The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery

Julija Raiz¹, Annette Damert¹,³, Sergiu Chira³, Ulrike Held¹,², Sabine Klawitter², Matthias Hamdorf², Johannes Löwer¹, Wolf H. Strätling⁴, Roswitha Löwer¹ and Gerald G. Schumann¹,²,*

¹Section PR2/Retroelements, ²Division of Medical Biotechnology, Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, D-63225 Langen, Germany, ³Institute for Interdisciplinary Research in Bio-Nano-Sciences, Molecular Biology Center, Babes-Bolyai-University, Cluj-Napoca, Treboniu Laurean Street 42, RO-400271 Cluj-Napoca, Romania and ⁴Institut für Biochemie und Molekularbiologie, Universitätsklinikum Hamburg-Eppendorf, Martinistrasse 52, D-20246 Hamburg, Germany

Received February 21, 2011; Revised and Accepted September 27, 2011

ABSTRACT

SINE-VNTR-Alu (SVA) elements are non-autonomous, hominid-specific non-LTR retrotransposons and distinguished by their organization as composite mobile elements. They represent the evolutionarily youngest, currently active family of human non-LTR retrotransposons, and sporadically generate disease-causing insertions. Since preexisting, genomic SVA sequences are characterized by structural hallmarks of Long Interspersed Elements 1 (LINE-1, L1)-mediated retrotransposition, it has been hypothesized for several years that SVA elements are mobilized by the L1 protein machinery in trans. To test this hypothesis, we developed an SVA retrotransposition reporter assay in cell culture using three different human-specific SVA reporter elements. We demonstrate that SVA elements are mobilized in HeLa cells only in the presence of both L1-encoded proteins, ORF1p and ORF2p. SVA trans-mobilization rates exceeded pseudogene formation frequencies by 12- to 300-fold in HeLa-HA cells, indicating that SVA trans-mobilization represents a preferred substrate for L1 proteins. Acquisition of an AluSp element increased the trans-mobilization frequency of the SVA reporter element by ~25-fold. Deletion of (CCCTCT)ₙ repeats and Alu-like region of a canonical SVA reporter element caused significant attenuation of the SVA trans-mobilization rate. SVA de novo insertions were predominantly full-length, occurred preferentially in G+C-rich regions, and displayed all features of L1-mediated retrotransposition which are also observed in preexisting genomic SVA insertions.

INTRODUCTION

Three different families of non-LTR retrotransposons are actively mobilized in the human genome. These are Long Interspersed Elements 1 (LINE-1, L1), Alu elements (Short Interspersed Elements, SINE) and SVA (SINE-VNTR-Alu) elements. Their success is documented by the fact that non-LTR retrotransposons encompass ~34% of the human genome, making them the most populous group of transposable elements in the human genome (1). L1 elements are the only currently known retrotransposons in the human genome that are coding for the protein machinery required for their own mobilization. Despite the cis-preference (2) of L1 proteins for their own encoding RNA, RNA polymerase III transcripts [Alu, 7SL, U6 and hY sequences (3–6)] mutated full-length L1 RNAs (2), and cellular mRNAs [resulting in processed pseudogene formation (7–9)] were experimentally demonstrated to be trans-mobilized by hijacking the L1-encoded protein machinery.

Since genomic preexisting insertions of the hominid-specific non-autonomous non-LTR retrotransposon SVA exhibit the classical hallmarks of L1-mediated
retrotransposition (10–12,24), SVA was also assumed to use the L1 protein machinery for its own mobilization (13). These hallmarks are: (i) insertion at a 5'-TITT/AA-3’ consensus sequence of the L1 endonuclease recognition motif, (ii) 4- to 20-bp target site duplications (TSDs) flanking each SVA insertion, (iii) poly(A) tails of varying lengths at their 3'-ends, (iv) presence of 5'-truncated SVA insertions, (v) internal rearrangements and inversions (13–15) and (vi) 3'-transductions (16,17). SVA is a composite retrotransposon and present in about 2700 copies (12) in the human genome reference sequence. Approximately 30% of them have been integrated after the divergence of humans and chimpanzees (18). The origin of SVA elements can be traced back to the beginnings of hominid primate evolution, ~18–25 Mya. Starting at the 5'-end, a full-length SVA element is composed of a (CCCTCT)_n hexamer repeat region; an Alu-like region consisting of two antisense Alu fragments and an intervening unique sequence; a variable number of tandem repeats (VNTR) region, which is made up of copies of a 36- to 42-bp sequence or of a 49- to 51-bp sequence (13), presumably derived from the SVA2 element found in Rhesus macaques and humans (19–22); and a short interspersed element of retroviral origin (SINE-R) region (22). A poly(A) tail is positioned downstream of the predicted conserved polyadenylation signal AATAAA (13). Seven different SVA variants exist in hominid genomes, including 5'- or 3'-transductions, or both 5' and 3' transductions (21,23). The in vivo retrotransposition rate of the SVA retrotransposon family was recently estimated to be one in 916 births (25).

Several observations indicate that SVA elements constitute a highly active family of hominid-specific non-LTR retrotransposons whose mobilization rate exceeds processed pseudogene formation frequencies: First, seven SVA insertions were found to be associated with disease [for review see (24)] suggesting that SVA mobilization in vivo is currently more efficient than the formation of high-copy pseudogenes, which have not been found to be associated with any human disease so far. Second, the identified disease-causing SVA insertions are derived from different source elements of the SVA subfamilies D, E, F and F1, and SVAs from these subfamilies are polymorphic in humans (11,12,21). It was estimated that ~40% of the SVA elements in the human genome are polymorphic (11). Also, 14 SVA insertions were recently identified in the HuRef sequence that are not present in the haploid human genome reference sequence from the HGP (25). Lastly, each mRNA pseudogene originates from primarily one source locus, while retrotransposed SVAs are derived from multiple SVA source loci (16,21).

We set out to test the hypothesis that the L1 protein machinery is mobilizing SVA elements in trans by establishing an SVA retrotransposition reporter assay in cells. We compared the rate of processed pseudogene formation with the trans-mobilization frequencies of two human-specific SVA elements that were identified as potential source elements. We found that SVA RNAs transcribed from the retrotransposition reporter plasmids are trans-mobilized 12- to 300-fold more efficiently than RNA-Pol II transcripts expressed from a pseudogene formation reporter plasmid in HeLa-HA cells. Furthermore, we demonstrate that the hexameric (CCCTCT)_n repeat/Alu-like region at the 5'-end of canonical SVA elements and the 3'-transduced AluSp sequence of an SVA source element have different effects on SVA retrotransposition frequencies in cell culture. Marked SVA de novo insertions were predominantly full-length and exhibited all structural features of L1-mediated retrotransposition that are observed in pre-existing genomic SVA insertions.

MATERIALS AND METHODS

Isolation of genomic SVA_E and SVA_F1 elements

Based on two recent publications (26,21), we selected two potentially functional human-specific SVA elements as retrotransposition reporter elements. In order to isolate the member of the SVA subfamily E (SVA_E) that served as source element of a reported retrotransposition event into the LDLRAP1 gene (26), we performed a BLAT search of the human genome reference sequence (hg17; May 2004), using the partially published sequence of this disease-causing SVA insertion as a query to identify its potential source element. We observed 100% identity between query sequence and genomic SVA_E element H19_27 (21). Since the human genome is polymorphic for this SVA insertion, we amplified this SVA element from a BAC clone (RP11-420K14 [AC092364] obtained from the Roswell Park Cancer Institute (Buffalo, NY, USA) via the RZPD-Deutsches Ressourcenzentrum für Genomforschung) by PCR using primers chrom19-gen-FW and chrom19-gen-REV (Supplementary Table S1) which are specific for the genomic sequences flanking SVA H19_27. PCRs were performed using the Expand Long Template PCR System (Roche) and controlled by sequence analyses. The resulting ~2-kb PCR product was subcloned into the pGEM-T Easy vector (Promega). Subsequent sequence analysis of the cloned PCR product revealed that the amplified SVA element harbors only 21 VNTR subunits instead of 33 VNTR subunits specified in the human genome reference sequence. This finding was confirmed by sequence analysis of three independent PCR amplifications of SVA H19_27 located on the BAC clone performed with three different primer pairs. We conclude that the nucleotide sequence of the VNTR region of SVA H19_27 on the BAC clone differs from the corresponding sequence of the human genome reference sequence. To isolate the SVA element H10_1 (21) which is a source element of the SVA subfamily F1 (SVA_F1) together with its 3'-flanking AluSp sequence, we amplified the corresponding genomic fragment from the BAC clone RPCI753F0114Q [AL392107] (ImaGenes) by PCR with primers H10_1 For1 and H10_1 Rev1 (Supplementary Table S1). The resulting 4291-bp PCR product was subcloned into the pTZ57R vector (Fermentas) leading to pTZ H10_1. The subcloned PCR product was verified by sequence analyses.
Retrotransposition reporter constructs and mutant L1 protein donor plasmids

*pCEPneo*. The *mneoI* indicator cassette of pJM101/L1.3 (61) was PCR amplified using the primers Neo-Nhe-FW and Neo-BamHI-REV (Supplementary Table S1), and subcloned into pGEM-T Easy. Next, the *mneoI* cassette was removed by NheI/BamHI digestion and inserted into pCEP4 (Invitrogen) downstream of the CMV promoter (CMV<sub>p</sub>) and in opposite transcriptional orientation relative to the CMV promoter (Figure 1A).

*pAD3/SVA<sub>E</sub>*. The subcloned SVA H19_27 element was reamplified by PCR using primers CT-up-KpnI and DOWN-Δ-pA (Supplementary Table S1). The resulting fragment is devoid of the SINE-R-encoded polyadenylation signal and was inserted between CMVP and the *mneoI* indicator cassette of pCEPneo via KpnI/NheI. CMVP-driven SVA transcription in pAD3/SVA<sub>E</sub> reads into the *mneo* indicator cassette to be terminated by the pCEP4-encoded SV40 polyadenylation signal.

*pAD4/SVA<sub>E</sub>*. pAD4/SVA<sub>E</sub> differs from pAD3/SVA<sub>E</sub> in the absence of the initial 499 nt of the 5′-end of SVA H19_27 covering the (CCCTCT)<sub>n</sub> hexameric repeats and the Alu-like region. To generate pAD4/SVA<sub>E</sub>, the reamplified SVA H19_27 fragment was digested with AlwNI and NheI. The resulting 1380-bp fragment was cloned between the CMVP and the *mneoI* indicator cassette of pCEPneo via KpnI/blunt/AlwNI and NheI (Figure 1A).

*pSC3/SVA<sub>F1</sub>*. To remove the transcriptional termination signal at the SINE-R 3′-end of SVA<sub>F1</sub> source element H10_1 for subsequent cloning steps, the H10_1 sequence was reamplified from pTZ H10_1 using the primer pair H10_1 For2 and H10_1 Rev2 including KpnI and NheI restriction sites, respectively (Supplementary Table S1). The resulting 3527-bp PCR product was inserted into pGEM-T Easy, excised from the resulting plasmid pGEM H10_1 by KpnI/NheI digestion, and inserted between the CMVP and the *mneoI* indicator cassette of pCEPneo via KpnI/NheI, yielding pSC3/SVA<sub>F1</sub>.

*pSC4/SVA<sub>F1</sub>*. The synthetic oligonucleotide sequence A<sub>14</sub>TTTA<sub>26</sub> (Generi Biotech) was fused as NheI/SpeI fragment into the NheI site of the SINE-R 3′-end of pGEM H10_1, substituting for the AATAAA-containing polyA tail at the 3′-end of the genomic SVA H10_1 element. The resulting plasmid pGEM-H10_1/ΔpA<sub>A</sub>. AluSp was amplified from pTZ H10_1 using oligonucleotides H10_1 AluFor2 and H10_1 AluRev2 including SpeI and SalI restriction sites, respectively (Supplementary Table S1), and subcloned into pGEM-T Easy. The 389-bp SpeI/SalI AluSp fragment was fused with the 3′-end of the A<sub>14</sub>TTTA<sub>26</sub> stretch of pGEM-H10_1/ΔpA<sub>A</sub> via SpeI/SalI resulting in pGEM-H10_1+Alu. The 3904-bp H10_1/A<sub>14</sub>TTTA<sub>26</sub>/AluSp-fragment was inserted between CMVP and *mneoI* cassette of pCEPneo via KpnI and NheI/SalI blunt, yielding pSC4/SVA<sub>F1</sub>.

*pJM101/L1<sub>Rp</sub>nΔneoΔORF1* (*L1<sub>Rp</sub>nΔORF1*). pJM101/L1<sub>Rp</sub>nΔneoΔORF1 was generated by introducing a 330-bp in-frame deletion in L1 ORF1 of pJM101/L1<sub>Rp</sub>nΔneo (2). The deletion was accomplished by XhoI/SapI-restriction of pJM101/L1<sub>Rp</sub>nΔneo and subsequent religation after blunting.

Cell culture, SVA retrotransposition reporter assays and statistical analyses

Cell lines HeLa-JVM and HeLa-HA (28) were cultured in DMEM High Glucose (Biochrom AG, Berlin, Germany) supplemented with 10% FCS (Biowest, Nuaillé, France), 100 μg/ml streptomycin and 100 U/ml penicillin. To perform retrotransposition reporter assays with initial hygromycin selection for the presence of retrotransposition reporter plasmid and L1 protein donor plasmid, 1.8×10<sup>6</sup> cells were plated on 10-cm dishes or T75-flasks. Plated cells were cotransfected with 3 μg of an SVA retrotransposition reporter plasmid or pCEPneo and 3 μg of an L1 protein donor construct (pJM101/L1<sub>Rp</sub>nΔneo, pJM101/L1<sub>Rp</sub>nΔneoΔORF1) or pCEP4 (negative control) using FUGENE 6 (Roche) according to the manufacturer’s instructions. Each cotransfection was performed twice or thrice in triplicate: In each case, three cotransfections were used to quantify retrotransposition rates of the SVA reporter elements and pseudogene formation rates of the pCEPneo construct. The fourth cotransfection was used to isolate cell lysates and total RNA in order to analyze expression of retrotransposition reporter cassettes and L1-specific gene products expressed from the L1 donor plasmids. Starting 24 h post-transfection, cells were subjected to hygromycin (200 μg/ml, Invitrogen) selection for 14 days. After trypsinization and reseeding, cells were selected for L1-mediated retrotransposition events in medium containing 400 μg/ml G418 (Invitrogen). After 11-12 days of selection, G418<sup>+</sup> colonies were either fixed and stained with Giemsa (Merck) to quantify retrotransposition events as described previously (27), or individual G418<sup>+</sup> colonies were isolated and expanded to characterize SVA de novo retrotransposition events.

To perform retrotransposition reporter assays after transient cotransfection of the respective reporter plasmid and L1 donor plasmid (without hygromycin selection), 2.8×10<sup>6</sup> cells were plated on 15-cm dishes, and transiently cotransfected with 8 μg of the reporter plasmids pAD3/SVA<sub>E</sub>, pAD4/SVA<sub>E</sub> or pCEPneo and 8 μg of either pJM101/L1<sub>Rp</sub>nΔneo or pCEP4 using FUGENE 6. Each cotransfection was performed four times in parallel. HeLa cells resulting from one transient cotransfection experiment each were used to analyze the expression of ORF1p encoded by pJM101/L1<sub>Rp</sub>nΔneo. Cells were trypsinized 2 days after transfection, re-seeded, and cultivated in G418 containing medium for 14 days. The *cis*-retrotransposition rate which was observed after cotransfection of pJM101/L1<sub>Rp</sub>n and pCEP4 served as positive control and was set as 100%. To obtain countable results for retrotransposition in *cis*, only 1×10<sup>4</sup> cells were plated for G418 selection. Transfection efficiency was determined by cotransfecting...
Figure 1. Rationale of the SVA trans-mobilization assay. (A) Schematic representation of SVA retrotransposition reporter plasmids, and pseudogene-formation control construct pCEPneo. SVA reporter elements in pAD3/SVA_E, pAD4/SVA_E, pSC3/SVA_F1 and pSC4/SVA_F1 and the processed pseudogene formation cassette in pCEPneo were each tagged with the indicator gene \textit{mneoI}, and set under transcriptional control of the human CMV promoter (CMVP). Splice donor (SD) and splice acceptor (SA) sites of the oppositely oriented \( g \)-globin intron are indicated. \textit{mneoI} is flanked by a heterologous promoter (P') and a polyadenylation signal (A'). Transcripts originating from CMVP driving SVA \textit{mneoI} or CEP \textit{mneoI} transcription can splice the intron, but contain an antisense copy of the \textit{neo} gene. G418 resistant (G418R) colonies arise only if this transcript is reverse transcribed, integrated into chromosomal DNA, and expressed from its own promoter P'. Each SVA reporter element was inserted between CMVP and the \textit{mneoI} cassette. pAD4/SVA_E differs from pAD3/SVA_E in the deletion of the initial 498 nt of the SVA 5'-end covering (CCCTCT)\textsubscript{n}- and \textit{Alu}-like region. pSC4/SVA_F1 varies from pSC3/SVA_F1 in the insertion of the 389-bp \textit{AluSp} element fused to the A\textsubscript{14}TTTA\textsubscript{26} stretch between (continued)
4 μg of pEGFP-N1 (Clontech), 4 μg of pAD3/SVAE, pAD4/SVAE or pCEPneo and 8 μg of the L1 donor plasmid into 2.8 x 10^6 HeLa-JVM cells using FUGENE 6. EGFP expressing cells were counted 24 h post-transfection by flow cytometry. The percentage of green fluorescent cells was used to determine the transfection efficiency of each sample (2,29). Statistical evaluation was performed by means of an Analysis of Variance (ANOVA). To control the overall type I error α = 0.05, P-values were adjusted according to Dunnett for multiple comparisons. The statistical analysis was performed with SAS®/STAT software, version 9.2 SAS system for Windows.

**Quantitative real-time RT-PCR**

Fourteen days after cotransfection of HeLa cells with pJM101/L1mneo and either SVA retrotransposition reporter plasmid or pCEPneo, total RNA was extracted from hygromycin-selected HeLa cells using TRIZOL® (Invitrogen) following the manufacturer’s instructions. One microgram of total RNA was incubated with 2 μl of RNase-free DNaseI (Invitrogen) for 30 min at room temperature. Using the SuperScript III™ First-Strand Synthesis System for RT–PCR (Invitrogen) in combination with an oligo(dT)16–18 primer, first-strand cDNA was synthesized from 0.5 μg of DNaseI-digested, total RNA according to the manufacturer’s instructions. To quantify levels of spliced transcripts expressed from the mneoI-tagged reporter elements in pAD3/SVAE, pAD4/SVAE, pSC3/SVAF1, pSC4/SCAF1 and pCEPneo, real-time PCR was performed in triplicate applying TaqMan™ chemistry (Applied Biosystems) in an Applied Biosystems 7900HT Fast Real-Time PCR System base unit. We used a primer/probe combination (Neofor: 5'-GCTATTCGGCTATGACTGG-3', Neorev: 5'-GCGCA CGATAGCGCGCTGC-3'; probe: 5'-FAM-CCTCGTC GTAAGCTTCC-3') specifically recognizing the spliced mneoI cassette. For normalization, eukaryotic 18s rRNA was used as internal control. Cycling conditions were as follows: 95°C for 15 min (initial cycle), 95°C for 15 s and 60°C for 1 min (40 cycles). The software applied to analyze real-time and end point fluorescence was RQ manager 1.2. Relative quantification of RNA expression was carried out using the ΔΔCt method (30).

**Immunoblot analysis**

To assess L1 ORF1p and L1 ORF2p expression, HeLa cells were cotransfected as described above and harvested 14 days (with hygromycin selection) or 2 days later (without hygromycin selection), respectively. Cells were lysed in RIPA buffer (25 mM Tris, pH 8, 137 mM NaCl, 1% glycerol, 0.5% sodium deoxycholate, 1% Nonidet P-40, 2 mM EDTA, pH 8, 0.1% SDS and protease inhibitors), and lysates were cleared by centrifugation. In the case of transient cotransfection of SVA reporter plasmid and L1 protein donor, 2.8 x 10^6 cells were plated on 15-cm dishes and transiently cotransfected as described above. Cells of one 15-cm dish carrying the respective plasmids were trypsinized 2 days after cotransfection and lysed with RIPA lysis buffer. Twenty micrograms of each protein lysate were boiled in Laemmli buffer, loaded on 12% polyacrylamide gels, subjected to SDS–PAGE, and electroblotted onto nitrocellulose membranes. After protein transfer, membranes were blocked for 2 h at room temperature in a 10% solution of non-fat milk powder in 1× PBS-T [137 mM NaCl, 3 mM KCl, 16.5 mM Na2HPO4, 1.5 mM KH2PO4, 0.05 % Tween 20 (Sigma)], washed in 1× PBS-T, and incubated overnight with the respective primary antibody at 4°C. To detect L1 ORF1p, the polyclonal rabbit-anti-L1ORF1p antibody #984 (see Supplementary Data) was used in a 1:2000 dilution in 1× PBS-T containing 5% milk powder. L1 ORF2p expression was verified using a rabbit anti-ORF2p-N antibody (31) at a 1:1000 dilution in 1× PBS, 5% milk powder, 0.05 % Tween 20. Membranes were washed thrice in 1× PBS-T and incubated with an HRP-conjugated, secondary anti-rabbit IgG antibody (Amersham Biosciences) at a dilution of 1:30 000 in 1× PBS-T/5% milk powder for 2 h. Subsequently, the membrane was washed three times for 10 min in 1× PBS-T. β-actin and α-tubulin expression were detected using a monoclonal anti-β-actin antibody (clone AC-74, Sigma-Aldrich) and a polyclonal anti-α-tubulin antibody (ab4074, Abcam) as primary antibodies at dilutions of 1:30 000 and 1:10 000, respectively. Anti-mouse HRP-linked species-specific antibody (from sheep) at a dilution of 1:10 000 and anti-rabbit HRP-linked specific antibody at a dilution of 1:5 000 served as secondary antibodies specific for anti-β-actin and anti-α-tubulin antibody, respectively. Immunocomplexes were visualized using lumino-based ECL immunoblot reagent (Amersham Biosciences).

**Analysis of SVA de novo insertions**

Genomic DNA from expanded G418R-HeLa colonies was isolated applying the Qiagen DNeasy® Tissue Kit according to the manufacturer’s protocol. To test for the...
presence of the spliced mneol indicator cassette, a diagnostic PCR was performed using the intron-flanking primer pair GS86/GS87 (Figure 4 and Supplementary Table S1). PCR cycling conditions were as follows: 3 min at 96°C, (30 s at 96°C, 15 s at 56°C; 2 min at 72°C) 25 cycles, 7 min at 72°C. To determine genomic pre- and post-integration sites of SVA de novo insertions, we used a modified version (32) of a previously published extension primer tag selection preceding solid-phase ligation-mediated PCR (EPTS/LM-PCR) (33) to isolate 5' junctions of these insertions. Products of the final PCR were separated in a 1% agarose gel, isolated from the gel using the QIAquick Gel Extraction Kit (Qiagen), and sequenced either directly or after subcloning into pGEM-T Easy. Obtained sequences were mapped to the human genome using the UCSC genome browser at http://genome.ucsc.edu. To characterize 5'-junctions of each SVA de novo insertion, primers specific for the genomic sequence adjacent to the 5'-end of the de novo insertions were designed. The second PCR primer used was GS87, GS88 or, alternatively, HERV-K REV because these primers bind specifically to the retrotransposed SVAE reporter cassette. All oligonucleotides used in this study are listed in Supplementary Table S1. Genomic pre-integration sites and surrounding sequences were characterized using the UCSC genome browser annotation for genes, the Repeatmasker and G+C content tracks. For localization of the integration sites to particular isochores/G+C-content regions the table published by Costantini et al. (34) was used.

Sequence logos
To display patterns of sequence conservation between genomic target sequences of pre-existing SVA F/SVA F1, L1-Ta and AluYa5 elements and isolated SVA F and SVA E de novo retrotransposition events, sequence logos were generated (Figure 6) applying the program WebLogo (35, http://weblogo.berkeley.edu/logo.cgi). We randomly picked 70 genomic target sequences of preexisting members from each of the retrotransposon subfamilies L1Hs-Ta, AluYa5 and SVA E or SVA F. Genomic target sequences of L1Hs-Ta and AluYa5 elements (Supplementary Table S3) were identified from databases (L1_selection.xls; L1_TSD.txt; L1_coord_seq.txt; http://batzerlab.lsu.edu) published recently (36,37). Genomic target sequences of 70 preexisting members of the SVA subfamilies E and F (Supplementary Table S3) were determined from a recently published list of human endogenous SVA insertions (21).

RESULTS
Identification, isolation and engineering of functional human-specific SVA reporter elements
In order to test the hypothesis that SVA elements are mobilized by the L1-encoded protein machinery, we set out to establish a trans-mobilization assay in which potentially functional marked SVA elements are tested for retrotransposition in the presence of the overexpressed L1 protein machinery. As a first step we identified human-specific, genomic SVA elements that were likely to be retrotransposition-competent (RC). Since sequence and/or structural characteristics of RC SVA source elements were unknown and RC SVA source elements had not been identified when we started our study, we firstly selected SVA insertions reported earlier to be the cause of single cases of genetic disorders. These SVAs were caught red-handed after they were launched from RC source elements and did not have time to accumulate disfiguring mutations in the human genome. We picked the sequence of an SVA E insertion into the LDLRAP1 gene that caused a case of autosomal recessive hypercholesterolemia (ARH) (26), as query, and performed a BLAT search of the human genome reference sequence (hg17; May 2004) to identify the potential source element of the disease-causing insertion. We observed 100% identity between the query sequence and genomic SVA E element H19_27 (Supplementary Figure S1) which displays presence/absence polymorphism in the human genome (21). Given the sequence correlation between the ARH-causing SVA insertion and SVA H19_27, and the polymorphic state of H19_27, we concluded that either H19_27 is the source element of the disease-causing SVA insertion, or both H19_27 and the ARH-causing insertion are derived from the same source element which is not present in the analyzed human genome reference sequence. Therefore, we chose the SVA E element H19_27 as a reporter element for the planned SVA retrotransposition reporter assays. To generate the SVA retrotransposition reporter plasmid pAD3/SVA E, the genomic canonical SVA E element H19_27 was amplified by PCR, tagged at its 3'-end with the mneol indicator cassette (38), and inserted into the episomal pCEP4 expression vector, where the SVA reporter cassette was set under the control of the human CMV promoter (Figure 1A; see ‘Materials and Methods’ section). To address the question if the (CCCTCT) n repeats and/or the 300-bp Alu-like region at the 5'-end of SVA elements play a role in the efficiency of any potential trans-mobilization of SVA elements, we deleted 498 nt of the 5'-end sequence of the SVA element in pAD3/SVA E to generate pAD4/SVA E (Figures 1 and Supplementary Figure S1).

We chose the genomic SVA F1 element H10_1 as second retrotransposition reporter, because it has been identified recently as source element of at least 13 SVA F1 subfamily members (21,23), including one SVA F1 element that is the progenitor of a disease-associated insertion into the HLA-A gene (39). The human-specific SVA F1 subfamily was generated by the acquisition of a disease-associated insertion into the HLA-A gene (39). The human-specific SVA F1 subfamily was generated by the acquisition of a disease-associated insertion into the HLA-A gene (39). The human-specific SVA F1 subfamily was generated by the acquisition of a disease-associated insertion into the HLA-A gene (39).
were retrotransposed, leading to at least 13 genomic 3'-transduced SVA\text{F1} insertions (21).

Since there is evidence that both the H10.1 element alone and H10.1 derivatives including the 3'-transduced AluSp sequence have served as source elements for SVA retrotransposition (22,23), we tested both SVA\text{F1} members (Figure 1A; pSC3/SVA\text{F1}, pSC4/SVA\text{F1}) in our trans-mobilization assay.

For that purpose, we first fused the entire 3.5-kb H10.1 element at its SINE-R 3'-end with the mneoI indicator cassette in opposite orientation, and inserted the marked SVA\text{F1} reporter element into the pCEP4 expression vector, generating pSC3/SVA\text{F1} (Figure 1A). Next, we fused the genomic AluSp element to the 3'-end of SVA H10.1, which resulted in a structure that also constitutes the genomic situation. The fusion was tagged with the mneoI indicator cassette and the entire 6.2-kb retrotransposition reporter cassette was inserted into pCEP4, resulting in pSC4/SVA\text{F1} (Figure 1A). To ensure that the indicator cassette is transcribed as efficiently as the upstream SVA sequences, transcriptional termination signals located at the SINE-R 3'-end and in the sequence downstream of the AluSp 3' TSD were removed (ΔAATAAA; Figure 1A). Since pSC3/SVA\text{F1} and pSC4/SVA\text{F1} differ exclusively in the presence of an intact AluSp element at the SINE-R 3'-end, we will be able to evaluate if this element plays any role in SVA trans-mobilization efficiency.

L1-encoded proteins facilitate retrotransposition of both SVA\text{E} reporter elements

To evaluate if the SVA\text{E} reporter elements are trans-mobilized by the L1-encoded protein machinery, we cotransfected pAD3/SVA\text{E} and pAD4/SVA\text{E} with the L1 protein donor plasmid pJM101/L1\text{RP}\Delta\text{neo} or pCEP4 into HeLa-JVM cells (Figure 1B) because it was demonstrated that HeLa cells can efficiently accommodate and express proteins from different expression vectors (2). Cotransfected cells were subjected to hygromycin selection for the presence of the plasmids (Figure 1B) and subsequently selected for G418 resistance. Each SVA reporter element was tagged with an antisense copy of the selectable marker gene neo encoding neomycin phosphotransferase, a heterologous promoter (P\text{'}), and a polyadenylation signal (A\text{'}-end) (Figure 1A). This arrangement ensures that G418-resistant cells (G418\text{R}) will only arise if a transcript initiated from the human CMV promoter (CMVP) driving SVA\text{mneoI} or CEP\text{mneoI} expression is spliced, reverse transcribed, reintegrated into chromosomal DNA, and expressed from promoter P\text{'}.

G418\text{R} foci indicating retrotransposition of an SVA reporter cassette, could be observed only if the L1 expression plasmid was cotransfected suggesting that L1 proteins are required for SVA mobilization (Figures 2A and Supplementary Figure S3).

Next, we asked if L1 proteins prefer our reporter SVA mRNA to any other PolIII transcript as substrate for trans-mobilization. If this was the case, trans-mobilization frequency of the SVA reporter element would be expected to exceed the frequency of processed pseudogene formation of a regular PolII gene. In order to determine the pseudogene formation rate, we generated the reporter plasmid pCEP\text{neo} which differs from pAD3/SVA\text{E} exclusively by the absence of the 1876-bp SVA sequence (Figure 1A). Since transcripts expressed from the CMVP of pCEP\text{neo} consist exclusively of the mneoI indicator cassette in antisense orientation and do not include any retrotransposon sequences, these transcripts should be trans-mobilized as frequent as random mRNAs encoded by host PolII genes. Overall, trans-mobilization frequencies of the pAD3/SVA\text{E}-encoded canonical SVA element exceeded pseudogene formation of the reverse mneoI cassette by 2- to 5-fold whereas the 5'-truncated SVA\text{E} cassette encoded by pAD4/SVA\text{E} outnumbered the pseudogene formation rate by only 1- to 3-fold (Figures 2A and Supplementary Figure S3). While the differences in trans-mobilization rates between pAD3/SVA\text{E} and pCEP\text{neo} were statistically significant (p1 = 0.0001), those between pAD4/SVA\text{E} and pCEP\text{neo} were not (p2 = 0.2212). To verify that the observed differences in trans-mobilization rates between pAD3/SVA\text{E} and pAD4/SVA\text{E} are not resulting from discrepancies in mRNA production or stability between the different SVA reporter constructs, we tested for the presence of similar amounts of spliced, tagged SVA mRNAs by quantitative real-time RT-PCR (qRT-PCR) (Figure 2B). Using primer/probe combinations specifically recognizing the spliced mneoI reporter cassette, we quantified the relative amounts of spliced mRNA expressed from the reporter plasmids after 14 days of hygromycin selection (Figure 2B).

Clearly, the observed differences in the amounts of spliced SVA reporter mRNA of a 1.1-fold between pAD3/SVA\text{E} and pAD4/SVA\text{E}-transfected cells (Figure 2B) is only minor compared to the 2- to 5-fold differences in trans-mobilization rates (Figure 2A). Therefore, we draw the conclusion that the 498-nt fragment deleted from the 5'-end of the canonical SVA element in pAD3/SVA\text{E} makes the 5'-truncated element in pAD4/SVA\text{E} a somewhat less attractive substrate for trans-mobilization.

Interestingly, although the amount of spliced pCEP\text{neo} transcripts exceeds those derived from transcription of the SVA\text{E} reporter cassettes, trans-mobilization of SVA\text{E} transcripts is still more efficient, emphasizing that SVA\text{E} RNA is a preferred substrate for the L1 protein machinery. In order to evaluate if the observed differences in trans-mobilization rates can be attributed to varying L1 protein levels, we assessed L1 ORF1p and ORF2p expression in a parallel set of cotransfected hygR-selected HeLa cell cultures (Figure 2C). Immunoblot analysis of cell extracts isolated the day before the onset of G418 selection of the remaining cotransfected cultures, shows that comparable amounts of ORF2p are expressed in the differently cotransfected cells. Although L1 ORF1p levels are elevated in HeLa cells cotransfected with pCEP\text{neo} relative to pAD3/SVA\text{E} and pAD4/SVA\text{E}-cotransfected cells, the observed trans-mobilization rate is the highest in cells cotransfected with pAD3/SVA\text{E}. This indicates that the observed differences between pseudogene formation rate and SVA\text{E} trans-mobilization frequencies did not result from diverse L1 protein levels.
Figure 2. Trans-mobilization of mneoI-tagged SVA_E reporter elements. (A) SVA_E retrotransposition reporter assay after hygromycin selection for the presence of expression plasmids. SVA_E reporter plasmids pAD3/SVA_E, pAD4/SVA_E, or pCEPneo were cotransfected with the L1 protein donor pJM101/L1RPΔneo (L1RP) or the empty vector pCEP4. After hygR selection, G418R selection for retrotransposition events followed and retrotransposition rates were determined by counting G418R HeLa colonies. Each cotransfection experiment and subsequent retrotransposition reporter assay was carried out three times in triplicates. Retrotransposition frequencies per 10^6 cells are listed and relative retrotransposition rates are indicated as bar diagram. Cis retrotransposition rate of the L1 reporter element pJM101/L1RP was set as 100%. Each bar depicts the arithmetic mean ± SD of the relative retrotransposition rates obtained from nine individual cotransfection experiments (n = 9). (B) qRT-PCR analyses to quantify the relative amounts of spliced transcripts expressed from retrotransposition reporter cassettes. Total RNA was isolated 48 h after cotransfection of pCEPneo, pAD3/SVA_E, and pAD4/SVA_E with the L1 protein donor plasmid pJM101/L1RPΔneo. The used primer/probe combination (see ‘Materials and Methods’ section) is specific for the spliced mneoI-cassette (black box with arrow). Relative amounts of mRNA expression refer to the signal obtained from total RNA of untransfected HeLa cells which was set as 1 (C1); Total RNA from mock-transfected HeLa cells served as negative control. (C) Immunoblot analysis of L1 protein expression in HeLa cells after cotransfection of the L1 protein donor (L1RP) with retrotransposition reporter plasmids pAD3/SVA_E, pAD4/SVA_E and pCEPneo. Whole-cell lysates were prepared 14 days after cotransfection upon completion of hygromycin selection and subjected to immunoblot analysis using antibodies against either L1 ORF1p (αORF1p) or L1 ORF2p (αORF2p). An amount of 20 μg of whole-cell extracts were loaded per lane. α-tubulin protein levels (~50 kDa) were analyzed as loading control. Lysates from untransfected HeLa cells and from the germ cell tumor cell line NTera-2 served as negative and positive control for L1 protein detection, respectively. (D) SVA_E trans-mobilization assay after transient cotransfection of expression plasmids. pAD3/SVA_E, pAD4/SVA_E or pCEPneo were transiently cotransfected with pJM101/L1RPΔneo (L1RP) or pCEP4 into HeLa cells. Two days later, cells were G418-selected for de novo retrotransposition events for 14 days. G418R HeLa colonies were Giemsa-stained and counted. Each cotransfection experiment was done in quadruplicate. Subsequent retrotransposition reporter assays were performed in triplicate. Retrotransposition frequencies per 10^6 cells are listed and relative retrotransposition rates are indicated as bar diagram. Cis retrotransposition rate of L1 reporter element pJM101/L1RP was set as 100%. Each bar depicts the arithmetic mean ± SD of the relative retrotransposition rates obtained from three individual cotransfection experiments (n = 3).
Next, we wanted to evaluate if we can confirm these results on trans-mobilization of the SVA_E reporter elements in an experimental setup, in which we transiently cotransfect HeLa cells with retrotransposition reporter and L1 protein donor plasmid, and select for retrotransposition events 2 days later (Figure 1B, approach 2). In compliance with the results obtained in the case of approach 1, we observed in this transient-cotransfection setup trans-mobilization rates of the SVA reporter cassette in pAD3/SVA_E and pAD4/SVA_E which exceeded pseudogene formation rates by 2- to 5-fold (arithmetic mean: 3.5-fold) and 1- to 3-fold (arithmetic mean: 2.4-fold), respectively (Figure 2D). Again, the difference in trans-mobilization rates between pAD3/SVA_E and pCEPneo was statistically significant ($p_3 = 0.0021$), whereas differences between pAD4/SVA_E and pCEPneo were not ($p_2 = 0.0354$). To test for comparable expression levels of the L1 protein machinery, we assessed L1 ORF1p expression in a parallel set of transiently cotransfected HeLa cultures. Immunoblot analysis of cell lysates harvested three days after cotransfection uncovered that there were no detectable differences in L1 ORF1p expression levels between the differently cotransfected cells (Supplementary Figure S5). This confirms that also in the transient-cotransfection setting, the observed discrepancies in trans-mobilization rates (Figure 2D and Supplementary Figure S4) are not a consequence of varying expression levels of the L1 protein machinery.

The 3′-transduced AluSp element increases the trans-mobilization rate of the SVA_F1 source element by ~25-fold

In order to evaluate if the canonical SVA_E reporter element differs in its trans-mobilization rate from SVA_F elements that were reported to be source elements of the highly successful human-specific SVA_F1 subfamily, we tested pAD3/SVA_E, pSC3/SVA_F1 and pSC4/SVA_F1 (Figure 1) in our trans-mobilization assay in parallel. Cotransfection of pAD3/SVA_E, pSC3/SVA_F1 and pCEPneo with the L1 protein donor plasmid pJM101/L1RPΔneo into HeLa-HA cells uncovered that SVA_E reporter and SVA_F1 source element H10_1 are trans-mobilized by ~18 and ~12-fold relative to pseudogene formation rate, respectively (Figure 3A and B). Surprisingly, the mobilization rate of the pSC4/SVA_F1-encoded SVA_F1 source element that includes the 3′-transduced AluSp sequence was exceeding pseudogene formation rate by ~300-fold. This corresponds with a relative retrotransposition rate of ~11% of L1 cis activity (Figure 3A and B). Data suggest that the acquired AluSp sequence makes the SVA element a significantly more attractive substrate for the L1 protein machinery resulting in a mobilization rate that corresponds to AluY reporter elements (3).

SVA trans-mobilization requires both L1 ORF1p and L1 ORF2p

While it is obvious that the generation of G418R foci after expression of the L1 protein machinery requires ORF2-encoded RT activity, it was unclear if L1 ORF1p is essential for SVA trans-mobilization. To uncover, if L1 ORF1p is required for trans-mobilization of SVA elements, we generated a second L1 protein donor plasmid, termed pJM101/L1RPΔneoΔORF1p, which differs from pJM101/L1RPΔneo exclusively in a 330-bp in-frame deletion in L1 ORF1 (Figure 1B). The in-frame deletion of pJM101/L1RPΔneoΔORF1p ensures that an ORF1p mutant which lacks amino acid positions 99–208 is expressed, and that initiation of ORF2 translation within the bicistronic RNA is not perturbed. The deletion comprises the C-terminal half of the coiled coil (cc) domain and the N-terminal half of the RNA-recognition motif (RRM) domain of ORF1p (40). We confirmed by immunoblot analysis that both L1 protein donor plasmids facilitated expression of similar amounts of ORF2p after cotransfection with each of the three SVA reporter plasmids (Figures 3C). Coexpression of ORF1p deletion mutant and functional ORF2p from pJM101/L1RPΔneoΔORF1p did not result in trans-mobilization of the SVA reporter elements encoded by pAD3/SVA_E and pSC3/SVA_F1 indicating that an intact ORF1p is essential for SVA retrotransposition. Surprisingly, in the absence of both overexpressed L1 proteins or ORF1p alone, trans-mobilization of the SVA_F1 reporter element of pSC4/SVA_F1 was still exceeding pseudogene formation by 2- to 3-fold (Figure 3A). Since the 3′-transduced AluSp sequence obviously makes the SVA RNA a preferred substrate for L1 proteins, mobilization in the absence of overexpressed L1 proteins could be explained by trans-mobilization of the pSC4/SVA_F1 reporter cassette beyond pseudogene formation by the moderately expressed endogenous L1 protein machinery in HeLa cells (41,42). Quantification of spliced SVA reporter RNAs by qRT–PCR (Figure 3D) and of overexpressed L1 proteins by immunoblot analysis (Figure 3C) show that the observed differences in G418R colonies are neither a consequence of differences in the amounts of spliced retrotransposition reporter transcripts nor the result of varying amounts of overexpressed L1 proteins.

Structural features of SVA de novo integrants

To verify that the G418R HeLa colonies that resulted from SVA trans-mobilization assays were a consequence of marked SVA de novo retrotransposition events, genomic DNA was extracted from 12 randomly chosen single expanded G418R HeLa cell colonies which had resulted from cotransfection of the SVA_E reporter plasmids with pJM101/L1RPΔneo (Figure 2A). First, we carried out a diagnostic PCR on genomic DNA with primers spanning the intron of mneoI (Figure 4). The generation of a 792-bp PCR product shows that transcription, splicing, and L1-mediated reverse transcription of the SVA reporter elements as well as integration of the SVA_E-mneoI cDNA into the genome has occurred. A 1694-bp PCR product that would be specific for the unspliced indicator cassette encoded by the SVA_E retrotransposition reporter plasmids could not be detected (Figure 4). Since PCR products obtained from HeLa clone X differed from the expected pattern, this
Figure 3. L1 ORF1p is required for trans-mobilization of SVA reporter elements. (A) Trans-mobilization of SVA<sub>E</sub> and SVA<sub>F1</sub> reporter elements in presence/absence of the entire L1 protein machinery. SVA reporter plasmids pAD3/SVA<sub>E</sub>, pSC3/SVA<sub>F1</sub> and pSC4/SVA<sub>F1</sub> were each cotransfected with intact (L1<sub>Rp</sub>) and mutant (ΔORF1) L1 protein donor plasmid and pCEP4. pCEPneo was cotransfected with pJM101/L1<sub>Rp</sub>Δneo (L1<sub>Rp</sub>) or the empty vector pCEP4. After hyg<sup>B</sup> selection, cotransfected HeLa cells were G418-selected for retrotransposition events and retrotransposition rates were determined. Each cotransfection experiment and subsequent retrotransposition reporter assay was carried out in triplicates twice (n=6). Retrotransposition rates per 10<sup>6</sup> cells including standard deviations (±SD) are listed. Arithmetic means of mobilization rates relative to the cis-activity of pJM101/L1<sub>Rp</sub> (L1 cis activity) are specified and depicted as bar diagram. Error bars, ± SD; Primary data of trans-mobilization assays are summarized in Supplementary Figure S6. (B) The 3′ terminal Alu<sup>Sp</sup> sequence of SVAF1 element H10_1 increases its trans-mobilization rate by 25-fold. For reasons of clarity, a subset of the information presented in Figure 3A is displayed. Relative retrotransposition rates (L1 cis activity [%]) and retrotransposition frequencies that resulted from cotransfections with the L1 protein donor pJM101/L1<sub>Rp</sub> (L1<sub>Rp</sub> cis activity) are specified and depicted as bar diagram. Error bars, ± SD; (C) Immunoblot analysis of L1 protein expression after cotransfection of L1 protein donors with SVA retrotransposition reporter plasmids or pCEPneo. Whole-cell lysates were prepared 14 days after cotransfection upon completion of hygromycin selection and subjected to immunoblot analysis using antibodies against either L1 ORF1p (αORF1p) or L1 ORF2p (αORF2p). An amount of 70 μg of whole-cell lysates were loaded per lane. α-tubulin protein levels (~50kDa) were analyzed as loading control. Lysates from untransfected HeLa cells and from the germ cell tumor cell line 2102Ep served as negative and positive control for L1 protein detection, respectively. Shorter (exp.1) and longer exposures (exp.2) of the αORF1p immunoblot are presented to demonstrate expression of endogenous L1 ORF1p. (D) qRT–PCR analyses to quantify relative amounts of spliced transcripts encoded by the diverse retrotransposition reporter cassettes. Total RNA was isolated after 14 days of hygromycin selection following cotransfection of the reporter constructs pSC3/SVA<sub>F1</sub>, pSC4/SVA<sub>F1</sub> and pCEPneo with pJM101/L1<sub>Rp</sub>Δneo. The used primer/probe combination is specific for the spliced mneoI-cassette (Figure 2B). Relative amounts of mRNA expression refer to the signal obtained from total RNA of untransfected HeLa cells which was set as 1 (C1); Total RNA from mock-transfected HeLa cells served as negative control.
clone was excluded from further analyses. Next, we examined the structures of the genomic mneo-tagged SVAE de novo insertions for attributes of L1-mediated retrotransposition. Analysis of pre- and post-integration sites of eight pAD3/SVAE-derived and three pAD4/SVAE-derived de novo insertions isolated from G418R HeLa colonies (Figure 5) revealed that 5/11 insertions occurred into genes (Supplementary Table S2) and each insertion is indeed distinguished by structural hallmarks of L1-mediated retrotransposition such as a 30- to 78-nt poly(A) tail. The nucleotide profile of the target sites of SVAE de novo insertions resembles the L1 EN consensus target sequence 5'-TTTT/AA-3' of L1-Ta de novo insertions and preexisting L1-Ta, AluYa5 and SVA EF insertions (Figure 6). This indicates that the sequence specificity of L1 EN determines SVA integration site properties at the local level. With one exception, all characterized insertions are flanked by TSDs ranging from 8 to 19 nt (Figure 5). SVA de novo insertion 9 led to the formation of an 11-bp target site deletion, and the deleted sequence was replaced by a 5'-truncated mneo-tagged SVA retrotransposition event lacking TSDs. The formation of similar genomic target site deletions ranging from 1 bp to >11 kb associated with 5'-truncated de novo insertions has been reported for EN-dependent L1 retrotransposition events earlier (43-45). Target site deletions associated with L1-mediated retrotransposition events are believed to arise as a consequence of a second-strand cleavage event that occurred upstream of the initial first-strand cleavage (43,46).

Only 2 out of the 11 characterized SVA de novo insertions were characterized by 5'-truncations encompassing 1716 and 680 bp (insertions 6 and 9; Figure 5), respectively. The remaining nine insertions are full-length and cover 3347 (pAD3/SVAE-derived) and 2850 nt (pAD4/SVAE derived). The 5'-ends of SVA full length de novo insertions derived from both SVA reporter elements coincide with positions 3 or 4 downstream of the CMV promoter transcription initiation site (47) indicating that transcription of the SVA reporter elements is controlled by CMVP and not by any potential SVA-specific internal promoter. Full-length SVA de novo insertions include the same 44-45 nt of non-SVA sequences at their 5'-ends which result from transcriptional initiation within the CMVp region (Figure 1A).

In three cases we observed an untemplated G nucleotide between the 3'-end of the 5' TSD and the 5'-end of full-length SVA insertions. (Figure 5A; insertions 2, 5 and 8). We identified short patches of microcomple- mentarity of 1-2 nt at the 5'-genomic DNA/SVA junction of 5/11 SVA de novo insertions which is consistent with previous studies analyzing preexisting 5'-genomic DNA/L1 junctions (48,36) and 5'-junctions of L1 de novo insertions (44,45). Taken together, each of the described structural features of SVA de novo insertions have been reported for L1 de novo insertions before. This is indicating that the analyzed SVA de novo insertions are a consequence of the trans-activity of the L1 protein machinery acting on SVA transcripts.

**Target site preferences of de novo SVA integrants**

SVA de novo insertions showed a clear integration preference for TSDs or sequences flanking endogenous non-LTR retrotransposons. Five out of the 11 characterized SVA insertions occurred into Alu-TSDs or adjacent sequences (Figure 5). SVA copies inserted into an L1 3'-end (insertion 10) and within the first 100 nt of the 3'-flanking region of an L1 element (insertion 5), respectively.

We calculated the average G+C content of the genomic sequences flanking de novo SVA insertion within 5-kb windows and 30-kb windows, and found that it amounted to 44% and 44.2%, respectively (Table 1). This was significantly higher than the genome average of 41% (p<0.025; p<0.005), and is in accordance with the overall distribution of preexisting members of the SVA subfamilies E and F which were reported to have accumulated in G+C-rich regions of the human genome (12). Several scenarios have been proposed for the apparent enrichment of SVAE and SVAF elements in G+C-rich regions (12). Sequence analyses of the 5-kb and 30-kb windows also indicate that SVA de novo insertions occurred into genomic regions with an increased Alu density and a relatively poor L1 density compared to the overall genomic situation (Table 1). Distribution of SVA de novo insertions to different isochores/ G+C-content domains of the genome is consistent with the recently
Figure 5. Genomic SVA<sub>e</sub> de novo insertions display hallmarks of L1-mediated retrotransposition. Both pre- and post-integration sites are presented. Insertions 1–8 and 9–11 are derived from SVA<sub>e</sub> reporter elements encoded by pAD3/SVA<sub>e</sub> and pAD4/SVA<sub>e</sub>, respectively. Nucleotide positions at the 5′-end of each SVA insertion refer to the reference sequence of SVA H19_27 [21]; Supplementary Figure S1]. Marked full-length insertions cover ~3.3 kb (pAD3/SVA<sub>e</sub> derived) and ~2.8 kb (pAD4/SVA<sub>e</sub> derived), respectively. Identified 5′-truncated insertions comprise 1580 bp (insertion 6) and 1620 bp (insertion 9), respectively. The L1EN cleavage site on the bottom strand is indicated in blue. Extra deoxyguanylates at the 5′-ends of de novo insertions are indicated in green. CMV<sub>p</sub>-derived sequences are highlighted in yellow. Nucleotides representing patches of microcomplementarity are underlined. Black bars, Alu TSD sequences. Red lettering, SVA TSD sequences; neo, neomycin-phosphotransferase gene.
described situation of genomic SVA copies (49). Eight integrants are found in isochore H1, while the remainder insertions are located in isochore L2 (two insertions) and H2 (one insertion).

**DISCUSSION**

**L1-mediated trans-mobilization of SVA elements requires both L1 encoded proteins**

In order to gain insight into the molecular mechanisms of the mobilization of the non-autonomous SVA elements, we tested the hypothesis that SVA elements are trans-mobilized by the protein machinery encoded by functional L1 elements. We established an SVA retrotransposition reporter assay which enabled us to experimentally confirm that SVA RNA recruits the L1 protein machinery for its own mobilization. Trans-mobilization frequencies of the tested human-specific SVA reporter elements exceeded the processed pseudogene formation rate of cellular mRNAs by 2- to 5-fold in HeLa-JVM cells and by 12- to 300-fold in HeLa-HA cells. During the revision of this manuscript, Hancks and coworkers published in rough accordance with our results that, in their hands,
L1-mediated SVA retrotransposition exceeds processed pseudogene formation in HeLa-HA cells by 1- to 54-fold (50). We observed a strict requirement of L1 ORF1p and L1 ORF2p for the mobilization of the three tested SVA reporter elements in HeLa cells. While ORF1p was also shown to be essential for processed pseudogene formation (2,7), it is dispensable for Alu retrotransposition (3). The stern need for ORF1p for the mobilization of both versions of the SVAF1 source element on chromosome 10 which differ from each other exclusively in the presence/absence of the ~390 bp AluSp 3′ transduction was confirmed by Hancks et al. (50). However, while we also observed the same requirement of ORF1p for the mobilization of the canonical SVAE reporter element H19_27, the Kazazian laboratory obtained conflicting results when they analyzed the role of ORF1p in the mobilization of a canonical SVAE source element (50): Consistent with our findings, they reported that expression of an L1 driver with a double mutation in the RRM domain of ORF1p blocked trans-mobilization of the canonical SVAE source element almost entirely in HeLa-HA cells, and that coexpression of ORF2p alone with the EGFP-marked SVAE element led to barely any detectable retrotransposition events. In contrast, coexpression of ORF2p with the mneol-marked SVAE element produced more G418R foci than transfection with the intact full-length L1 driver plasmid suggesting that ORF2p alone is essential for trans-mobilization (50). It is unlikely that the ~600-bp difference in the extension of the VNTR region between the ~1.9-kb SVAE element and the ~2.5-kb SVAE element played any role in the observed differences in ORF1p requirement.

**Structural features of SVA elements affecting trans-mobilization rates**

The observation that the trans-mobilization rates of the canonical SVAE reporter element in pAD3/SVAE exceeds retropseudogene formation by 2- to 5-fold in HeLa-JVM cells and by ~18-fold in HeLa-HA cells, raises the question if potential SVA-specific structural features might qualify SVA transcripts as preferred substrates for the L1 protein machinery. First, as indicated by Alus, tRNA-derived SINEs (51), and tailless tRNAs, the ability of an RNA to localize to the ribosome determines its retrotranspositional success. After transcription, the SVA RNA needs to come in contact with L1 ORF2p, and out-compete the L1 RNA for the attention of ORF2p. L1 and Alu RNA competition for ORF2p presumably takes place at the ribosome (52,53). It has been hypothesized that the SVA-encoded Alu-like domain localizes SVA RNA to the ribosome by annealing with Alu RNAs which were suggested to be docked on ribosomes via the SRP9/14 complex (52,54). This hypothesis is consistent with our observation that the retrotransposition rate of the SVA reporter element in pAD4/SVAE which is devoid of hexameric (CCCTCT)₆ repeats and Alu-like domain, is reduced in average by 32–46% (Figure 2A and D) relative to the full-length SVA reporter element in pAD3/SVAE. A potential relevance of the Alu-like region for trans-mobilization is also supported by the finding that our pSC3/SVAF1 encoded SVA source element which also lacks the (CCCTCT)₆ repeats and almost the entire Alu-like domain exceeds pseudogene formation rate by only ~12-fold while the canonical SVAE element is trans-mobilized ~18-fold more efficiently than the pseudogene in HeLa-HA cells (Figure 3A and B). Second, the structures of preexisting SVA insertions imply that several modules of a canonical full-length SVA element are dispensable and not essential for successive rounds of retrotransposition. The existence of SVA2 elements which consist exclusively of VNTRs fused to short non-SVA sequences and display hallmarks of L1-mediated retrotransposition (20,21,55), suggests that the VNTR region is essential for SVA mobilization. This is supported by the fact that the VNTR region which can vary in length significantly among different SVA insertions, is the only module all mobilized SVA RNAs have in common. It was suggested that the VNTR alone or within the context of SVA may increase RNA stability (24).

We show that the 3′ transduction-mediated acquisition of the AluSp sequence by the SVAF1 source element H10_1 (pSC4/SVAF1) led to a 3′-transduced SVAF1 RNA which is mobilized by the L1 protein machinery ~25-fold more efficiently than the same RNA lacking the AluSp sequence (pSC3/SVAF1) (Figure 3A and B). The relatively high trans-mobilization rate of the 3′-transduced SVAF1 element H10_1 is obviously based on the observed preference of L1 proteins for SVA transcripts carrying intact AluSp sequences at their 3′-ends and is consistent with the presence of numerous genomic SVAF1 copies characterized by the 3′-transduced AluSp sequences (21,23,39). The same mechanism hypothesized recently to be responsible for the preferential trans-mobilization of Alu elements by the L1-encoded protein machinery (3,52) could explain the favored mobilization of SVA RNAs harbouring intact Alu elements at their 3′-terminus (21). In this model, the Alu sequence is docked on ribosomes via the SRP9/14 complex and captures the L1 ORF2 protein as it is translated from an active L1 element mRNA (52). Provided that the 3′-terminal AluSp elements in SVA transcripts allow formation of the three-dimensional structure required for SRP9/14 interaction (56), Alu sequences could mediate docking of the respective SVA RNA to the ribosome via SRP9/14 and thus facilitate efficient capture of ORF2 proteins (3,52,54). Interestingly, the AluSc element in the 5′-region of the SVAF1 source element (Figure 1A) does not seem to be beneficial for trans-mobilization, because the canonical SVAE reporter element in pAD3/SVAE which is devoid of AluSc, is mobilized at a similar frequency (~18-fold) as the SVAF1 source element in pSC3/SVAF1 (~12-fold) (Figure 3B).

**Comparison of trans-mobilization rates of SVA and Alu elements**

The trans-mobilization frequency of our canonical SVAE reporter element in HeLa cells after hygromycin selection equates to a relative retrotransposition rate of 0.4–0.9%
as to the cis retrotransposition frequency of the pJM101/L1RP-encoded functional L1RP element which was set as 100% (Figures 2A and 3B). Trans-mobilization of the full-length SVA\textsubscript{E} reporter element exceeded pseudogene formation rates of 0.01–0.04% in HeLa-JVM and HeLa-HA cells by 2- to 5-fold and ~18-fold, respectively (Figures 2D and 3B) and coincides well with the processed pseudogene formation frequencies of 0.01–0.05% reported previously by Moran et al. (2). Differences in retrotransposition activity between human cell lines have been reported earlier (50) and are also known to exist between different HeLa cell lines (Moran and Deininger, personal communication). The ~25-fold increase of the trans-mobilization rate of the SVA\textsubscript{F1} source element H10-1 after the addition of the intact Alu\textsubscript{Sp} sequence (Figure 3) equates to a relative retrotransposition rate of ~11%. This corresponds approximately to the relative trans-mobilization frequency (~10%) of the AluYa5a2NF1 element (3) which is one of the most active Alus described to date (53). Trans-mobilization rates of the canonical SVA\textsubscript{E} reporter element (pAD3/SVA\textsubscript{E}, 0.68%) and the SVA\textsubscript{F1} source element that lacks Alu\textsubscript{Sp} (pSC4/SVA\textsubscript{F1}, 0.43%) are close to the activity of a subset of polymorphic Alu\textit{Y} elements which were found to reach only 10% of the AluYa5a2NF1 retrotransposition rate (53). The relative retrotransposition rates of L1, Alu and SVA elements determined in cell culture assays do not reflect their recently estimated in vivo retrotransposition rates of one in 212, 21 and 916 births, respectively (25). One possible explanation for discrepancies, for example, in the case of SVA and L1 elements, could be an excessive upregulation of SVA source element transcription in the germ line or during early stages of embryonal development relative to L1 transcription.

Hancks et al. (50) referred only a 2-fold increase in trans-mobilization rate of the SVA\textsubscript{F1} source element after acquisition of the 3'-transduced Alu\textsubscript{Sp} sequence and that Alu\textit{Y} is trans-mobilized 30-fold more efficiently than the Alu\textsubscript{Sp} including SVA\textsubscript{F1} element. There are several differences in the design of SVA reporter and L1 donor plasmids between the two reports which make a comparison of the presented results rather complicated. First, the 3'-transduced sequence including the Alu\textsubscript{Sp} element in the SVA\textsubscript{F1} reporter used by Hancks and coworkers (50) differed from the genomic sequence in several nucleotide substitutions which were reported to affect trans-mobilization rates significantly. In contrast, nucleotide sequences of all SVA reporter elements tested in our study match genomic sequences. Second, unlike Hancks et al., we removed transcriptional terminations signals at the SINE-R 3'-ends and in the 3'-flanking sequence of the Alu\textsubscript{Sp} 3' TSD to ensure transcriptional read-through into the mneoI cassettes and consistent polyadenylation at the pCEP4-encoded SV40 polyadenylation signal (Figure 1A). Third, we fused the mneoI indicator cassette always with the 3'-end of the respective SVA reporter element, while Hancks and coworkers inserted this cassette between SINE-R and Alu\textsubscript{Sp} sequence of the SVA\textsubscript{F1} element. As a result, the transcriptional start site of the mneoI cassette is located ~430-bp upstream of the 3'-end of the SVA\textsubscript{F1} RNA, while the transcription start site of the same cassette in our analogous construct pSC3/SVA\textsubscript{F1} is located at the RNA 3'-end. Since 5'-truncations occur during SVA trans-mobilization, it could well be that the location of the indicator cassette upstream of the SVA 3'-end leads to the formation of less G418\textsuperscript{R} foci or EGFP expressing cells in cell culture assays.

**SVA de novo insertions bear the hallmarks of mobilization by the L1 protein machinery**

Analysis of pre- and post-integration sites of 11 SVA de novo insertions uncovered the hallmarks of L1-mediated retrotransposition. We found variable TSDs of 8–19 bp in length which correspond well to the TSD lengths of preexisting genomic L1 insertions ranging from 9 to 27 bp (15). The nucleotide profile of the target sites of the 11 analyzed SVA\textsubscript{E} de novo insertions resembles the L1 consensus target sequence 5'-TTTT/AA-3' (Figure 6). SVA insertions into or next to Alu TSDs can be explained by the fact that their AT-rich TSDs represent recognition sequences for the L1 endonuclease in an otherwise AT-poor environment (Figure 5).

Poly(A) tail lengths of SVA de novo insertions (30–78 nt) exceeded those of pre-existing SVA\textsubscript{E} (2–72 nt) (21) significantly. Similar differences have been reported for L1 de novo insertions (3–150 nt) and preexisting L1s (approximately 13 adenosines) (43,57). One reason for the differences in polyA tail lengths between preexisting and de novo insertions might be the fact that each de novo insertion derived from an SVA\textsubscript{E} reporter cassette was polyadenylated at the sole SV40pA site present at the 3'-end of the reporter construct (Figure 1A), while RNAs derived from genomic SVA insertions are polyadenylated at sites encoded by the SINE-R region.

We found 9/11 (~82%) SVA de novo insertions to be full-length covering 3.35 and 2.85 kb, respectively (Figure 5 and Supplementary Table S2). Full-length insertions included the entire spliced transcript expressed from the SVA retrotransposition reporter construct, starting with position +3 or +4 relative to the transcriptional initiation site of the CMV promoter. The fact that the percentage of 63% of preexisting full-length SVA insertions differs from the proportion of de novo full-length insertions could be a consequence of the relative small number of analyzed de novo insertions relative to the statistically more significant pool of ~2760 preexisting SVAs analyzed by Wang et al. (12). Since it was reported that the length of de novo insertions generated by L1 proteins in trans depends on the activity of this particular L1 element (58), the observed difference might alternatively be attributed to the use of L1\textsubscript{RP}, one of the most active human L1 elements identified to date (59), as L1 protein donor in our reporter assays. In vivo trans-mobilization of genomic SVA elements, however, is probably mediated by a multitude of different functional L1 elements with most of them being less active than L1\textsubscript{RP} (59). Therefore, lower frequencies of endogenous full-length SVA retrotransposition events would be expected. In the case of L1, only ~6% of all de novo retrotransposition events (43–45) and ~5% of preexisting copies (15) are full-length.
Clearly, one reason for the significantly larger fraction of full-length SVA elements compared to L1 is the fact that L1 RT can process more of the comparatively short full-length SVA-encoded mRNAs which are only 700–4000 bp in length (12,23) than 6-kb transcripts that are encoded by full-length L1s. Since the average size of recovered L1RN de novo insertions is ≈3165 bp (44), and our de novo SVA full-length insertions are 3303 or 2805 bp in length (Supplementary Table S2), one would expect an increased rate of full-length SVA insertions. The fact that in length (Supplementary Table S2), one would expect an increased rate of full-length SVA insertions. The fact that L1 RT successfully reverse transcribes the GC-rich VNTR region of the SVA reporter elements militates against the hypothesis that L1 5′-truncations are a consequence of an extenuated L1 RT activity. Hancks et al. reported that mobilization of an SVAE source element that was not under transcriptional control of an external promoter generated only 5′-truncated de novo insertions (50). This result suggests that SVA elements per se do not include strong promoter sequences that could facilitate full-length transcription, and that mmeol1 insertions derived from the promoterless SVAE reporter might have been produced from a cryptic promoter located upstream of the SVA sequence on the reporter plasmid. Alternatively, these SVAE insertions might not represent 5′-truncations at all and SVA transcription started within the VNTR region as numerous transcription start sites exist throughout the SVA sequence (23,50).

We found untemplated G nucleotides at the 5′-end of all full-length SVA de novo insertions. They have been described originally for L1 insertions (60) and were attributed to reverse transcription of the 7-methyl guanosine cap (44). Endogenous SVAs are very likely transcribed by RNA polymerase II. 5′ capping of SVA RNAs was suggested in earlier reports because of the presence of guanosine residues at the 5′-end of about 33% of pre-existing SVA insertions (12). The 5′/11 insertions are characterized by short patches of microcomplementarity of 1–2 nt at the junctions between SVA 5′-end and TSD. Such short patches of microhomology were reported earlier for L1 5′ junctions (36) and indicate the involvement of double strand break repair by error-prone non homologous endjoining (NHEJ) in the attachment of the SVA 5′-end to the chromosomal DNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6, Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6, Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6, Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6, Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6, Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6.

ACKNOWLEDGEMENTS

The authors thank John Moran for providing plasmids pJM101/L13, pJM101/L1RP, and pJM101/L1pAneo and the cell lines HeLa-HA and HeLa-JVM. The authors are grateful to John Goodier and Haig Kazazian who provided us with the reliable zL1ORF2p-N antibody. The authors want to thank Mark Batzer for giving us access to his genomic AluYa5 database. The authors are indebted to Kay-Martin Hanschmann for statistical analyses.

FUNDING

Deutsche Forschungsgemeinschaft (DA 545/2-1 to A.D. and G.G.S.). Funding for open access charge: Paul-Ehrlich-Institut and Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

Selectively inhibition of Alu retrotransposition by APOBEC3G.


