ORIGINAL ARTICLE

Structure and Expression Analyses of SVA Elements in Relation to Functional Genes

Yun-Jeong Kwon, Yuri Choi, Jungwoo Eo, Yu-Na Noh, Jeong-An Gim,
Yi-Deun Jung, Ja-Rang Lee, Heui-Soo Kim*

Department of Biological Sciences, College of Natural Sciences, Pusan National University, Busan 609-735, Korea

SINE-VNTR-*Alu* (SVA) elements are present in hominoid primates and are divided into 6 subfamilies (SVA-A to SVA-F) and active in the human population. Using a bioinformatic tool, 22 SVA element-associated genes are identified in the human genome. In an analysis of genomic structure, SVA elements are detected in the 5′ untranslated region (UTR) of *HGSNAT* (SVA-B), *MRGPRX3* (SVA-D), *HYAL1* (SVA-F), *TCHH* (SVA-F), and *ATXN2L* (SVA-F) genes, while some elements are observed in the 3′UTR of *SPICE1* (SVA-B), *TDRKH* (SVA-C), *GOSR1* (SVA-D), *BBS5* (SVA-D), *NEK5* (SVA-D), *ABHD2* (SVA-F), *C1QTNF7* (SVA-F), *ORC6L* (SVA-F), *TMEM69* (SVA-F), and *CCDC137* (SVA-F) genes. They could contribute to exon extension or supplying poly A signals. *LEPR* (SVA-C), *ALOX5* (SVA-D), *PDS5B* (SVA-D), and *ABCA10* (SVA-F) genes also showed alternative transcripts by SVA exonization events. Dominant expression of *HYAL1*_SVA appeared in lung tissues, while *HYAL1*_noSVA showed ubiquitous expression in various human tissues. Expression of both transcripts (*TDRKH*_SVA and *TDRKH*_noSVA) of the *TDRKH* gene appeared to be ubiquitous. Taken together, these data suggest that SVA elements cause transcript isoforms that contribute to modulation of gene regulation in various human tissues.

Keywords: gene expression profiling, genomics, messenger RNA isoform, SVA, tissue specificity

Introduction

The human genome contains actively moving retrotransposons (human endogenous retrovirus [HERV], long interspersed nucleotide element [LINE], and short interspersed nucleotide element [SINE]). Approximately 36% of the human genome has been generated through retrotransposition of LINE and other RNA species by the LINE reverse transcriptase [1, 2]. Retrotransposition is progressing in the human population by L1, *Alu*, and SINE-VNTR-*Alu* (SVA) insertions [3]. L1 is an autonomous retrotransposon that contains an internal RNA polymerase II promoter and a reverse transcriptase, whereas *Alu* and SVA lack activities for independent mobilization [4, 5]. Therefore, *Alu* and SVA are assumed to use the L1 protein machinery for their own mobilization [3], and retrotransposition events of marked SVA elements occur indeed by L1 elements in human cultured cells [5, 6].

The SVA elements were originally named from the SINE-R retroposon, derived from an endogenous retrovirus, the HERV-K LTR element. SINE-R11, 14, and 19 have been isolated by colony blot hybridization using the LTR element as probe [7]. SINE-R.C2 has been found in the third intron of the C2 gene on the short arm of human chromosome 6, which was a human-specific element [8]. Within the Xq21.3 block, two SINE-R retroposons (HS307 and HS408) were identified [9]. Multiple copy numbers of retroposons have been successively detected in hominoid primates and humans [9-16]. Other similar sequences have been associated with *Alu*-like sequences within the *HLA-RP1* (*STK19*) gene [17]. These composite retroposons with the entire structure are named SVA (SINE-R, VNTR, and *Alu*).

The SVA families are evolutionarily young and hominoid-specific retroelements. They have ability to influence a genomic locus in which they reside and cause various human diseases as insertional mutagens. In fact, Fukuyama muscular dystrophy (FCMD) is derived from the polymorphic SVA-E insertion in the 3' untranslated region (UTR) of the *FUKUTIN* gene [18]. SVA elements could also drive transcription of functional genes. In the 5' upstream region of

TBPL2, the SVA element serves as a promoter, while the SVA-D element in the *WDR66* gene promotes the transcription of a human-specific transcript variant [19]. Here, we analyzed structure variants of functional genes mediated by SVA subfamilies and examined their expression patterns in various human tissues.

Methods

Bioinformatic analysis

To identify SVA consensus sequences in the human genome, we obtained SVA sequences from the Giri database (http://www.girinst.org). The SVA subfamilies were aligned using the BioEdit program [20]. Then, we identified SVAs in each region. RepeatMasker (http://www.repeatmasker.org) and the UCSC genome site (http://genome.ucsc.edu) were employed to analyze isoform structures of functional genes. The human expressed sequence tag and RefSeq mRNA were also used to identify alternatively spliced transcripts. The expression pattern of SVA fusion genes in normal human tissues was analyzed using GeneCard (http://www.gene-cards.org). We obtained microarray data from the BioGPS database, and then we generated a heatmap according to microarray values. High expression levels were indicated by brighter color, and low expression levels were indicated by darker color.

Human RNA samples

A human 20-RNA tissue master panel (1, adrenal gland; 2, bone marrow; 3, cerebellum; 4, whole brain; 5, fetal brain; 6, fetal liver; 7, heart; 8, kidney; 9, liver; 10, lung; 11, placenta; 12, prostate; 13, salivary gland; 14, skeletal muscle; 15,

spinal cord; 16, testis; 17, thymus; 18, thyroid; 19, trachea; 20, uterus) was purchased from Clontech (Mountain View, CA, USA).

Reverse-transcription (RT) and reverse transcription polymerase chain reaction (RT-PCR) amplification

To eliminate possible DNA contamination of purchased RNA samples, Turbo DNA-free (Ambion, Austin, TX, USA) was used according to the manufacturer's instructions. A no-RT control was also amplified to double-check the absence of DNA contamination. Quantity of RNA samples was measured using a ND-1000 UV-Vis spectrophotometer (NanoDrop, Wilmington, DE, USA). Moloney-Murine-Leukemia-Virus reverse transcriptase with an annealing temperature of 42°C was used for the RT reaction with RNase inhibitor (Promega, Madison, WI, USA). To develop the specific primers for individual alternative transcripts, primer pairs were designed with the aid of Primer3 (http://frodo.wi.mit.edu/) (Table 1). In each run, 1 μL of cDNA was used as template for amplification per reaction. RT-PCR was performed using reactions containing a mixed cDNA template, representing a combination of different tissues examined. RT-PCR amplification for functional genes and a housekeeping gene was carried out for 30 cycles of 94°C for 3 minutes, 56–60°C for 1 minute, and 72°C for 3 minutes. As a standard control, *G3PDH* was amplified through RT-PCR in human tissues. PCR products were loaded on 1–2% agarose gels and stained with ethidium bromide.

Results and Discussion

SVAs are composite elements consisting of multiple

Table 1. List of RT-PCR primer sets for expression analysis of SVA fusion genes

Gene	Accession no.	Location	Product size (bp)	Primer sequence	Tm (°C)
*HYAL1*_SVA	NM_153281	3p21.3	1,631	F: GCCCCGTCTGAGAAGTGA	53.5
				R: CCAGGGTAGCATCGACATTT	
*HYAL1*_noSVA	NM_033159	3p21.3	378	F: CATCCTGAACGTGACCAGTG	53
				R: TGAATGGTGTCTGCTGTGGT	
*TDRKH*_SVA	AK225160	1q21	538	F: AGTTTGCTCAGCCATCTGCT	54
				R: TACAGTCCAGCTTCGGCTTT	
*TDRKH*_noSVA	NM_006862	1q21	1,170	F: GAGGATGGAGACCGAAGGAT	54
				R: GGGTGCTCAGAAGCAGAAAC	
*ORC6L*_SVA	AK_024077	16q12	853	F: GCAGTGAACATGGCTTCAAA	53
				R: TCCCCCTCTGTGAGAAACAC	
*ORC6L*_noSVA	NM_014321	16q12	490	F: GCAGTGAACATGGCTTCAAA	53
				R: TTCCTGGAGATGGCAATGTA	
G3PDH	NM_002046	12p13.31	249	F: GAAATCCCATCACCATCTTCCAGG	55
				R: GAGCCCCAGCCTTCTCCATG	

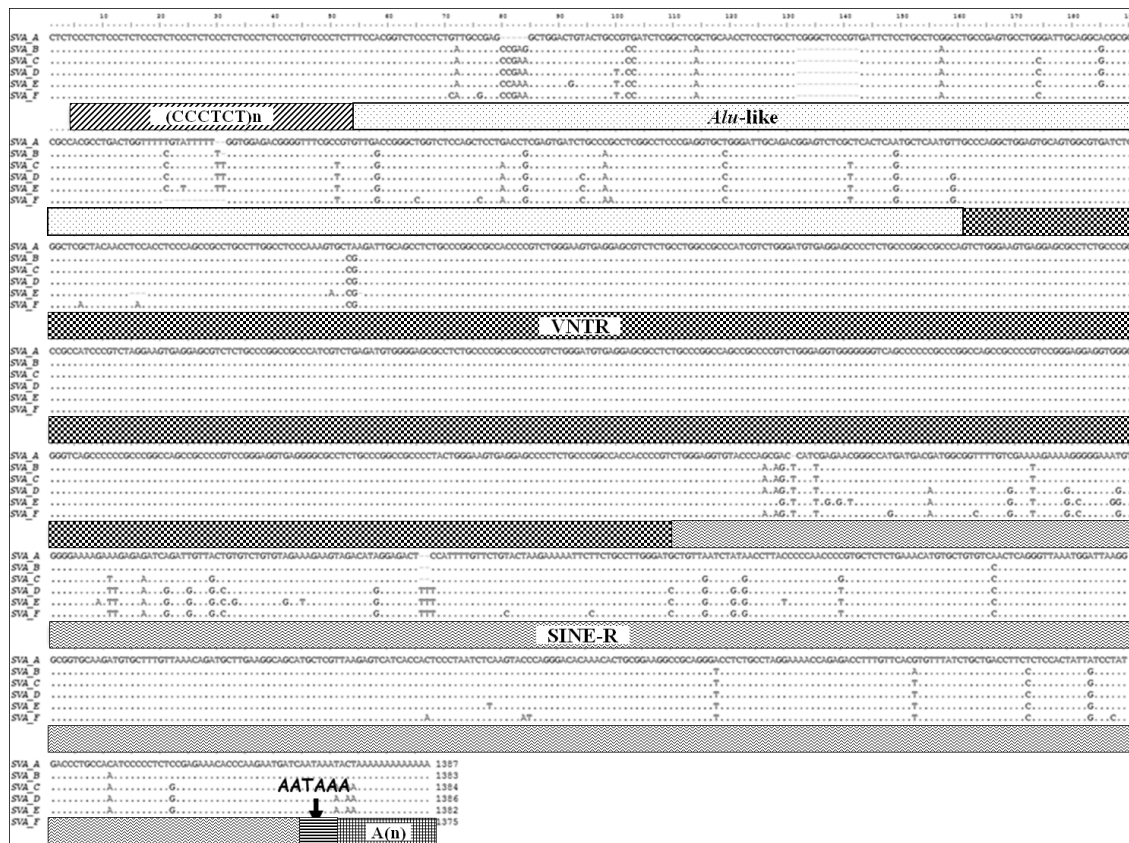RT-PCR, reverse transcription polymerase chain reaction; SVA, SINE-VNTR-*Alu*.

Fig. 1. Nucleotide sequence alignments of SINE-VNTR-*Alu* (SVA) families. Dots indicate no change to the SVA-A sequences, and dashes indicate gaps. The consensus sequences of SVA-A, -B, -C, -D, -E, and -F were taken from Giri DB (http://www.girinst.org). Main domains (*Alu*, VNTR, and SINE-R) are indicated.

domains: a CCCTCT repeat, *Alu*-like domain, a GC-rich variable number of tandem repeat (VNTR), and SINE-R derived from the HERV-K LTR element [7-9, 21, 22]. They are flanked by target site duplications and terminate in a poly(A) tail (Fig. 1). In genomic sequence analysis, SVA elements are present in G + C-rich regions; however, they do not have any preferences for inter- or intragenic regions. SVA families are separated into six subfamilies (SVA-A to SVA-F), based upon point mutation and insertion and deletion events within the SINE-R [22]. Among them, four subfamilies (SVA-A, SVA-B, SVA-C, SVA-D) are present in gibbons and orangutans, while two subfamilies (SVA-E and SVA-F) are restricted to the human lineage [22].

SVA elements residing in genes are potentially disruptive in either orientation. Approximately 1/3 of all SVA elements in the human genome reside in genic regions, with 20% of those SVA elements being in the same orientation as a gene [23]. As shown in Fig. 2, we analyzed the genomic structure of SVA fusion genes using bioinformatic tools (Repeat-Masker program and University of California, Santa Cruz [UCSC] genome browser). SVA elements are detected in the 5'UTR of the *HGSNAT* (SVA-B, AK_057293), *MRGPRX3*

(SVA-D, NM_054031.2), *HYAL1* (SVA-F, NM_153281), *TCHH* (SVA-F, AK_307946), and *ATXN2L* (SVA-F, AY_188334-8) genes, while some elements are observed in the 3'UTR of the *SPICE1* (SVA-B, NM_144718), *TDRKH* (SVA-C, AK_225160), *GOSR1* (SVA-D, NM_001007024), *BBS5* (SVA-D, NM_1523842), *NEK5* (SVA-D, NM_199289; AK_126330), *ABHD2* (SVA-F, NM_007011), *C1QTNF7* (SVA-F, NM_031911.3), *ORC6L* (SVA-F, AK_024077.1), *TMEM69* (SVA-F, NM_016486.2), and *CCDC137* (SVA-F, NM_199287.2) genes. Within the genic region, SVA elements seem to prefer the 3'UTR compared to another regions. Recently, a novel promoter derived from the SVA-D element was identified in the 5'UTR of the *TBPL2* gene (DB089735) [19]. In order to understand whether the SVA element could act as a transcriptional regulator or not, the functional activity of the *PARK7* SVA element was performed, suggesting that the SVA with the SINE-R region deleted showed significant activity to enhance reporter gene activity in SK-N-AS and MCF-7 cells [24]. Likewise, we have also demonstrated the activity of an alternative promoter by the HERV-H LTR element in the human *GSDML* gene [25]. The transcripts of the LTR-derived promoter were widely dis-
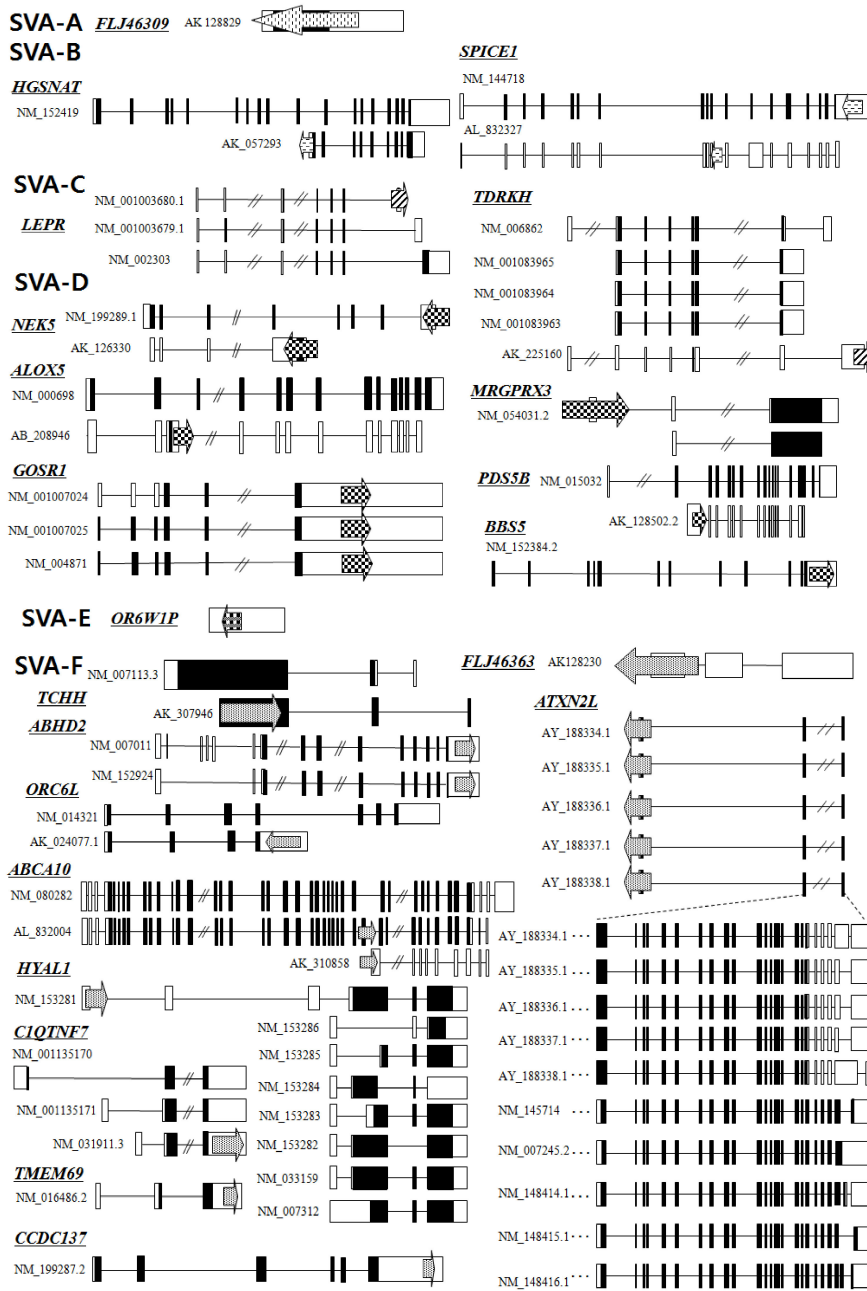
Fig. 2. Alternative splicing patterns of functional genes by SINE-VNTR-*Alu* (SVA) families. SVA element is indicated by the arrow.

tributed in various tissues, whereas the transcripts of the original promoter were found in stomach tissues, compared to those of various human tissues [25]. A human-specific solitary LTR element (L47334) was previously shown to have enhancer activity in Tera-1 human testicular carcinoma cells [26]. The varying genetic structure of the alternative promoter or enhancer results in different functional effects [27]. Those events could regulate alternative transcript variants, resulting in the production of different protein isoforms [28]. In the case of leptin receptor isoform 219.1, SVA sequences form the C-terminal coding exon. Using the

SVA-specific probe, Southern blot analysis in hominoids and Old World monkeys was performed, indicating that it was found to be a human-specific SVA element [29]. Furthermore, new exons are generated from intronic retrotransposons or the elements inserts into the exon, leading to a new transcript [30]. In the present study, the *LEPR* (SVA-C, NM_001003680), *ALOX5* (SVA-D, AB_208946), *PDS5B* (SVA-D, AK_128502.2), and *ABCA10* (SVA-F, AL_832004) genes showed alternative transcripts by the SVA exonization event. To investigate the supply of alternative splicing sites at SVA elements, we analyzed related splicing sites using
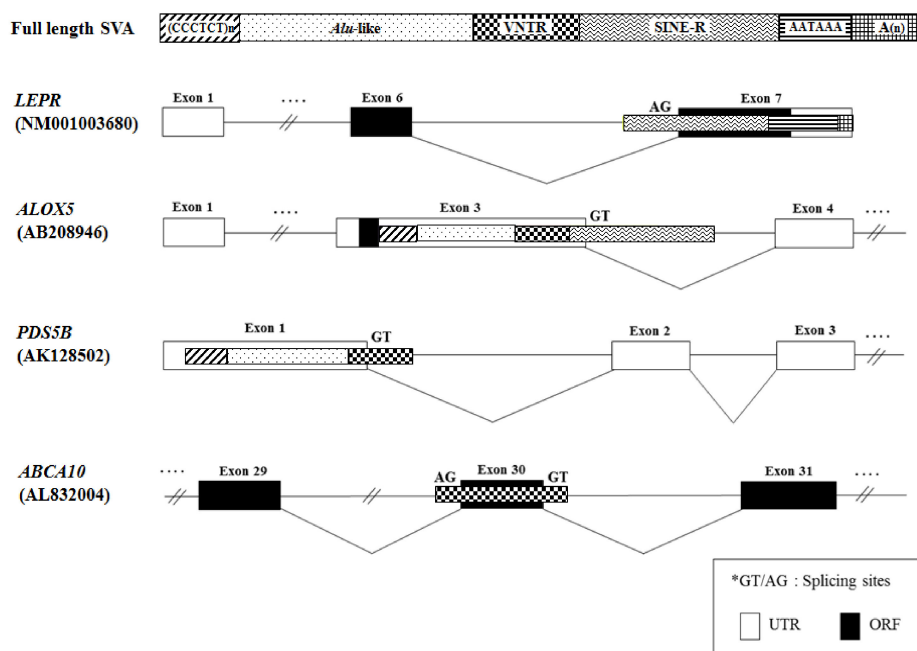
Fig. 3. Exonization and splice site analysis by SINE-VNTR-*Alu* (SVA) element in functional genes. UTR, untranslated region; ORF, open reading frame.

human genomic sequences (http://genome.ucsc.edu/) (Fig. 3). In the case of *ABCA10* (AL832004) genes, the SVA element provided both a splicing donor and acceptor site for the cassette exon. In the case of the *ALOX5* (AB208946) and *PDS5B* (AK128502) genes, the SVA element provided a splicing donor site for alternative 5' splicing. Conversely, the *LEPR* (NM001003680) gene was provided a splicing acceptor site for alternative 3' splicing. These phenomena conformed to the canonical splicing site. Also, the *PDS5B* (AK128502) and *ABCA10* (AL832004) genes induced exonization by providing alternative splicing sites in the VNTR region of SVA elements. Exonization of the *LEPR* (NM001003680) and *ALOX5* (AB208946) genes occurred by the SINE-R region of SVA elements. This integration event of SVA elements into the exon or intron causes the promotion of transcriptional variants of functional genes. In *WDR66* (NM_144668), the SVA-F element, inserted in intron 19, could produce human-specific transcripts that spliced to the last three exons [19]. Those transcripts, derived from SVA elements, could have important biological function in humans, therefore deserving further investigation in various tissues of hominoid primates.

SVA elements are capable of generating individual variation in gene expression at loci in which they are present. RT-PCR amplification could be a good indicator to detect alternative variants derived from SVA elements. As shown in Fig. 4, dominant expression of *HYAL1*_SVA appeared in lung tissues, while *HYAL1*_noSVA showed ubiquitous expression in various human tissues. Expression of both transcripts (*TDRKH*_SVA and TDRKH_noSVA) of the *TDRKH* gene
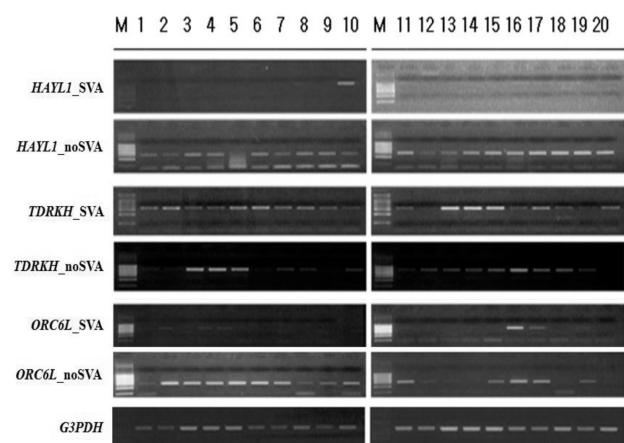

Fig. 4. Reverse transcription polymerase chain reaction analysis of transcript variants of functional genes in various human tissues. 1, adrenal gland; 2, bone marrow; 3, cerebellum; 4, whole brain; 5, fetal brain; 6, fetal liver; 7, heart; 8, kidney; 9, liver; 10, lung; 11, placenta; 12, prostate; 13, salivary gland; 14, skeletal muscle; 15, spinal cord; 16, testis; 17, thymus; 18, thyroid; 19, trachea; 20, uterus. *G3PDH* indicates the positive control.

appeared to be ubiquitous. We also examined *in silico* expression of SVA fusion genes using microarray data obtained by the BioGPS database. Transcripts of the *TCHH* and *HYAL1* genes appeared dominant in placenta and liver, respectively. The *ALOX5* gene showed dominant expression in whole blood and lung. Also, dominant expression of the *ABHD2* gene appeared in testis and prostate tissues, whereas the *BBS5, NEK5, MRGPRX3, PDS5B, OR6W1P, ABCA10, TMEM69, C1QTNF7,* and *ATXN2L* genes indicated ubiqui-
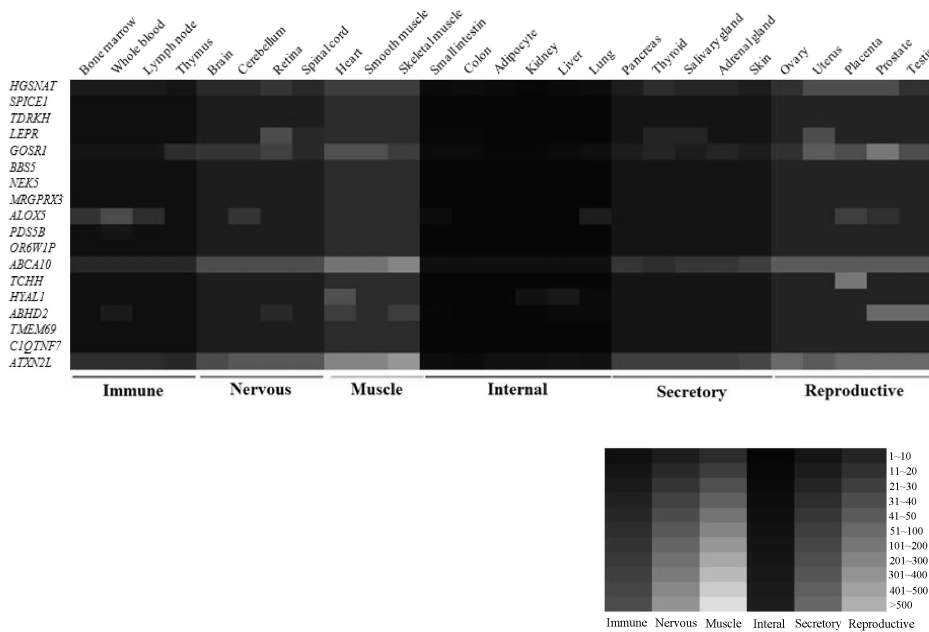
Fig. 5. Heatmap of SINE-VNTR-*Alu* (SVA) fusion gene expression profiles. Horizontal lines separate the 6 clusters via biological systems. SVA fusion genes are ordered according to SVA subfamily type at vertical lines. Values increase from dark to bright.

tous expression (Fig. 5). SVA elements are nonautonomous retrotransposons that cause diseases in humans, and they are mobilized in trans by active L1 elements. The quantitative real-time polymerase chain reaction analysis of fukutin mRNA in lymphoblasts from FCMD patients indicated that the disease results from SVA-E insertion in the 3'UTR of the *FUKUTIN* gene. Sequence data demonstrated an abnormal splicing event by the integration of an SVA-E element [18]. In the case of the *PMS2* gene, SVA-F has inserted in intron 7, and causes Lynch syndrome. Sequence analysis of the RT-PCR product revealed a 71-bp SVA-F element between *PMS2* exon 7 and 8 in the aberrant transcripts [31]. Exon 3 of the *PNPLA2* gene was interrupted by a 1.8-kb SVA-F insertion, which causes lipid storage disease with subclinical myopathy [32], while an SVA-F element inserted in intron 1 of the *ARH* gene causes hypercholesterolemia [33]. Taken together, alternative promoter, enhancer, polyadenylation, and exonization events by SVA elements cause various transcript isoforms and evolutionary dynamics that contribute to flexibility in the regulation of gene expression and hominoid radiation, including human disease.

Acknowledgments

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
2. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 2009;10:691-703.
3. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* 2003;73:1444-1451.
4. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, *et al*. LINE-1 retrotransposition activity in human genomes. *Cell* 2010;141:1159-1170.
5. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, *et al*. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* 2012;40:1666-1683.
6. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet* 2011;20:3386-3400.
7. Ono M, Kawakami M, Takezawa T. A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* 1987;15:8725-8737.
8. Zhu ZB, Jian B, Volanakis JE. Ancestry of SINE-R.C2 a human-specific retroposon. *Hum Genet* 1994;93:545-551.
9. Kim HS, Wadekar RV, Takenaka O, Winstanley C, Mitsunaga F, Kageyama T, *et al*. SINE-R.C2 (a *Homo sapiens* specific retroposon) is homologous to CDNA from postmortem brain in schizophrenia and to two loci in the Xq21.3/Yp block linked to handedness and psychosis. *Am J Med Genet* 1999;88:560-566.
10. Kim HS, Crow TJ. Presence and phylogenetic relationships of a hominoid-specific retroposon family on the human Y chromosome. *Zool Sci* 1999;16:963-970.
11. Kim H, Crow TJ. Identification and phylogeny of novel human endogenous retroviral sequences belonging to the HERV-W family on the human X chromosome. *Arch Virol* 1999;144:2403-2413.

12. Kim HS, Crow TJ. Presence and phylogenetic analysis of HERV-K LTR on human X and Y chromosomes: evidence for recent proliferation. *Genes Genet Syst* 1999;74:267-270.

13. Kim HS, Crow TJ. Phylogenetic relationships of a class of hominoid-specific retro-elements (SINE-R) on human chromosomes 7 and 17. *Ann Hum Biol* 2000;27:83-93.

14. Kim HS, Crow TJ. Cloning and phylogeny of endogenous retroviral elements belonging to the HERV-K LTR in cDNA library of human fetal brain and Xq21.3 region linked to psychosis. *Korean J Genet* 2001;23:129-134.

15. Kim HS, Wadekar RV, Takenaka O, Hyun BH, Crow TJ. Phylogenetic analysis of a retroposon family in african great apes. *J Mol Evol* 1999;49:699-702.

16. Kim HS, Takenaka O, Crow TJ. Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J Gen Virol* 1999;80(Pt 10):2613-2619.

17. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, *et al*. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. *J Biol Chem* 1994;269:8466-8476.

18. Taniguchi-Ikeda M, Kobayashi K, Kanagawa M, Yu CC, Mori K, Oda T, *et al*. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature* 2011;478:127-131.

19. Kim DS, Hahn Y. Identification of human-specific transcript variants induced by DNA insertions in the human genome. *Bioinformatics* 2011;27:14-21.

20. Tippmann HF. Analysis for free: comparing programs for sequence analysis. *Brief Bioinform* 2004;5:82-87.

21. Kim HS, Choi JY, Lee WH, Jang KL, Hyun BH. Nucleotide sequence and phylogenetic analysis of long terminal repeats of human endogenous retrovirus K family (HERV-K) on human chromosomes. *Microb Comp Genomics* 2000;5:121-127.

22. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, *et al*. SVA elements: a hominid-specific retroposon family. *J Mol Biol* 2005;354:994-1007.

23. Hormozdiari F, Hajirasouliha I, Dao P, Hach F, Yorukoglu D, Alkan C, *et al*. Next-generation VariationHunter: combinatorial algorithms for transposon insertion discovery. *Bioinformatics* 2010;26:i350-i357.

24. Savage AL, Bubb VJ, Breen G, Quinn JP. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* 2013;13:101.

25. Sin HS, Huh JW, Kim DS, Kang DW, Min DS, Kim TH, *et al*. Transcriptional control of the HERV-H LTR element of the *GSDML* gene in human tissues and cancer cells. *Arch Virol* 2006;151:1985-1994.

26. Ruda VM, Akopov SB, Trubetskoy DO, Manuylov NL, Vetchinova AS, Zavalova LL, *et al*. Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Res* 2004;104:11-16.

27. Gerlo S, Davis JR, Mager DL, Kooijman R. Prolactin in man: a tale of two promoters. *Bioessays* 2006;28:1051-1055.

28. Xin D, Hu L, Kong X. Alternative promoters influence alternative splicing at the genomic level. *PLoS One* 2008;3:e2377.

29. Damert A, Löwer J, Löwer R. Leptin receptor isoform 219.1: an example of protein evolution by LINE-1-mediated human-specific retrotransposition of a coding SVA element. *Mol Biol Evol* 2004;21:647-651.

30. Mersch B, Sela N, Ast G, Suhai S, Hotz-Wagenblatt A. SERpredict: detection of tissue- or tumor-specific isoforms generated through exonization of transposable elements. *BMC Genet* 2007;8:78.

31. van der Klift HM, Tops CM, Hes FJ, Devilee P, Wijnen JT. Insertion of an SVA element, a nonautonomous retrotransposon, in *PMS2* intron 7 as a novel cause of Lynch syndrome. *Hum Mutat* 2012;33:1051-1055.

32. Akman HO, Davidzon G, Tanji K, Macdermott EJ, Larsen L, Davidson MM, *et al*. Neutral lipid storage disease with subclinical myopathy due to a retrotranspospal insertion in the *PNPLA2* gene. *Neuromuscl Disord* 2010;20:397-402.

33. Wilund KR, Yi M, Campagna F, Arca M, Zuliani G, Fellin R, *et al*. Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet* 2002;11:3019-3030.