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 and drove transcription of the adjacent genomic segment, creating a novel antisense transcriptional unit (for example, 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_abellii/), 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
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 SYN2E 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 SYN2E.

Other notable cases of transposable element-initiated antisense transcription include the CAMK4, LSAMP, GLRA3 and ATP6V1G1 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.

<table>
<thead>
<tr>
<th>No</th>
<th>Contribution to antisense transcript</th>
<th>Human genome position (hg18)</th>
<th>Origin of insert</th>
<th>Number of antisense transcripts</th>
<th>Accession numbers of antisense transcripts</th>
<th>Partner gene name</th>
<th>Partner gene accession number</th>
<th>Partner gene title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Promoter</td>
<td>chr2:6976077-6976372</td>
<td>AluYd8</td>
<td>14</td>
<td>AK057135 BI755810 BU155881 BX103200 BX353322 DA012236 DA013141 DA071863 DA104120 DA342957 DA362957 DA759273 DA931401 DN989632</td>
<td>RNF144A</td>
<td>NM_014746</td>
<td>Ring finger protein 144A</td>
</tr>
<tr>
<td>2</td>
<td>Promoter</td>
<td>chr3:117561525-117567573</td>
<td>L1Hs</td>
<td>3</td>
<td>AL288335 BF509771 BX10108</td>
<td>LSAMP</td>
<td>NM_002339</td>
<td>Limbic system-associated membrane protein</td>
</tr>
<tr>
<td>3</td>
<td>Terminal exon</td>
<td>chr3:179982257-179984497</td>
<td>L1PA2</td>
<td>1</td>
<td>DA746847</td>
<td>KCNMB2</td>
<td>NM_181361</td>
<td>Calcium-activated potassium channel β2 subunit</td>
</tr>
<tr>
<td>4</td>
<td>Promoter</td>
<td>chr4:175936178-175942219</td>
<td>L1PA2</td>
<td>1</td>
<td>BF028106</td>
<td>GLRA3</td>
<td>NM_006529</td>
<td>Glycine receptor γ3</td>
</tr>
<tr>
<td>5</td>
<td>Promoter</td>
<td>chr5:10665810-110878183</td>
<td>L1Hs</td>
<td>2</td>
<td>DA454582 DB019405</td>
<td>CAMK4</td>
<td>NM_001744</td>
<td>Calcium/calmodulin-dependent protein kinase IV</td>
</tr>
<tr>
<td>6</td>
<td>Terminal exon</td>
<td>chr6:78483373-78492801</td>
<td>HERVK</td>
<td>1</td>
<td>DB094606</td>
<td>-</td>
<td>DR731261</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>Internal exon</td>
<td>chr7:104175600-104180502</td>
<td>HERVK</td>
<td>14</td>
<td>BE410585 BE898131 BM459062 BQ437485 BQ680043 BU184136 BX089953 H83743 N26114 N26360 N39695 N41549 W02818 W02825</td>
<td>LHFPL3</td>
<td>NM_199000</td>
<td>Lipoma HMGIC fusion partner-like 3</td>
</tr>
<tr>
<td>8</td>
<td>Promoter</td>
<td>chr9:16399281-16399281</td>
<td>AluY</td>
<td>1</td>
<td>BU949417</td>
<td>ATP6V1G1</td>
<td>NM_004888</td>
<td>Vacuolar H+ ATPase G1</td>
</tr>
<tr>
<td>9</td>
<td>Promoter</td>
<td>chr10:130008004-130008314</td>
<td>AluYb8</td>
<td>2</td>
<td>AI149915 BX089661</td>
<td>-</td>
<td>AK124226</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Promoter</td>
<td>chr14:63573593-63573560</td>
<td>SVA</td>
<td>6</td>
<td>BX098046 H29276 R41832 R56239 T16573 T17168</td>
<td>SYNE2</td>
<td>NM_015180</td>
<td>Spectrin repeat containing nuclear envelope 2</td>
</tr>
<tr>
<td>11</td>
<td>Promoter</td>
<td>chr17:64506845-645512885</td>
<td>L1PA3</td>
<td>2</td>
<td>CB960713 CB961243</td>
<td>ABCA9</td>
<td>NM_080283</td>
<td>ATP-binding cassette sub-family A member 9</td>
</tr>
<tr>
<td>12</td>
<td>Internal exon</td>
<td>chr18:27181918-27187963</td>
<td>L1PA2</td>
<td>3</td>
<td>AW468896 DB461091 DB526502 CN270894</td>
<td>DSG1</td>
<td>NM_001942</td>
<td>Desmoglein 1 preproprotein</td>
</tr>
<tr>
<td>13</td>
<td>Terminal exon</td>
<td>chrX:69815491-69817169</td>
<td>SVA</td>
<td>1</td>
<td>-</td>
<td>TEX11</td>
<td>NM_031276</td>
<td>Testis expressed sequence 11</td>
</tr>
</tbody>
</table>
Figure 1. (A) AluYd8-induced human-specific antisense transcription in the RNF144A gene. Top panel: The full-length exonic structure 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.
Modification of existing antisense transcripts. We also identified several cases where the inserted transposable elements interact with existing antisense transcripts, providing 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.
Another example of the terminal exon addition is the antisense EST DA746847 in the \textit{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