we searched for TSDs of each element and examined the chimpanzee reference sequence (Chimpanzee Sequencing and Analysis Consortium 2005). Only the *AluSp*, hereafter referred to as 3′ *Alu*, was present in the chimp reference sequence, suggesting that it was the first insertion on CH10 (Fig. 3A) and that the SVA insertion occurred, with or without the *AluSc*, hereafter referred to as 5′ *Alu*, sometime since our last common ancestor with chimp. Furthermore, the 3′ *Alu* on CH10 is at least 25 million years (Myr) old because it is present in the *Rhesus macaque* genome sequence (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007). The 5′ *Alu* on Ch10 could be traced back to a locus on CH9, due to 185 nt present directly upstream of the 5′ *Alu* on CH10 (Fig. 3E), which represents a 5′ transduction from the CH9 *AluSc* source locus (Fig. 3C). On CH10 there is a 13-nt TSD flanking the 5′ *Alu* containing the 5′ transduction and the SVA, suggesting that the 5′ *Alu* and SVA retrotransposed as one unit (Fig. 3C). However, at the CH9 locus there is no SVA downstream from the *AluSc* in the human reference sequence.

Identification of multiple SVA TSSs

SVA is a nonautonomous retrotransposon and was previously thought to rely on an internal promoter to initiate its transcription, similar to LINE-1 (Swergold 1990) and *Alu* (Di Segni et al. 1981; Duncan et al. 1981; Fritsch et al. 1981). Previous attempts in our laboratory to locate the SVA promoter have led to ambiguous results (MC Seleme and HH Kazazian, unpubl.). We set out to identify the SVA TSS for insight into how SVA mRNA is transcribed.

We used 5′ RACE to identify novel SVA 5′ ends from total RNA extracted from cell lines (see Methods) and chimpanzee testes. Currently, both the requirements for SVA transcription and the repertoire of expressed SVAs are unknown. We identified a total of 56 unique SVA-associated TSSs after sequencing and analysis of human and chimp SVA 5′ RACE products (Table 1). We grouped the TSSs into three classes: (1) internal SVA TSSs (Supplemental Table 3); (2) 5′ TSSs, defined as any position upstream of SVA annotated sequence (Supplemental Table 4); and (3) examples in which part of the sequence aligned within the SVA and part aligned upstream with a large gap in between, representing transcripts where 5′ sequences are spliced into the SVA part of the transcript (Table 2). The 26 class I TSS are scattered throughout the SVA but tend to cluster toward the 5′ end of the element (Fig. 4). The 14 class II TSSs start 76–440 bp upstream of SVAs in the human or chimp genome (Fig. 4; Supplemental Table 2).

Class III SVA-associated TSSs represent splicing into the SVA

We identified 17 class III TSSs, 16 of which are unique (13 human and four chimp) where the 5′ ends mapped upstream and represented splicing into SVA at 10 different 3′ SS (Table 2). SVA splicing events are listed in Table 2, with the 3′ SS position and 3′ SS sequence annotated relative to $\mathrm{SVA_{Rep}}$. Twelve of 17 class III TSSs involved exons from known genes/ESTs. One gene, *AFF1* on CH4 had two SVA alternative splicing events identified by 5′ RACE, suggesting the same SVA may contain multiple functional 3′ SS.

Many SVA splicing events are present in human ESTs

To identify further examples of SVA splicing, we performed a computational survey to identify splicing events. We focused on splicing events involving the 228 SVA elements present in intronic regions oriented in the same direction as the surrounding gene. EST databases (Boguski et al. 1993) were mined for uniquely aligned sequences with evidence of intronic SVA expression. Spliced ESTs were selected where one or more blocks aligned within the intronic SVA sequence, and the exon–SVA junctions were examined. The SVA sequences upstream and downstream from the EST junctions were compared to the SS consensus sequence, and those containing the canonical "(T/C)AG" trinucleotide (Wang and Cooper 2007) at the junction preceded by a reasonable polypyrimidine tract were kept. We defined an event as a unique exon 5′ SS and a unique SVA 3′ SS. If multiple overlapping ESTs having the same 5′ SS and SVA 3′ SS existed, we called it one event. ESTs with the same 5′ SS and two different SVA 3′ SS were called two events.

In total, 16 events, involving 14 genes, at eight different SVA 3′ SS were detected, supporting the notion that splicing into SVAs

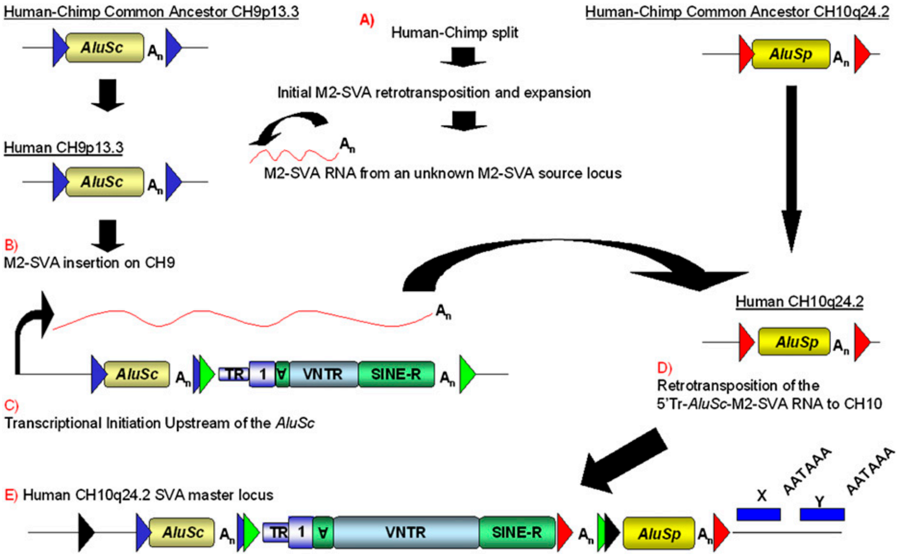
Human-Chimp Common Ancestor CH9p13.3
AluSc
A$_n$
A)
Human-Chimp split
Initial M2-SVA retrotransposition and expansion
A$_n$
M2-SVA RNA from an unknown M2-SVA source locus
Human-Chimp Common Ancestor CH10q24.2
AluSp
A$_n$
Human CH9p13.3
AluSc
A$_n$
B)
M2-SVA insertion on CH9
A$_n$
Human CH10q24.2
AluSp
A$_n$
AluSc
A$_n$
TR
1
∀
VNTR
SINE-R
A$_n$
D)
Retrotransposition of the
5'Tr-AluSc-M2-SVA RNA to CH10
C)
Transcriptional Initiation Upstream of the AluSc
E) Human CH10q24.2 SVA master locus
AATAAA
AATAAA
X
Y
AluSc
A$_n$
TR
1
∀
VNTR
SINE-R
A$_n$
AluSp
A$_n$

Figure 3. Genesis of the SVA master locus on CH10. The *AluSc* on CH9 and the *AluSp* on CH10, hereafter 5′ *Alu* and 3′ *Alu*, respectively, are shown present in the human–chimp common ancestor. The original target-site duplications (TSDs) for the 5′ *Alu* and the 3′ *Alu* are indicated by blue and red arrowheads, respectively. (*A*) After the human–chimp split, the initial SVA$_{F1}$ was formed due to an alternative splicing event. A SVA mRNA was transcribed from either the founder SVA$_{F1}$ insertion or some other unknown SVA$_{F1}$ locus in the human genome, derived from the original SVA$_{F1}$ founder insertion. (*B*) The SVA mRNA is retrotransposed directly downstream from the 5′ *Alu* on CH9 as displayed by the SVA insertion flanked by TSDs (green arrowheads). (*C*) Transcriptional initiation occurred upstream of the 5′ *Alu*, generating a composite mRNA containing in order (1) 185 nt of 5′ transduced sequence from CH9, (2) the 5′ *Alu*, and (3) the SVA$_{F1}$. (*D*) The composite mRNA from CH9 retrotransposed to CH10 directly upstream of the 3′ *Alu*. (*E*) The CH10 master element present in the human genome reference is shown. Directly downstream from the 3′ *Alu* TSD, unique sequence blocks (blue boxes), labeled X and Y, are present in the daughter insertions derived from this locus. The polyA signals utilized in the 3′ transductions are indicated by AATAAA relative to the unique sequence blocks.

occurs with some frequency across the genome (Table 2). We found one gene, *C2CD3*, which had ESTs aligning to three different 3′ SS locations at the *C2CD3* locus, AG138, AG319, and AG386, further indicating that multiple functional 3′ SS exist within SVA.

Gene-trapping occurs in primates

The upstream sequence of all class II TSSs were aligned to either the human or chimp reference sequence using BLAT (Kent 2002) to determine whether the sequence was present elsewhere in the genome, which would indicate potential SVA retrotransposed 5′ transductions. Of the 14 class II TSSs, only the lone example from chimp aligned elsewhere in the chimp reference sequence. The sequence consisted of 423 bp upstream of a truncated SVA on 3q11.2 (Fig. 5A) that aligned to multiple genomic locations with 91%–95% identity, one of which was a 342-bp hit on 15q14 (Fig. 5B). Further analysis revealed that the SVAs on CH3 and CH15 were insertions derived from two different SVA alternative splicing events, and that the transduced exons were derived from the transmembrane phosphatase with tensin homology (*TPTE*) gene on CH22 in chimp (CH21 in humans). *TPTE* is a testis-specific gene that shares significant homology with *PTEN* (Chen et al. 1999). However, similar to *MAST2*, no SVA is present in the *TPTE* gene reference sequence. These SVA insertions differ in that the 5′ transduction of the CH3 insertion is 531 bp and contains exons 1, 18, 19, and 20, while the 5′ transduction of the CH15 insertion is 561 bp and contains five exons, an unannotated exon in intron 16 referred to as 16a, and exons 17, 18, 19, and 20 (Fig. 5B). The CH3 insertion spliced into AG 336 of SVA, while the CH15 insertion spliced into AG 386. Both 3′ SS were identified previously in our data. Both the CH3 and CH15 SVA insertions are present in the human reference genome; however, both are absent from the orangutan reference genome. We used the ensemble browser to identify the presence of the CH15 SVA$_{TPTE}$ insertion in the gorilla genome sequence, while the CH3 SVA$_{TPTE}$ insertion was not located.

After finding SVA$_{F1}$ and SVA$_{TPTE}$, we examined the human genome to see if we could identify additional examples of splicing followed by retrotransposition. We searched the human genome reference sequence (Lander et al. 2001) for SVAs that had a non-SVA sequence upstream present multiple times in the genome. This approach would identify examples where an SVA provided a 3′ SS and this mRNA retrotransposed and then likely jumped again. Computational analysis identified an SVA subgroup, SVA$_{RHOT1}$, where the first six exons, 532 bp, of the *RHOT1* gene on chromosome 17, also known as mitochondrial Rho GTPase 1 (*MIRO-1*) (Fransson et al. 2003), were spliced into SVA at AG 336 and subsequently retrotransposed (Supplemental Fig. 1). However, unlike the *MAST2* and *TPTE* examples, there is an SVA residing in the sixth intron of *RHOT1* in the human reference genome. There are three SVAs in the human genome containing upstream *RHOT1* processed exons, located on 13q11, 18p11.21, and 21q11.2. Two SVA$_{RHOT1}$ copies were identified in the chimp genome on CH13 and CH18 and one on CH13 in the gorilla genome. Both the human and chimp SVA$_{RHOT1}$ insertions are missing the first 36 nt of the *RHOT1* 5′ UTR and share identical 13-nt TSDs. Surprisingly, these different insertions represent duplications and not individual retrotransposition events. The human SVA$_{RHOT1}$ insertions are within large CNVs, present in both human and chimp several times, yet only the CH13, CH18, CH21 CNVs contain the SVA$_{RHOT1}$ insertion (Supplemental Fig. 1). We concluded that the CH13 SVA$_{RHOT1}$ insertion most likely represents the original SVA$_{RHOT1}$ insertion based upon its presence in the gorilla genome draft sequence; however, this is contingent upon the CNVs containing the other SVA$_{RHOT1}$ insertions not being polymorphic in these species. Both the SVA insertion in intron 6 of *RHOT1* and all of the SVA$_{RHOT1}$ insertions are absent from the orangutan genome sequence, suggesting that the SVA at the *RHOT1* locus and the

Table 1. SVA-associated transcriptional start sites

Class	Human	Chimp	Total unique
I: Internal	16	10	26
II: Upstream	13	1	14
III: Splicing	13	4	16[a]
Total			56

[a] CH17 SVA splicing event was identified in both species.

Table 2. 3′ Splice sites identified within SVA

Splice acceptor	Sequence	Gene/EST	Species	UCSC Genome Browser	SVA subfamily	Method
52	CTCTCTTCCCACAG	n/a	Hsa, Ptr	chr17: 78004366–78005622	A	5′ RACE
52	CTCTCTTTCCACAG	*FAM122C*	Hsa	chrX: 133800359–133802069	D	EST analysis
70	CCTCTCATGTGGAG	*CF145918*	Hsa	chr2: 27435097–27436626	F	5′ RACE
70	CCTCTGATGCCAAG	n/a	Hsa	chr3: 9956016–9957408	D	5′ RACE
76	ATGCCGAACCGAAG	*SgK494*	Hsa	chr17: 23954622–23956451	D	5′ RACE
138	TTCTCCTGCCTCAG	*AFF1*	Hsa	chr4: 88135456–88136923	D	5′ RACE
138	TTCTCCTGCCTCAG	*DA477153*	Hsa	chr12: 64814008–64816131	F	5′ RACE
138	TTCTCCTGCCTCAG	*SCYE1*	Hsa	chr4: 107461381–107462912	B	EST analysis
138	TTCTCCTGCCTCAG	*CLIC4*	Hsa	chr1: 24997705–25000164	F	EST analysis
138	TTCTCCTGCCTCAG	*C2CD3*	Hsa	chr11: 73517261–73518822	D	EST analysis
138	TTCTCCTGCCTCAG	*MAPKAP1*	Hsa	chr9: 127260318–127261973	D	EST analysis
138	TTCTCCTGCCTCAG	*ESCO1*	Hsa	chr18: 17425626–17427548	D	EST analysis
138	TTCTCCTGCCTCAG	*CF145918*	Hsa	chr2: 27435097–27436626	F	EST analysis
208	TTTTTTTGGTGGAG	*LOC650368*	Ptr	chr11: 3400235–3400597	B	5′ RACE
269	GAGTGATCTGCCAG	*RAB7A*	Hsa	chr3: 129974255–129975628	C	EST analysis
269	GAGTGATCCGCCAG	n/a	Ptr	chr16: 80967766–80968128	B	5′ RACE
319	TCTCGTTCACGCAG	*C2CD3*	Hsa	chr11: 73517261–73518822	D	EST analysis
336	Unknown	*TPTE*	Hsa	chr21: 9928614–10012791[a]	B	Genome
336	TCAATGGTGCCCAG	*MTFR1*	Hsa	chr8: 66736844–66738336	D	5′ RACE
336	TCAATGGTGCCCAG	*ELOVL5*	Hsa	chr6: 53277226–53279021	D	5′ RACE
336	TCAATCTTGCCCAG	*BU159250*	Hsa	chr19: 11648074–11649703	B	5′ RACE
336	TCAGTGTTGCCCAG	*RHO/T1*	Hsa	chr17: 27529349–27531779	A	Genome
336	TCAATGTTGCCCAG	*VPS24*	Hsa	chr2: 86628818–86631061	A	EST analysis
336	TCAATGTTGCCCAG	n/a	Ptr	chr8: 142697098–142697454	B	5′ RACE
386	Unknown	*MAST2*	Hsa	chr1: 46041872–46274383[b]	F	Genome
386	Unknown	*TPTE*	Hsa	chr21: 9928614–10012791[a]	B	Genome
386	CCTCCACCTCCCAG	*SEP15*	Hsa	chr1: 87137058–87138908	D	5′ RACE
386	CCTCCATCTCCCAG	*IARS*	Hsa	chr9: 94028908–94030246	D	5′ RACE
386	CCTCCACCTCCCAG	*MTFRI*	Hsa	chr8: 66736844–66738336	D	RT-PCR, ECR analysis
386	CCTCCACCTCCCAG	*SLC25A12*	Hsa	chr2: 172430609–172432417	D	EST analysis
386	CCTCCACCTCCCAG	*C2CD3*	Hsa	chr11: 73517261–73518822	D	EST analysis
386	CCTCCACCTCCCAG	*SLC38A9*	Hsa	chr5: 55026855–55028866	D	EST analysis
389	CCACCTCCCAGCAG	*AFF1*	Hsa	chr4: 88135456–88136923	D	5′ RACE
450	GCCATCCCATCTAG	*AK096668*	Hsa	chr5: 43613499–43615016	D	5′ RACE
450	GCCATCCCATCTAG	*PPARD*	Hsa	chr6: 35459078–35461415	F	EST analysis
450	ACCACCCCATCTAG	*ZNF611*	Hsa	chr19: 57914830–57916788	A	EST analysis
450	ACCATCCCATCTAG	*ORCIL*	Ptr	chr1: 53274568–53274862	E	RT-PCR

[a] Location of *TPTE* gene in HG18.
[b] Location of *MAST2* gene in HG18.

original SVA_{RHOT1} event occurred some time after the orangutan diverged from the human–gorilla last common ancestor, aging the insertion between 8 and 15 Myr.

SVA splicing is not rare

To study the potential mutagenic potential of SVA splicing, we cloned two SVAs from the human genome, SVA_{C2CD3} and SVA_{MTFR1}, containing multiple functional 3′ SS. We cloned the SVAs into the intron of a splicing minigene construct named pPKC-EGFP (Fig. 6A), hereafter pPKC-SVA. 293T cells were transfected with pPKC-SVA, and total RNA was harvested after 1 d. To characterize SVA splicing and identify functional 3′ SSs, we performed RT-PCR with a forward primer in the first exon, PKC, and then used three reverse primers (6A) in independent reactions to answer three questions: (1) are SVAs exonized (PKC For + 1R primers); (2) what 3′ SS exist in SVA (PKC For + 2R); and (3) can we detect SVA gene-trapping (PKC For + 3R).

PCR products were analyzed on a 2% agarose gel (Fig. 6B); bands were cloned and sequenced. Bands from the lane labeled 1R for pPKC-SVA_{C2CD3} corresponded to the normal splicing, PKC exon to EGFP exon, and also to SVA exonization events. Five SVA exonization events utilizing three different 3′ SS and three different 5′ SS with SVA exons ranging from 159–359 nt (Fig. 6B; Table 3) were identified. Three out of the five SVA exonizations shift the reading frame, while all five SVA exonization events introduce premature stop codons (PTCs) located in the exonized SVA sequence.

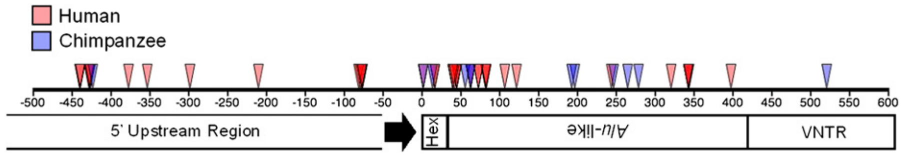

Figure 4. SVAs contain many transcriptional start sites (TSSs). 5′ RACE was performed on total RNA extracted from 293T, HeLa, and nTera cell lines, along with chimp testes. A nested PCR was performed followed by Sanger sequencing of Topo clones. The 5′ RACE adaptor was identified, and the first nucleotide after it was annotated as the TSS. Human (red triangles) and chimp TSS (blue triangles) within SVA are aligned relative to the position in SVA_{Rep}. The CCCTCT hexamer (Hex), *Alu*-like region, and part of the VNTR are shown. TSSs aligning upstream of SVA are displayed relative to the first nucleotide of the SVA insertion in the genome. Both human (red triangles) and chimp (blue triangles) TSSs identified by 5′ RACE are shown.

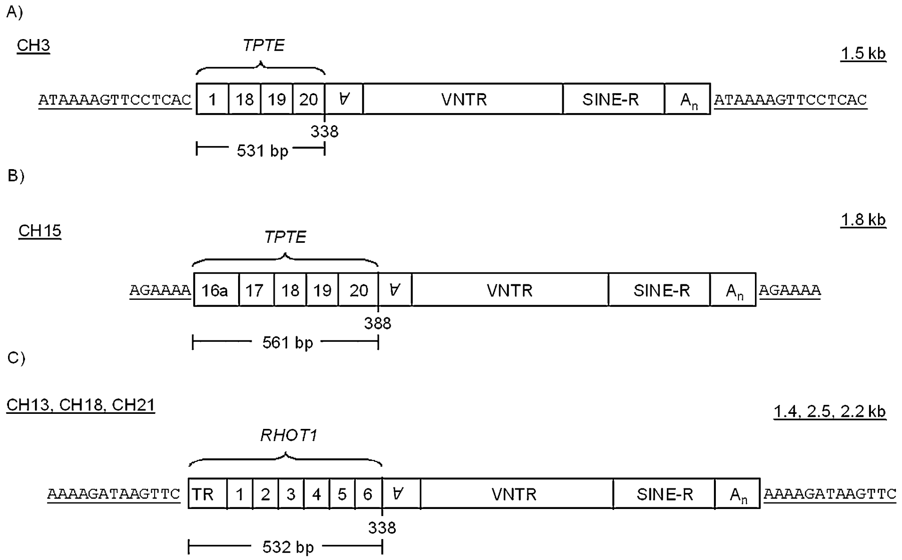

Figure 5. SVA gene-trapping has occurred in other hominids. (*A,B*) Two different SVA splicing events followed by retrotransposition derived from the *TPTE* locus on chromosome 21 flanked by a target-site duplication. (*A*) CH3 SVA containing four processed *TPTE* exons that utilized AG 336. (*B*) CH15 SVA containing five processed exons that utilized AG 386. (*C*) The structure of SVA insertions on CH13, CH18, and CH21 containing the six processed exons from the *RHOT1* gene on CH17 with identical target-site duplications.

SVA splicing events using pPKC-SVA and verified by sequencing are listed in Table 3. It is noteworthy that a 3′ SS site was identified in the SINE-R domain of SVA (Fig. 6B, lane 3, lower band). These PCR results suggest that SVA splicing is not rare and that both SVA exonization and SVA gene-trapping can occur in the same SVA.

Semi-quantitative RT-PCR followed by Southern blotting using amplicons from a PKC for and 1R PCR (Fig. 6A,C) was carried out for pPKC-EGFP, pPKC-SVA$_{C2CD3}$, and pPKC-SVA$_{MTFR1}$ to estimate SVA exonization. The intensities for normal splicing varied across the samples, so the Southern blot was exposed overnight in order to ensure no bands were present in the vector-only lane (data not shown). The ratio of the higher molecular weight bands indicative of SVA exonization relative to PKC-EGFP splicing within that lane was determined using a phosphorimager (Fig. 6C, bottom panel). The ratio of total SVA$_{C2CD3}$ exonization relative to PKC-EGFP splicing was 0.19:1, while total SVA$_{MTFR1}$ exonization to PKC-EGFP was 0.12:1.

Discussion

These data are the first to provide insight into how SVA retrotransposons are expressed and how they might impact gene expression. Our data suggest that SVAs are expressed in a variety of ways in humans and chimps. Recently, a study identified many TSSs in LINE-1s and SINEs in human and mouse tissues and cell lines. (Faulkner et al. 2009). Whether or not internal TSSs identified here represent retrotransposition-competent SVA transcripts is unknown. Many SVAs have transduced sequence 5′ of their location in the genome to other locations (Damert et al. 2009); this is consistent with our observation of upstream TSSs. What is unclear is whether upstream TSSs represent solely upstream promoters driving SVA expression or whether something inherent to SVA directs transcriptional initiation upstream.

The SVA$_{F1}$ subfamily, SVA$_{TPTE}$, and SVA$_{RHOT1}$ together indicate that if an SVA loses the CCCTCT hexamer and most of the *Alu*-like region due to alternative splicing into it, the remaining SVA sequence is able to retrotranspose. Furthermore, the lack of most of the *Alu*-like region suggests that the model suggested by Mills et al. (2007), adapted from Boeke's model (Boeke 1997), where the SVA *Alu*-like region hybridizes to *Alu* RNAs at the ribosome in order to compete for the LINE-1 ORF2 reverse transcriptase, may not be case. However, it is possible that SVA RNA may be located at the ribosome where competition for the LINE-1 ORF2 takes place, but that it is not hybridizing to *Alu* RNA.

The lack of SVA$_{F1}$s with a complete *MAST2* 5′ UTR and first exon suggests that a full-length *MAST2* 5′ UTR first exon is not required for transcription or retrotransposition. Exactly, how the *MAST2* sequence contributed to the expansion of SVA$_{F1}$s in the absence of the CCCTCT hexamer and the *Alu*-like region still needs to be determined. One possibility is that the *MAST2* sequence, in combination with certain SVA sequence variants or in a specific genomic context, enhances transcription or retrotransposition relative to a canonical SVA. It is worthwhile to note that *TPTE* is a testis-specific gene (Chen et al. 1999) and that we found the CH3 SVA$_{TPTE}$ by 5′ RACE, and the TSS was in exon 1 of the transduced *TPTE* sequence.

We have identified 11 3′ SS throughout SVA, in addition to multiple 3′ SS in the VNTR, including examples from all subfamilies except E, and we have shown that exonization can occur. Whether or not older SVAs are still retrotransposing in humans is currently unknown; however, the older SVAs are still able to be transcribed and may influence transcription if residing within an intron in the same orientation as the gene.

SSs within retrotransposons are not uncommon. *Alu*s are primate-specific retrotransposons that are known to exonized (Sorek et al. 2002; Lev-Maor et al. 2003). In addition to *Alu*, internal splicing has been observed in the human L1 (Belancio et al. 2006) and the zebrafish LINE (Tamura et al. 2007). Although, it appears that if an SVA undergoes a splicing event, it can still carry out subsequent rounds of retrotransposition, as indicated by SVA$_{F1}$.

SVA splicing followed by retrotransposition may be rare based on only four examples identified in the human genome, three of which are present in the chimpanzee. Additional splicing followed by retrotransposition events may have occurred, but the results are undetectable due to truncation upon insertion or low allele frequency. On the contrary, splicing into the SVA is not rare, as indicated by our semi-quantitative PCR data, which suggest that SVA exonization events may occur 12%–19% of the time relative to normal splicing in our minigene.

Currently, the ratio of SVA gene-trapping to SVA exonization has not been determined. Here we provide a low-end estimate for SVA splicing by assessing SVA exonization using a splicing minigene. Our SVA exonization estimate may be an underestimate because SVAs contain more than 10 nonsense codons in each reading frame on the sense strand and exonization of these sequences may induce nonsense-mediated decay if the exonized SVA sequence is more than 50–55 nt upstream of the 3′ most

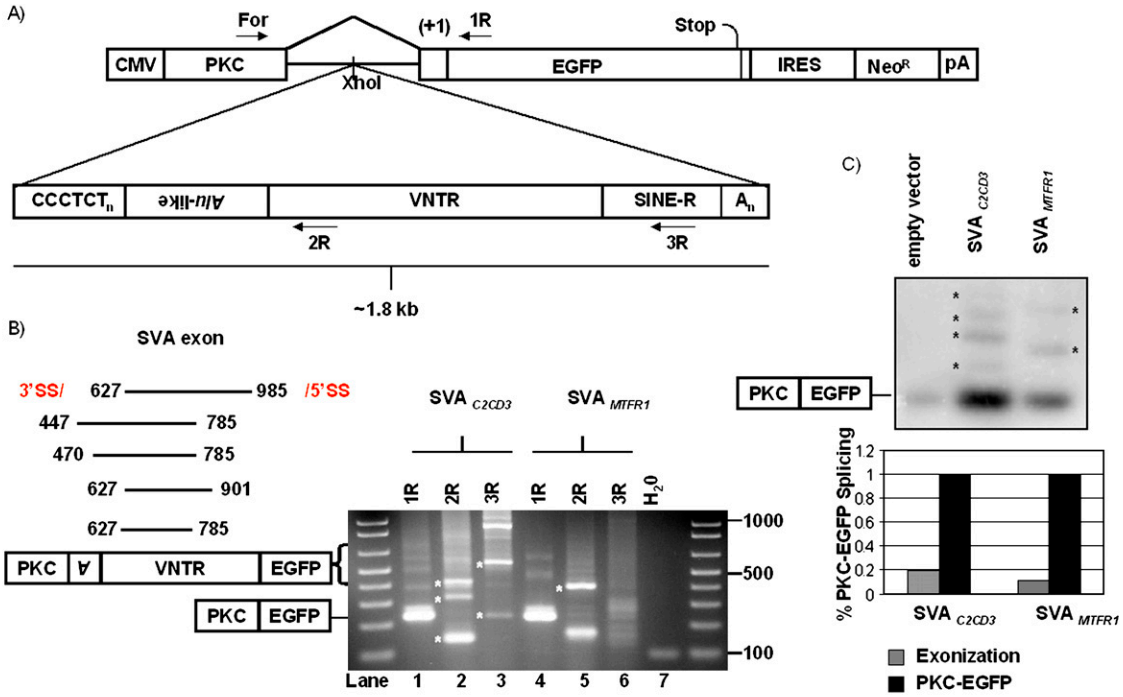

Figure 6. SVA gene-trapping and exonization are not rare. (*A*) Two SVAs were cloned into PKC-EGFP (Newman et al. 2006) to test the mutagenic potential of SVA splicing. Primers used for RT-PCR are marked. (*B*) RT-PCR was performed on total RNA extracted from 293T cells transiently transfected with pPKC-EGFP containing one of two different SVAs cloned into the intron. SVA exonization events (*left* panel) are annotated with the first and last nucleotide of the SVA exon, all of which occur within the *Alu*-like and VNTR domains. A representative agarose gel displaying SVA alternative splicing events is shown (*right* panel) (see Table 3). (*) Indicates bands verified by DNA sequencing to be SVA splicing events. (*C*) Semi-quantitative PCR to determine the frequency of SVA exonization. Ten cycles of PCR on cDNA from individual pPKC-EGFP, pPKC-SVA$_{C2CD3}$, and pPKC-SVA$_{MTFR1}$ transfections were carried out using PKC For and 1R. PCR products were resolved on a 2% agarose gel, followed by overnight transfer to a membrane, and subsequent probing using a DNA probe targeting the PKC exon (*top* panel). (*) Indicates bands quantified by a phosphorimager. Total SVA exonization was normalized to PKC-EGFP splicing within each respective lane and graphed (*bottom* panel).

exon–exon junction (Nagy and Maquat 1998). Be that as it may, an SVA splicing event, exonization or trapping (Fig. 7), will likely lead to a dead-end to the protein-coding capacity of the mRNA because either event has the capability to produce truncated proteins.

Most SSs identified using the splicing minigene were not identified by 5′ RACE, such as the 3′ SSs in the VNTR and SINE-R. This is likely due to the nested PCR approach utilized in 5′ RACE. However, downstream SVA 3′ SS may be selected for in the splicing minigene due to the small size of the intron. Each SVA was cloned into pPKC-EGFP with less than 100 bp of flanking DNA to ensure splicing was inherent to the SVA and not due to intronic splicing enhancers. If SVA is cloned in as a larger fragment, one may see 3′ SS selection shift toward the 5′ end of SVA residing in the *Alu*-like domain.

If SVAs impact gene expression by being alternatively spliced, then one would expect to observe either a depletion of SVAs in genes or on the coding strand of genes. An underrepresentation was observed for SVAs on the coding-strand in the human genome; 1060/2772 SVAs are in RefSeq genes, with 228/1060 on the coding strand (introns or exons) and 832/1060 on the antisense strand. This underrepresentation of intronic SVA insertions on the coding strand is highly significant ($P < 2.2 \times 10^{-16}$) under a null hypothesis of random orientation. Likewise, a similar significant underrepresentation is observed in the chimp reference genome, with 228 (partial overlap with the human 228) of 1024 intronic SVAs oriented on the sense strand with respect to the surrounding gene. Nevertheless, this SVA strand bias may be due to a factor other than selection, such as SVA insertional preference.

Altogether, these data show that SVAs are alternatively spliced in cell culture, in tissue, and in vivo. We speculate that SVAs may influence local gene expression by providing alternative SSs and might even account for some of the variation in gene expression observed within and across hominids. As more primate genomes are sequenced along with more studies on SVA, the impact of this retrotransposon will become clear. Thus, although SVAs effect on genome evolution may be less than that of L1 and *Alu* because of their smaller numbers, SVA has had recent effects that are likely growing with their continued expansion as indicated by the SVA$_{F1}$ subfamily and the CH10 subgroup. In another 50 Myr, the SVA effect on genome evolution may be much greater than that of L1 and *Alu*.

Table 3. Splices sites identified in SVA splicing mini-gene

Event	3′ SS (*C2CD3*)	3′ SS Sequence	5′ SS (*C2CD3*)	Exon size
Exon	386 (445)	CCTCCACCTCCCAG	VNTR (786)	339
Exon	468 (411)	TTGGCCTCCCAAAG	VNTR (786)	316
Exon	VNTR (625)	GCCATCCCATCTAG	VNTR (986)	359
Exon	VNTR (625)	GCCATCCCATCTAG	VNTR (786)	159
Exon	VNTR (625)	GCCATCCCATCTAG	VNTR (902)	275
Trap	VNTR (625)	GCCATCCCATCTAG	—	—
Trap	1372 (1352)	CTGTGTCCACTCAG	—	—
N/A	386 (445)	CCTCCACCTCCCAG	—	—
N/A	VNTR (625)	GCCATCCCATCTAG	—	—
N/A	VNTR (629)	TCCCATCTAGGAAG	—	—

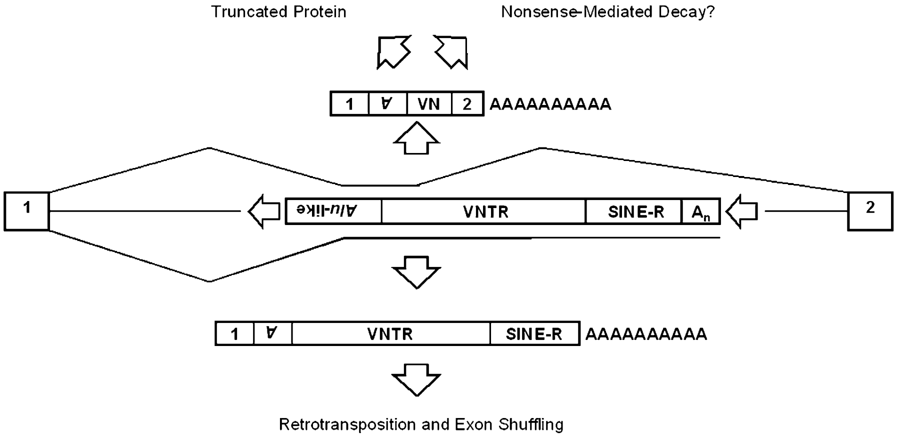

Figure 7. SVA alternative splicing outcomes. An intronic truncated SVA is shown (*middle*). The SVA is truncated because these SVAs are still likely to be spliced. If SVAs are exonized, they will likely generate a truncated protein or subject the mRNA to nonsense-mediated decay due to the inclusion of SVA nonsense codons (*top*). If SVAs mimic an endogenous gene-trap, that is provide a 3′ SS followed by termination at the SVA or downstream polyA signal, this may result in truncated proteins, but more importantly the retrotransposition of exons.

Methods

Sequence analysis

BLAT (Kent 2002) and BLAST (Altshul et al. 1990) were used in mapping sequences to the reference genomes. Censor (Kohany et al. 2006) and RepeatMasker (Smit et al. 1996) were used to identify relative positions in SVA and subfamily classification, respectively.

Cell culture

293T and HeLa cells were grown in a humidified, 5% CO_2 incubator at 37°C in DMEM (GIBCO) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and 100 U/mL penicillin, 0.1 mg/mL streptomycin. nTera cells were grown as described above except that the media was supplemented with nonessential amino acids.

5′ RACE, cDNA synthesis, and PCR

RNA extraction was performed using the RNeasy kit (Qiagen) according to the manufacturer's instructions. DNase treatment consisted of using twice the recommended amount of RQ1 RNase-Free DNase (Promega) followed by ethanol precipitation of the RNA. Chimp testis was used for RNA extraction (Department of Veterinary Medicine and Surgery, University of Texas M.D. Anderson Cancer Center, Houston, TX).

5′ RACE was performed using the GeneRacer Kit (Invitrogen) with 5 μg DNase-treated RNA as the starting material. First-strand cDNA synthesis was performed using the supplied SuperScript III RT kit with random hexamer primers or Array Script Reverse Transcriptase (Ambion). All steps were carried out according to the manufacturer's instructions. A two-round PCR scheme was utilized in order to enrich for SVA containing transcripts using GoTaq (Promega) or Expand Long (Roche) according to the manufacturer's instructions with 1 μL of cDNA containing the 5′ RACE adaptor. The first round of PCR consisted of reverse primers complementary to the SINE-R region of SVA_{Rep}. Primers complementary to the *Alu*-like region were used for the second round of PCR. PCR cycling parameters consisted of variations on touchdown PCR with the initial annealing temperature at 60°C and cycled down to 50°C over 40 cycles. PCR reactions were analyzed on 1%–1.5% agarose gels. Bands of varying size were cut out, gel purified using QIAquick Gel Extraction kit (Qiagen), Topo cloned (Invitrogen), and sequenced.

RT-PCR

DNase-treated RNA was reverse transcribed using random primers with the SuperScript III First-Strand Synthesis SuperMix (Invitrogen) according to the manufacturer's instructions. One microliter of cDNA was used in PCR with GoTaq (Promega) or Expand Long (Roche).

EST analysis

EST locations corresponding to human genome assembly hg18 were obtained from the UCSC Genome Browser and stored in a local relational database along with SVA and RefSeq gene locations. ESTs with blocks aligning unambiguously within SVA locations present on the same strand as a RefSeq intron were further analyzed using Perl scripts locate EST splice junctions within the SVA. Junctions corresponding to splicing patterns consistent with SVA gene-trapping were compared to the SS consensus sequence to ensure presence of the relevant nucleotides.

SVA splicing minigene, transfection, and RT-PCR

pPKC-EGFP has been previously reported (Newman et al. 2006). SVAs were amplified from human genomic DNA and subcloned into Topo (Invitrogen). The SVA was then amplified as a XhoI fragment and cloned into the XhoI site within the intron.

293T cells were seeded into T-75 flasks in order to be 50%–80% confluent upon transfection. Twenty-four hours later, 8 μg of each splicing minigene was transfected using 24 μL of Fugene6 (Roche) according to the manufacturer's instructions. Total RNA was isolated 1 d after transfection as described above. Five micrograms of DNase-treated RNA was reverse-transcribed with Array Script (Ambion) using an olio dT primer. Touchdown PCR from 59°C to 51°C over 40 cycles was performed with elongation at 72°C for 2 min using GoTaq Master Mix (Promega) with 1 μL of cDNA as template and primers at a final concentration of 0.2 μM per reaction. Amplicons were analyzed on 2% agarose gels.

Semi-quantitative PCR

One microliter of oligo dT primed cDNA derived from total RNA from splicing minigene transfections was amplified by 10 cycles of PCR (20 sec at 94°C, 30 sec at 57°C, 1 min at 72°C) using GoTaq MasterMix (Promega) with PKC forward and 1R primers in a 25 μL reaction. The entire reaction was resolved on a 2% agarose gel. Overnight alkaline transfer to N+ hybond membrane (Amersham) was performed followed by overnight hybridization with a 182-bp DNA probe labeled with [α-^{32}P]dCTP targeting the PKC exon at 65°C. Reaction products were imaged using a Storm 840 phosphorimager (GE Healthcare) and quantified with ImageQuant 5.2 (GE Healthcare). The intensity of each band was determined followed by the subtraction of background. SVA exonization band intensities were summed followed by normalization to PKC-EGFP splicing.

Acknowledgments

This work was supported by grants from the NIH. D.C.H. is funded by NIH training grant T32-M007229-27. We thank Dr. Yoshihide Ishikawa, Japanese Red Cross Central Blood Institute, for HLA-A deletion samples. We also thank Claude Warzecha and Dr. Russ Carstens for the PKC-EGFP construct.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Belancio VP, Hedges DJ, Deininger P. 2006. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* **34:** 1512–1521.

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.

Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. 2004. Natural genetic variation caused by transposable elements in humans. *Genetics* **168:** 933–951.

Boeke JD. 1997. LINEs and *Alus*—the polyA connection. *Nat Genet* **16:** 6–7.

Boguski MS, Lowe TMJ, Tolstoshev CM. 1993. dbEST—database for "expressed." *Nat Genet* **4:** 332–333.

Chen H, Rossier C, Morris MA, Scott HS, Gos A, Bairoch A, Antonarakis SE. 1999. A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. *Hum Genet* **105:** 399–409.

Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437:** 69–87.

Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes & Dev* **11:** 2755–2766.

Damert A, Raiz J, Horn AV, Löwer J, Wang H, Xing J, Batzer MA, Löwer R, Schumann GG. 2009. 5′-Transduced retrotransposons groups spread efficiently throughout the human genome. *Genome Res* doi: 10.1101/gr.093435.109.

Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked *Alu* sequences. *Nat Genet* **35:** 41–48.

Di Segni G, Carrara G, Tocchini-Valentini GR, Shoulders CC, Bralle FE. 1981. Selective in vitro transcription of one of the two *Alu* family repeats present in the 5′ flanking region of the human epsilon-globin gene. *Nucleic Acids Res* **9:** 6709–6722.

Duncan CH, Jagadeeswaran P, Wang RR, Weissman SM. 1981. Structural analysis of templates and RNA polymerase III transcripts of *Alu* family sequences interspersed among the human beta-like globin genes. *Gene* **13:** 185–196.

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **5:** 563–571.

Fransson A, Ruusala A, Aspenström P. 2003. Atypical Rho GTPases have roles in mitochondrial homeostasis and apoptosis. *J Biol Chem* **278:** 6495–6502.

Fritsch EF, Shen CK, Lawn RM, Maniatis T. 1981. The organization of repetitive sequences in mammalian globin gene clusters. *Cold Spring Harb Symp Quant Biol* **45:** 761–765.

Goodier JL, Kazazian HH Jr. 2008. Retrotransposons revisited: The restraint and rehabilitation of parasites. *Cell* **135:** 23–35.

Han JS, Szak ST, Boeke JD. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429:** 268–274.

Hassoun H, Coetzer TL, Vassiliadis JN, Sahr KE, Maalouf GJ, Saad ST, Catanzariti L, Palek J. 1994. A novel mobile element inserted in the alpha spectrin gene: Spectrin dayton. A truncated a spectrin associated with hereditary elliptocytosis. *J Clin Invest* **94:** 643–648.

Jurka J. 2000. Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet* **16:** 418–420.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110:** 462–467.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* **12:** 656–664.

Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa M, et al. 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* **394:** 388–392.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7:** 474.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. *Science* **300:** 1288–1291.

Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, Goto S, Matsumoto S, Tabuena D, Maranon E, Dantes M, et al. 2007. Reduced neuron-specific expression of the *TAF1* gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* **80:** 393–406.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16:** 793–805.

Mills RE, Bennett EA, Iskow RC, Devine SE. 2007. Which transposable elements are active in the human genome? *Trends Genet* **23:** 183–191.

Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* **283:** 1530–1534.

Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem Sci* **23:** 198–199.

Newman EA, Muh SJ, Hovhannisyan RH, Warzecha CC, Jones RB, McKeehan WL, Carstens RP. 2006. Identification of RNA-binding proteins that regulate FGFR2 splicing through the use of sensitive and specific dual color fluorescence minigene assays. *RNA* **12:** 1129–1141.

Ono M, Kawakami M, Takezawa T. 1987. A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* **15:** 8725–8737.

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.

Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the *Rhesus macaque* genome. *Science* **316:** 222–234.

Rohrer J, Minegishi Y, Richter D, Eguiguren J, Conley ME. 1999. Unusual mutations in Btk: An insertion, a duplication, an inversion, and four large deletions. *Clin Immunol* **90:** 28–37.

Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. 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.

Smit AFA, Hubley R, Green P. 1996. RepeatMasker Open-3.0. http://www.repeatmasker.org.

Sorek R, Ast G, Graur D. 2002. *Alu*-containing exons are alternatively spliced. *Genome Res* **12:** 1060–1067.

Swergold GD. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* **10:** 6718–6729.

Takasu M, Hayashi R, Maruya E, Ota M, Imura K, Kougo K, Kobayashi C, Saji H, Ishikawa Y, Asai T, et al. 2007. Deletion of entire *HLA-A* gene accompanied by an insertion of a retrotransposon. *Tissue Antigens* **70:** 144–150.

Tamura M, Kajikawa M, Okada N. 2007. Functional splice sites in a zebrafish LINE and their influence on zebrafish gene expression. *Gene* **390:** 221–231.

Wang GS, Cooper TA. 2007. Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8:** 749–761.

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. 2005. SVA elements: A hominid-specific retroposon family. *J Mol Biol* **354:** 994–1007.

Wilund KR, Ming Y, Campagna F, Arca M, Zuliani G, Fellin R, Ho Y, Garcia JV, Hobbs HH, Cohen JC. 2002. Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet* **11:** 3019–3030.

Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci* **103:** 17608–17613.

Received February 25, 2009; accepted in revised form July 20, 2009.