

Human-specific antisense transcripts induced by the insertion of transposable element

DONG SEON KIM and YOONSOO HAHN

Department of Life Science (BK21 Program) and Research Center for Biomolecules and Biosystems,
Chung-Ang University, Seoul 156-756, Korea

Received February 26, 2010; Accepted April 16, 2010

DOI: 10.3892/ijmm_00000447

Abstract. Antisense transcripts can influence the sense partner gene function by modifying the transcriptional and/or posttranscriptional regulation processes. Here, we report the identification of 13 cases of human-specific antisense transcripts induced by transposable element insertions from the analysis of primate genome alignment and human transcriptome data. The original sources of the insert included L1, Alu, SVA and human endogenous retrovirus (HERV). In the majority of the cases, insertion of a transposable element served as a promoter and drove transcription of the adjacent genomic segment, creating a novel antisense transcriptional unit (for examples, *RNF144A*, *SYNE2*, *CAMCK4* and *LSAMP*). In the remaining cases, an existing antisense transcript was modified upon insertion; the insert supplied a promoter (*ABCA9*), an internal exon (*LHFPL3* and *DSG1*) or a terminal exon (*TEX11*). We propose that creation of human-specific antisense transcripts may have altered the partner gene function and consequently may have played a role in the acquisition of various human-specific traits.

Introduction

The human genome contains a large number of genetic alterations acquired after the human-chimpanzee divergence. Many of these human-specific changes may have been associated with the evolution of human-specific phenotypes that are distinct from the other great apes. For example, *FOXP2*, a gene encoding the transcription factor which is known to be implicated in speech and language, exhibits an accelerated nonsynonymous substitution rate in humans (1). This change was suggested to cause a different transcription pattern of the central nervous system development genes between humans and chimpanzees (2). A noncoding RNA

HAR1F which is expressed during cortical development shows an accelerated nucleotide sequence substitution rate in humans (3). Loss of gene function is also related with emergence of some human-specific traits; the mechanism is often expressed as the 'less-is-more' hypothesis (4). For example, the *MYH16* gene was found to be inactivated by a frameshift mutation and was proposed to be associated with the smaller jaws and bigger brains of humans (5). The *CMAH* gene was found to be inactivated by an exon deletion (6) and is believed to be responsible for resistance to *Plasmodium reichenowi* malaria but for susceptibility to *P. falciparum* malaria (7,8).

Currently, the genome sequences of several primate species have been determined and are available including the human (9), chimpanzee (10), orangutan (http://genome.wustl.edu/genomes/view/pongo_abelii/), rhesus macaque (11) and marmoset (http://genome.wustl.edu/genomes/view/callithrix_jacchus/). By directly comparing these genome sequence data, one can systematically identify the genetic changes during human evolution; for example, genes showing accelerated rate of sequence substitution (3,12), generation of novel genes (13,14) and gene losses (15-18).

The paired sense-antisense transcription from a single genomic locus is rather common in the mammalian genomes including humans (19,20). Antisense transcription is regarded as one of the levels of the regulation to gene expression. For example, the controlled expression of the sense and antisense transcripts of a genetic locus plays an important role in many molecular biological processes, such as X inactivation (21) and genomic imprinting (22). The antisense transcripts can influence the expression of their sense partner genes by several mechanisms (23). The transcription of the sense partner gene can be disrupted due to a crash of the transcription machineries moving in opposite directions (24,25). When the mature forms of the sense and the antisense transcripts overlap with each other, they can form double-stranded RNAs which are potential substrates for the RNA interference (RNAi) machinery (26). Mutated expression of an antisense transcript can lead to gene silencing and cause a genetic disease (27).

Insertions of transposable elements such as short interspersed elements (SINEs), long interspersed elements (LINEs) and endogenous retroviruses are known to be drivers of the mammalian genome evolution (28-30). It is reported that a large number of antisense transcripts are initiated by transposable elements in the human genome (31). In this study, we report 13 cases of human-specific antisense transcripts of

Correspondence to: Dr Yoonsoo Hahn, Department of Life Science, Chung-Ang University, 47 Heukseok-ro, Dongjak-gu, Seoul 156-756, Korea
E-mail: hahny@cau.ac.kr

Key words: human genome, antisense transcript, transposable element

known genes that might have been induced by novel insertions of transposable elements in the human genome after the divergence of humans and chimpanzees.

Materials and methods

Identification of human-specific insertion. To collect novel antisense transcripts for the annotated genes that have newly emerged in the human genome after the human-chimpanzee divergence, we analyzed the human-chimpanzee genome alignment data, human-rhesus macaque genome alignment data and the human transcript-to-genome mapping data downloaded from the University of California Santa Cruz (UCSC) Genome Browser database (<http://genome.ucsc.edu>) (32). Initially, we aimed to identify human-specific transcripts induced by human-specific genomic insertions. The detailed procedure and results will be described elsewhere. Briefly, we downloaded the genome annotation databases for the human (hg18), the chimpanzee (panTro2) and the rhesus macaque (rheMac2). The human-chimpanzee and the human-rhesus macaque genome alignment data were obtained from the tables hg18.netPanTro2 and hg18.netRheMac2, respectively. The human genomic sequences that were not aligned with either the chimpanzee or the rhesus macaque genomes were collected as human-specific genomic inserts (10447 loci).

Identification of human-specific antisense transcripts. To examine whether the human-specific insert is expressed as transcripts in the human cell, we analyzed the human transcript-to-genome mapping data: RefSeqs (hg18.refSeqAli), mRNAs (hg18.chrN_mrna) and expressed sequence tags (ESTs) (hg18.chrN_intronEst), where N is the chromosome number (from 1 to 22 and X). When a human-specific insert overlapped with a transcribed sequence, it was considered as a human-specific transcript. In order to avoid the interference of any genomic DNA contamination during cDNA construction, unspliced mRNA or EST sequence data were removed. Finally, we manually examined the candidates to check whether the human-specific transcript was an antisense transcript of a known gene. As a result, we identified 13 human-specific antisense transcripts induced by insertion of transposable elements. Classification of the transposable element was obtained from the table hg18.chrN_rmsk, where N is the chromosome number.

Results and Discussion

Human-specific antisense transcripts. We identified 13 human-specific antisense transcripts associated with known protein-coding genes that have evolved to be expressed uniquely in humans upon insertion of transposable elements (Table I). The inserted transposable elements include L1, Alu, SVA and human endogenous retroviruses (HERVs). The 6 L1 elements collected in this study belong to subfamilies L1Hs, L1PA2 or L1PA3, which are reported to have been commonly amplified during the period of ape and human evolution (33,34) and are currently active in the human genome (35,36). All of the 3 Alu repeat elements observed in this analysis are young AluY subfamilies, which are actively

transposing in the human genome (35,36). We found 2 SVA-induced human-specific antisense transcripts. SVA elements are also actively amplified in the human genome (37,38). The 2 HERVs involved in the formation of novel antisense transcripts are of the HERVK subfamily which is an active HERV (35).

In the majority of the cases (8 out of 13), the human-specific antisense transcripts are initiated from the inserted transposable elements (Table I), implying that the transposable elements provide a novel promoter. It is well known that transposable elements drive transcription of adjacent genomic segment and interact with cellular genes in proximity (39,40). In the remaining cases, the inserted elements contribute a terminal exon (3 cases) or an internal exon (2 cases) to an existing antisense transcript. Addition of an exon cassette into and truncation of the transcripts are commonly induced by transposable elements (37,40,41).

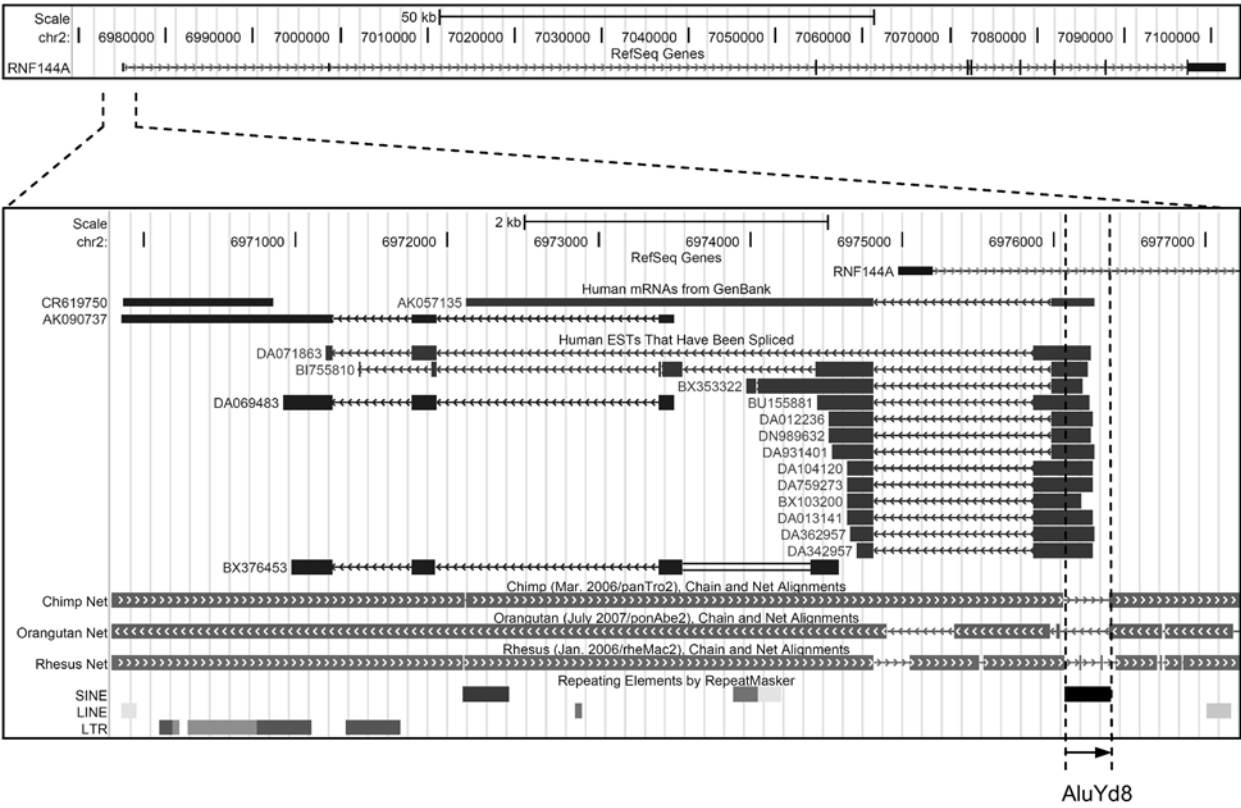
Transposable elements drive transcription of novel antisense transcripts. We identified 8 cases where the antisense transcripts are initiated from the inserted transposable elements (Table I and Fig. 1). In intron 1 of the *RNF144A* gene, there is a human-specific AluYd8 element from which 14 transcript sequences are initiated in the opposite orientation to the *RNF144A* gene (Fig. 1A). The transcriptional unit of the antisense transcript encompasses the first exon and the promoter region of the *RNF144A* gene. Exon 2 and others of the antisense transcript are in the 5' flanking region of *RNF144A*, possibly overlapping with the promoter region. It is possible that transcriptional activation of the antisense transcript may interfere with transcription of the *RNF144A* gene or block the binding of transcription factors in the promoter. In the *SYNE2* gene intron 45 (intron number is based on the RefSeq NM_182914), there is a human-specific SVA element from which 6 antisense ESTs are initiated (Fig. 1B). The antisense transcript consists of 2 exons both of which are within intron 45 of *SYNE2*.

Other notable cases of transposable element-initiated antisense transcription include the *CAMK4*, *LSAMP*, *GLRA3* and *ATP6VIG1* genes (Table I). In the *CAMK4* gene, intron 1, 2 ESTs are initiated from the antisense promoter of the human-specific L1Hs element. There are 2 more downstream antisense ESTs in this intron, probably representing truncated forms. The entire transcriptional unit of the antisense appears to be embedded in intron 1 of *CAMK4*. The antisense promoter of the L1 element is known to drive transcription of the adjacent genomic segment (39,40). The *LSAMP* gene contains a human-specific L1Hs insertion in its intron 1 from which 3 antisense ESTs are initiated. There are 5 additional antisense ESTs that share the last 2 or 3 exons out of the total 4. All of the 4 exons are positioned in the intron 1 of *LSAMP*. Since the antisense transcriptional unit is very close to the transcription start site of the *LSAMP* gene, it is highly likely that the transcription of *LSAMP* is effectively disrupted when both transcriptional units are simultaneously activated. In contrast to many examples where the antisense promoter of the L1 element is commonly involved in the transcription of nearby genomic segment, the *LSAMP* antisense transcripts are derived from the L1 sense promoter. The 5' untranslated region (UTR) of the L1 is spliced to cryptic exons in the proximal genomic DNA.

Table I. List of the human-specific antisense transcripts identified in this study.

No	Contribution to antisense transcript	Human genome position (hg18)	Origin of insert	Number of antisense transcripts	Accession numbers of antisense transcripts	Partner gene name	Partner gene accession number	Partner gene title
1	Promoter	chr2:6976077-6976372	AluYd8	14	AK057135 BI755810 BU155881 BX103200 BX353322 DA012236 DA013141 DA071863 DA104120 DA342957 DA362957 DA759273 DA931401 DN989632	<i>RNF144A</i>	NM_014746	Ring finger protein 144A
2	Promoter	chr3:117561525-117567573	L1Hs	3	AI828835 BF509771 BX110108	<i>LSAMP</i>	NM_002338	Limbic system-associated membrane protein
3	Terminal exon	chr3:179982257-179984497	L1PA2	1	DA746847	<i>KCNMB2</i>	NM_181361	Calcium-activated potassium channel $\beta 2$ subunit
4	Promoter	chr4:175936178-175942219	L1PA2	1	BF028106	<i>GLRA3</i>	NM_006529	Glycine receptor $\alpha 3$
5	Promoter	chr5:110665810-110671863	L1Hs	2	DA454582 DB019405	<i>CAMK4</i>	NM_001744	Calcium/calmodulin-dependent protein kinase IV
6	Terminal exon	chr6:78483373-78492801	HERVK	1	DB094606	-	DR731261	-
7	Internal exon	chr7:104175600-104180502	HERVK	14	BE410585 BE898131 BM459062 BQ437485 BQ680043 BU184136 BX089953 H83743 N26114 N26360 N39695 N41549 W02818 W02825	<i>LHFPL3</i>	NM_199000	Lipoma HMGIC fusion partner-like 3
8	Promoter	chr9:116399281-116399390	AluY	1	BU949417	<i>ATP6V1G1</i>	NM_004888	Vacuolar H ⁺ ATPase G1
9	Promoter	chr10:130008004-130008314	AluYb8	2	AI149915 BX089661	-	AK124226	-
10	Promoter	chr14:63573593-63575360	SVA	6	BX098046 H29276 R41832 R56239 T16573 T17168	<i>SYNE2</i>	NM_015180	Spectrin repeat containing nuclear envelope 2
11	Promoter	chr17:64506845-64512885	L1PA3	2	CB960713 CB961243	<i>ABCA9</i>	NM_080283	ATP-binding cassette sub-family A member 9
12	Internal exon	chr18:27181918-27187963	L1PA2	3	AW468896 DB461091 DB526502	<i>DSG1</i>	NM_001942	Desmoglein 1 preproprotein
13	Terminal exon	chrX:69815491-69817169	SVA	1	CN270894	<i>TEX11</i>	NM_031276	Testis expressed sequence 11

A *RNF144A*



B *SYNE2*

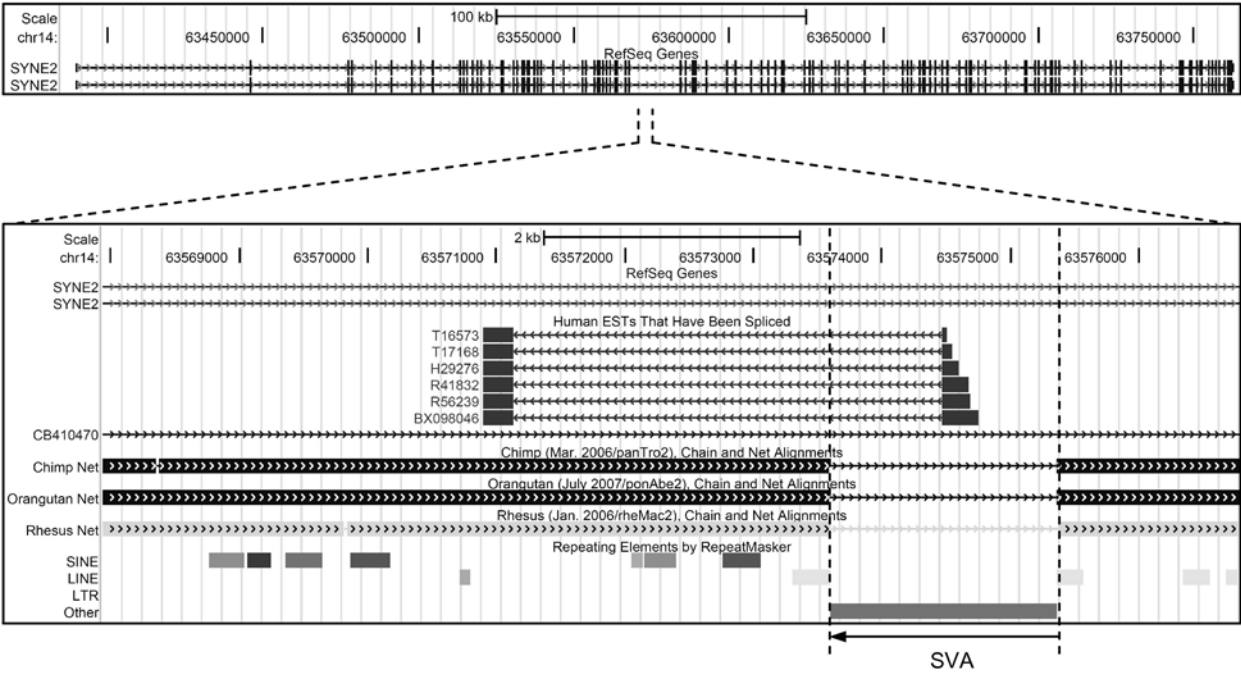


Figure 1. (A) AluYd8-induced human-specific antisense transcription in the *RNF144A* gene. Top panel: The full-length exonic structure of the *RNF144A* gene. Bottom panel: Detail of the promoter region of the *RNF144A* gene including its antisense transcripts. The known gene transcript is labeled with its gene name (the 'RefSeq Genes' track) and the antisense mRNAs and ESTs are labeled with their accession numbers (the 'Human mRNAs from GeneBank' and the 'Human ESTs That Have Been Spliced' tracks). Note that an mRNA and 13 ESTs initiate from the antisense strand of the human-specific AluYd8 element (bounded by vertical dotted lines). The other 4 transcripts (CR619750, AK090737, DA069483 and BX376453) appear to also be derived from the antisense transcriptional unit of which sequence data lack the insert region. The exons and introns are represented by thick and thin lines, respectively. Consecutive less-than signs (<) and greater-than signs (>) on the line indicate the transcription direction or orientation of the chromosome: from right to left and from left to right, respectively. The tracks labeled 'Chimp Net', 'Orangutan Net' and 'Rhesus Net' are human-chimpanzee, human-orangutan and human-rhesus macaque genome alignments, respectively. The thin line on the 'Net' tracks indicates an alignment gap, meaning the corresponding human genomic segment is not found in the subject genome due to a human-specific insertion. The orientation of the insert is indicated by an arrow at the bottom. (B) SVA-derived antisense transcripts in the *SYNE2* gene. The 2 RefSeq transcripts of the *SYNE2* gene are shown at the top. The antisense transcripts derived from the sense strand of the human-specific SVA element are positioned in the intron 45.

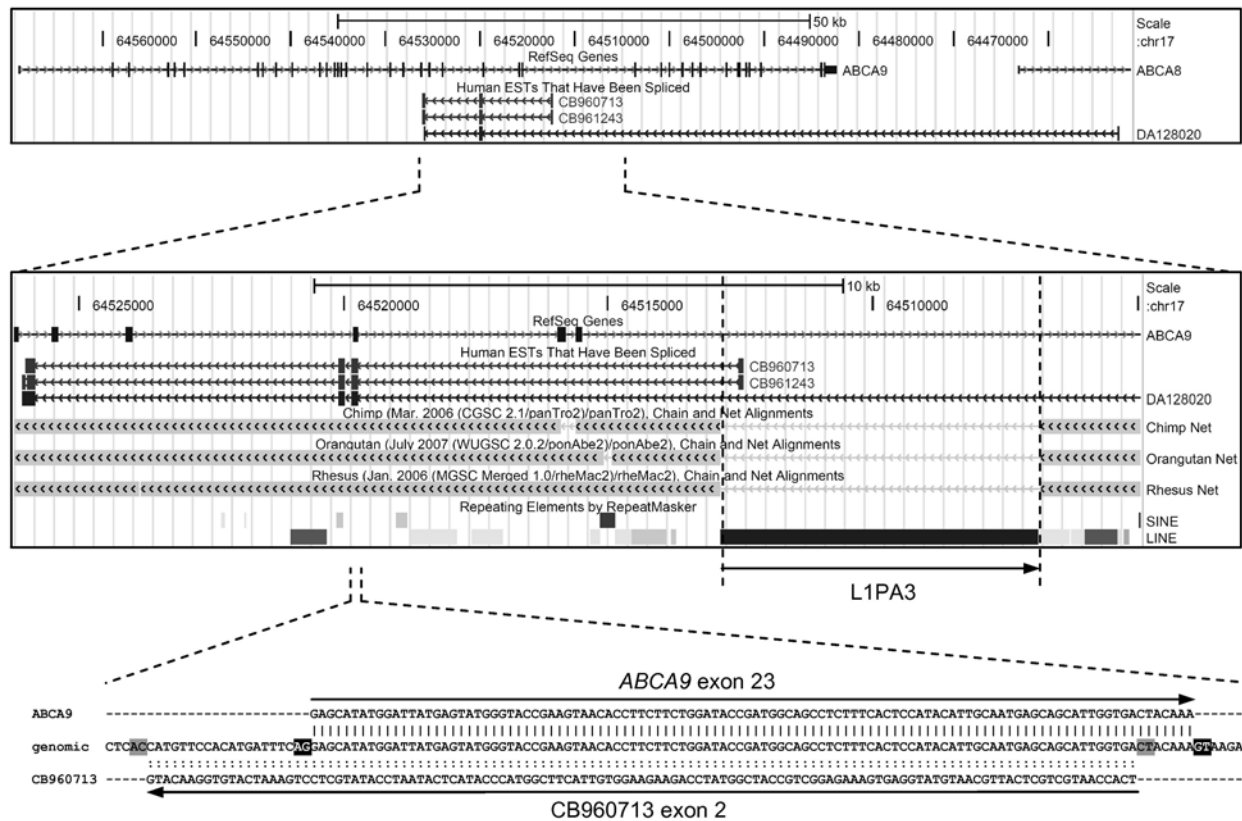
A *ABCA9*

Figure 2. Antisense transcripts in the *ABCA9* gene. (A) The exonic structures of the *ABCA9* gene and the 3 antisense ESTs are shown at the top panel. Note that the longer-spanning antisense EST DA128020 originates from the last intron of the adjacent gene *ABCA8*. The bottom panel shows the detail of the human-specific 2 antisense ESTs CB960713 and CB961243, which are initiated from the antisense promoter of the L1PA3 element. Note that the last 3 exons are shared by all of the 3 antisense ESTs. The orientation of the chromosome is reversed. (B) The second exon of the antisense transcripts overlaps with exon 23 of the *ABCA9* gene in a reversed orientation. The splice donor and acceptor sites for the *ABCA9* exon 23 and for the antisense transcript exon 2 are highlighted by a black and a gray background, respectively. The vertical bars and colons indicate direct matches and inverse complementary matches, respectively.

Modification of existing antisense transcripts. We also identified several cases where the inserted transposable elements interact with existing antisense transcripts, providing a novel promoter, a novel exon cassette or a novel terminal exon to them. In the *ABCA9* gene, there are 2 antisense ESTs (accession nos. CB960713 and CB961243) originating from the antisense promoter of the human-specific L1PA3 element in intron 20 (Fig. 2A). There is another antisense EST (accession no. DA128020) which is initiated from the last intron of the adjacent gene *ABCA8*. All of the 3 antisense ESTs share the last 3 exons out of the 4. It is assumed that DA128020 represents the existing antisense transcript in this locus, and insertion of the L1PA3 element provides an additional novel promoter to it. Notably, the second exon of the antisense transcript almost perfectly overlaps with exon 23 of the *ABCA9* gene in a reversed orientation (Fig. 2B). The mature products of the sense and the antisense transcripts are able to form a double-stranded RNA and may recruit the RNAi machinery (26). If this is the case, the human *ABCA9* gene would show reduced expression compared to its chimpanzee or other primate orthologs due to increased expression of its antisense transcript.

There are a large number of antisense ESTs in intron 2 of the *LHFPL3* gene (Fig. 3). The antisense transcripts undergo

extensive alternative splicing and show highly diverse exonic organization. It is assumed that there is a well-established intronic antisense transcriptional unit in this locus. Among the antisense ESTs, 14 contain a novel exon cassette derived from the 5' part of the human-specific HERVK element. All of the antisense ESTs spanning the HERVK incorporate the novel exon cassette, suggesting that the HERVK exon exhibits a strong exon definition. Another example of a human-specific addition of an exon cassette to an existing antisense transcript is found in the *DSG1* gene locus (Table I). The 3 antisense ESTs of the *DSG1* contain an exon cassette derived from a part of the human-specific L1PA2 element in the last intron. There are 2 additional antisense ESTs which span the insert but do not share the L1PA2-derived exon, indicating that this exon is alternatively spliced.

The Alu, L1 and SVA elements contain cryptic polyadenylation sites and often prematurely truncate the associated transcripts (37,40,41). The antisense EST CN270894 in the *TEX11* gene locus ends within the human-specific SVA element. A splice acceptor site for this exon is located outside the SVA element, which seems to be activated upon insertion. Notably, the terminal exon of the antisense EST overlaps with the *TEX11* gene exon 16 (number based on RefSeq NM_001003811), possibly triggering the RNAi pathway.

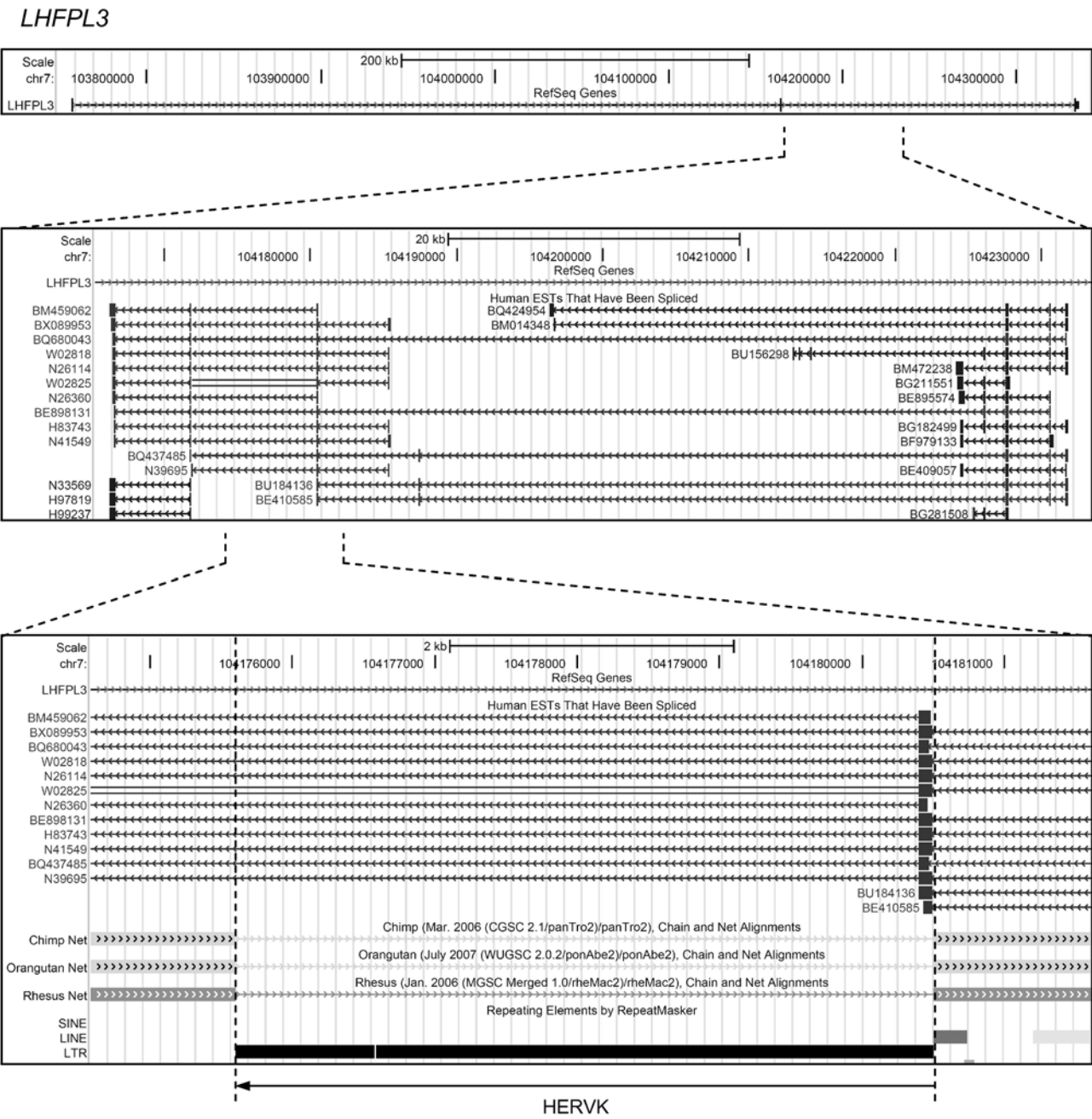


Figure 3. Antisense transcripts in the *LHFPL3* gene. (A) The *LHFPL3* gene is composed of 3 exons (top). In intron 2 many antisense transcripts (middle) are noted. The 14 antisense ESTs contain an exon cassette derived from the 5' part of the human-specific HERVK (bottom).

Another example of the terminal exon addition is the antisense EST DA746847 in the *KCNMB2* gene locus. A polyadenylation signal in the human-specific L1PA2 element is utilized. However, many variants of the antisense transcript in this locus span the inserted L1 and are not affected by the cryptic L1 polyadenylation signal, indicating the signal is rather weak.

In conclusion, we identified 13 cases of the creation and/or the modification of human-specific antisense transcripts induced by transposable element integration. The antisense transcription could have resulted in alteration of the regulation of the sense partner gene expression. We propose that some of these changes may have played a role in human evolution.

Acknowledgements

This research was supported by Chung-Ang University research grants in 2009.

References

1. Enard W, Przeworski M, Fisher SE, *et al*: Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418: 869-872, 2002.
2. Konopka G, Bomar JM, Winden K, *et al*: Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462: 213-217, 2009.
3. Pollard KS, Salama SR, Lambert N, *et al*: An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443: 167-172, 2006.

4. Olson MV: When less is more: Gene loss as an engine of evolutionary change. *Am J Hum Genet* 64: 18-23, 1999.
5. Stedman HH, Kozyak BW, Nelson A, *et al*: Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* 428: 415-418, 2004.
6. Chou HH, Takematsu H, Diaz S, *et al*: A mutation in human CMP-sialic acid hydroxylase occurred after the homo-pan divergence. *Proc Natl Acad Sci USA* 95: 11751-11756, 1998.
7. Martin MJ, Rayner JC, Gagneux P, Barnwell JW and Varki A: Evolution of human-chimpanzee differences in malaria susceptibility: Relationship to human genetic loss of N-glycolylneuraminic acid. *Proc Natl Acad Sci USA* 102: 12819-12824, 2005.
8. Rich SM, Leendertz FH, Xu G, *et al*: The origin of malignant malaria. *Proc Natl Acad Sci USA* 106: 14902-14907, 2009.
9. International Human Genome Sequencing Consortium: Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945, 2004.
10. The Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69-87, 2005.
11. Rhesus Macaque Genome Sequencing and Analysis Consortium: Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222-234, 2007.
12. Berglund J, Pollard KS and Webster MT: Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol* 7: e26, 2009.
13. Yu H, Jiang H, Zhou Q, *et al*: Origination and evolution of a human-specific transmembrane protein gene, *c1orf37-dup*. *Hum Mol Genet* 15: 1870-1875, 2006.
14. Knowles DG and McLysaght A: Recent de novo origin of human protein-coding genes. *Genome Res* 19: 1752-1759, 2009.
15. Hahn Y and Lee B: Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences. *Bioinformatics* 21 (Suppl 1): i186-i194, 2005.
16. Hahn Y and Lee B: Human-specific nonsense mutations identified by genome sequence comparisons. *Hum Genet* 119: 169-178, 2006.
17. Wang X, Grus WE and Zhang J: Gene losses during human origins. *PLoS Biol* 4: e52, 2006.
18. Hahn Y, Jeong S and Lee B: Inactivation of *MOXD2* and *S100A15A* by exon deletion during human evolution. *Mol Biol Evol* 24: 2203-2212, 2007.
19. Lehner B, Williams G, Campbell RD and Sanderson CM: Antisense transcripts in the human genome. *Trends Genet* 18: 63-65, 2002.
20. Shendure J and Church GM: Computational discovery of sense-antisense transcription in the human and mouse genomes. *Genome Biol* 3: Research0044, 2002.
21. Senner CE and Brockdorff N: Xist gene regulation at the onset of X inactivation. *Curr Opin Genet Dev* 19: 122-126, 2009.
22. Rougeulle C and Heard E: Antisense RNA in imprinting: Spreading silence through air. *Trends Genet* 18: 434-437, 2002.
23. Werner A and Sayer JA: Naturally occurring antisense RNA: Function and mechanisms of action. *Curr Opin Nephrol Hypertens* 18: 343-349, 2009.
24. Shearwin KE, Callen BP and Egan JB: Transcriptional interference - a crash course. *Trends Genet* 21: 339-345, 2005.
25. Osato N, Suzuki Y, Ikeo K and Gojobori T: Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics* 176: 1299-1306, 2007.
26. Meister G and Tuschl T: Mechanisms of gene silencing by double-stranded RNA. *Nature* 431: 343-349, 2004.
27. Tufarelli C, Stanley JA, Garrick D, *et al*: Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 34: 157-165, 2003.
28. Deininger PL, Moran JV, Batzer MA and Kazazian HH Jr: Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 13: 651-658, 2003.
29. Dewannieux M, Harper F, Richaud A, *et al*: Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res* 16: 1548-1556, 2006.
30. Kazazian HH Jr: Mobile elements: Drivers of genome evolution. *Science* 303: 1626-1632, 2004.
31. Conley AB, Miller WJ and Jordan IK: Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet* 24: 53-56, 2008.
32. Rhead B, Karolchik D, Kuhn RM, *et al*: The UCSC genome browser database: Update 2010. *Nucleic Acids Res* 38: D613-D619, 2010.
33. Lee J, Cordaux R, Han K, *et al*: Different evolutionary fates of recently integrated human and chimpanzee line-1 retrotransposons. *Gene* 390: 18-27, 2007.
34. Mathews LM, Chi SY, Greenberg N, Ovchinnikov I and Swergold GD: Large differences between line-1 amplification rates in the human and chimpanzee lineages. *Am J Hum Genet* 72: 739-748, 2003.
35. Mills RE, Bennett EA, Iskow RC and Devine SE: Which transposable elements are active in the human genome? *Trends Genet* 23: 183-191, 2007.
36. Mills RE, Bennett EA, Iskow RC, *et al*: Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* 78: 671-679, 2006.
37. Hancks DC, Ewing AD, Chen JE, Tokunaga K and Kazazian HH Jr: Exon-trapping mediated by the human retrotransposon SVA. *Genome Res* 19: 1983-1991, 2009.
38. Wang H, Xing J, Grover D, *et al*: SVA elements: A hominid-specific retroposon family. *J Mol Biol* 354: 994-1007, 2005.
39. Speek M: Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* 21: 1973-1985, 2001.
40. Wheelan SJ, Aizawa Y, Han JS and Boeke JD: Gene-breaking: A new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* 15: 1073-1078, 2005.
41. Chen C, Ara T and Gautheret D: Using Alu elements as polyadenylation sites: A case of retroposon exaptation. *Mol Biol Evol* 26: 327-334, 2009.