

Research Article

L1 Antisense Promoter Drives Tissue-Specific Transcription of Human Genes

Kert Mätlik, Kaja Redik, and Mart Speek

Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, Tallinn 19086, Estonia

Received 26 July 2005; Revised 11 November 2005; Accepted 16 November 2005

Transcription of transposable elements interspersed in the genome is controlled by complex interactions between their regulatory elements and host factors. However, the same regulatory elements may be occasionally used for the transcription of host genes. One such example is the human L1 retrotransposon, which contains an antisense promoter (ASP) driving transcription into adjacent genes yielding chimeric transcripts. We have characterized 49 chimeric mRNAs corresponding to sense and antisense strands of human genes. Here we show that L1 ASP is capable of functioning as an alternative promoter, giving rise to a chimeric transcript whose coding region is identical to the ORF of mRNA of the following genes: *KIAA1797*, *CLCN5*, and *SLCO1A2*. Furthermore, in these cases the activity of L1 ASP is tissue-specific and may expand the expression pattern of the respective gene. The activity of L1 ASP is tissue-specific also in cases where L1 ASP produces antisense RNAs complementary to *COL11A1* and *BOLL* mRNAs. Simultaneous assessment of the activity of L1 ASPs in multiple loci revealed the presence of L1 ASP-derived transcripts in all human tissues examined. We also demonstrate that L1 ASP can act as a promoter *in vivo* and predict that it has a heterogeneous transcription initiation site. Our data suggest that L1 ASP-driven transcription may increase the transcriptional flexibility of several human genes.

Copyright © 2006 Kert Mätlik et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Non-LTR and LTR retrotransposons are the two most abundant classes of transposable elements that contain regulatory regions (promoter, enhancer, and polyadenylation signal) necessary for their transcription and transposition [9]. Although most of the non-LTR retrotransposons and all the LTR retrotransposons in the human genome have lost their transpositional competence due to broken ORFs, a large number of them have retained regulatory sequences [10]. Scattered all over the chromosomes, retrotransposons can affect the regulation of host genes' transcription.

Recent studies carried out in several laboratories have revealed that LTR retrotransposons, such as an intracisternal A-particle in mice [11], endogenous retroviruses in humans and mice [12], and *Wis 2-1A* in wheat [13], can influence transcription of adjacent genes. Similarly, two families of non-LTR retrotransposons, L1 [3] and B2 SINE [14], have been shown to drive transcription of human and mouse genes, respectively. It has been shown that the effect of retrotransposons on the host gene expression depends on their epigenetic status and thus may cause phenotypic variation between genetically identical individuals [15].

Retrotransposons may provide alternative promoters for host genes. Here, the known examples include LTR-mediated transcription of *Agouti* [16], *PTN* [17], *apoC-I*, *EDNRB* [18], *CYP19* [19], β 3GAL-T5 [20], and *SPAMI* genes [21]. We have previously shown that in transformed cells many human genes are transcribed from the L1 ASPs located in introns of these genes [2, 3].

To reveal the possible function of L1 ASP as an alternative promoter of human genes, we carried out a systematic search for additional chimeric L1 ESTs/mRNAs deposited in GenBank. Here we describe 49 chimeric mRNAs generated by L1 ASP-driven transcription. Four of these chimeras differ from the bona fide mRNAs by 5' untranslated region (UTR) and another four (antisense RNAs) have regions complementary to exons of known mRNAs. Based on these bioinformatic data, we show that L1 ASP is capable of functioning as an alternative promoter in normal human tissues and drives tissue-specific transcription of several human genes.

METHODS

Computational analysis

The search and analysis of chimeric L1 transcript sequences derived from the human subset of EST division of GenBank,

TABLE 1: Primers used for the detection of mRNAs, chimeric mRNAs, and L1 splice forms.

mRNA	Forward primer	Reverse primer	Annealing temperature
AL711955	CTGTGGCAGAAGGGAGAAG	GCAGCAGAGAGGACTTTGG	65°C
(L1) <i>KIAA1797</i>	TCTCAGACTGCTGTGCTA	GCAGCAGAGAGGACTTTGG	60°C
<i>CLCN5</i> (uP)	GGAGAAAACAGGGCCACATA	CATGCTCAGAGTTCCAGCAA	60°C
<i>CLCN5</i> (dP)	GACCCTTTTGTCTCCCTTCC	CATGCTCAGAGTTCCAGCAA	60°C
L1- <i>CLCN5</i>	CTGCTGTGCTAGCAATCAGC	CATGCTCAGAGTTCCAGCAA	60°C
<i>SLCO1A2</i>	AAAGCGTTCAGGTATTTTGG	GCTCTTCAGGGTGTTCCTCAAG	55°C
L1- <i>SLCO1A2</i>	CTGCTGTGCTAGCAATCAGC	GCTCTTCAGGGTGTTCCTCAAG	60°C
<i>MET</i>	ACGGTCCAAAGGAACTCT	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-A</i>	CTGCTGTGCTAGCAATCAGC	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-B</i>	CTAAGCAAAGCCTGGGCAATG	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-C</i>	TTCCCGGCTGCTTTGTTTAC	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-D</i>	GGCTCCACCCAGTTCGAGCT	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-E</i>	AGGCAGGCCTCCTTGAGCTCTG	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-F</i>	AGGTGGAGCCTACAGAGGCAG	CCTTGTAGATTGCAGGCAGAC	60°C
L1- <i>MET-G</i>	TGCAGAGGTTACTGCTGTCT	CCTTGTAGATTGCAGGCAGAC	60°C
<i>COL11A1</i>	GGATTTC AAGGCAAGACCG	TTTGCACCTTCTTTTCCTGC	55°C
L1- <i>COL11A1</i>	CTGCTGTGCTAGCAATCAGC	TAGGGTGATCCAGGTCCTCA	60°C
<i>BOLL</i>	CGCAAACATCAAACCAGATG	TACTGTGTGGTGGCCTGGTA	60°C
L1- <i>BOLL</i>	CTGCTGTGCTAGCAATCAGC	GCCTTCAAATGCAGGACTGT	60°C
L1 II sp v ¹	CTCCCCAGCCTCGCTGC	GGTTCATCTCACTGGCTC	60°C
L1 IV sp v ¹	CTGCTGTGCTAGCAATCAGC	GGTTCATCTCACTGGAAA	55°C

¹sp v stand for splice variant

EMBL, and DDBJ) was carried out by using the strategy described earlier [2]. The alignment of EST and mRNA sequences to genomic contigs was done with SPIDEY [1] and confirmed with the human genome browser available at University of California, Santa Cruz [5]. BLAST [6], BLAST2 sequences [7], and SPIDEY programs, used in the analysis of sequences of RT-PCR products, were run on the National Center for Biotechnology Information BLAST network service using default parameters.

The transcriptional start sites in the DBTSS [22] were mapped using the BLASTN [6]. The accession numbers of the respective one-pass cDNA entries were OFR00417, CNR02292, KAR05296, TDR09332, T3R04859, TDR07820, KMR03236, HKR11044, KMR01202, COL02332, KMR02654, TDR05153, TDR04283, T3R08474, T3R07002, TDR08640, DMC04507, HKR03051, T7R06886, T3R04414, 29R05294, OFR01051, T3R00241, and HKR11121. Splice site search was done with NNSPLICE 0.9 [23] and NetGene2 [24].

RT-PCR, Southern blot, and sequence analysis

PCR amplification of the human cDNAs of the multiple tissue cDNA (MTC) panels I and II (BD Biosciences Clontech) was carried out using recombinant Taq polymerase and Taq buffer with (NH₄)₂SO₄, 2.0 mM Mg₂Cl₂, 0.2 mM dNTP

(Fermentas), and 0.75 μM primers. Each reaction contained 0.5 μl cDNA and 0.5 units of Taq polymerase in a final volume of 10 μl. After cDNA denaturation at 95°C for 1 minute, amplification (35–40 cycles) was carried out by using the following cycling profile: 95°C 30 s, 55°–65°C 30 s, and 72°C 30 s for products < 0.5 Kb or 1 minute for products > 0.5 Kb. Primers and annealing temperatures used are given in the supplementary table Table 1. The locations of primers are shown in Figures 1 and 2. PCR products were sized on 1–2% agarose gels and analyzed by restriction mapping. After gel elution, their sequences were determined from both ends using BigDye Terminator cycle sequencing kit (Applied Biosystems).

First strand L1-*MET* cDNA was synthesized with a reverse primer positioned in *MET* exon 5 (TATGGTCAGCCTTGTCCCTC) using total RNA isolated from human teratocarcinoma cell line (N-Tera2D1) and RevertAid H minus M-MuLV reverse transcriptase (Fermentas). This cDNA was denatured at 95°C for 1 minute and amplified (30 cycles, see above) using one of the primer pairs (L1-*MET-A-G*) shown in Table 1. For Southern blot analysis, the RT-PCR products obtained were sized on an agarose gel, transferred to a nylon membrane and hybridized with a riboprobe specific to *MET* exons 2–5. Hybridization-positive products were detected by autoradiography.

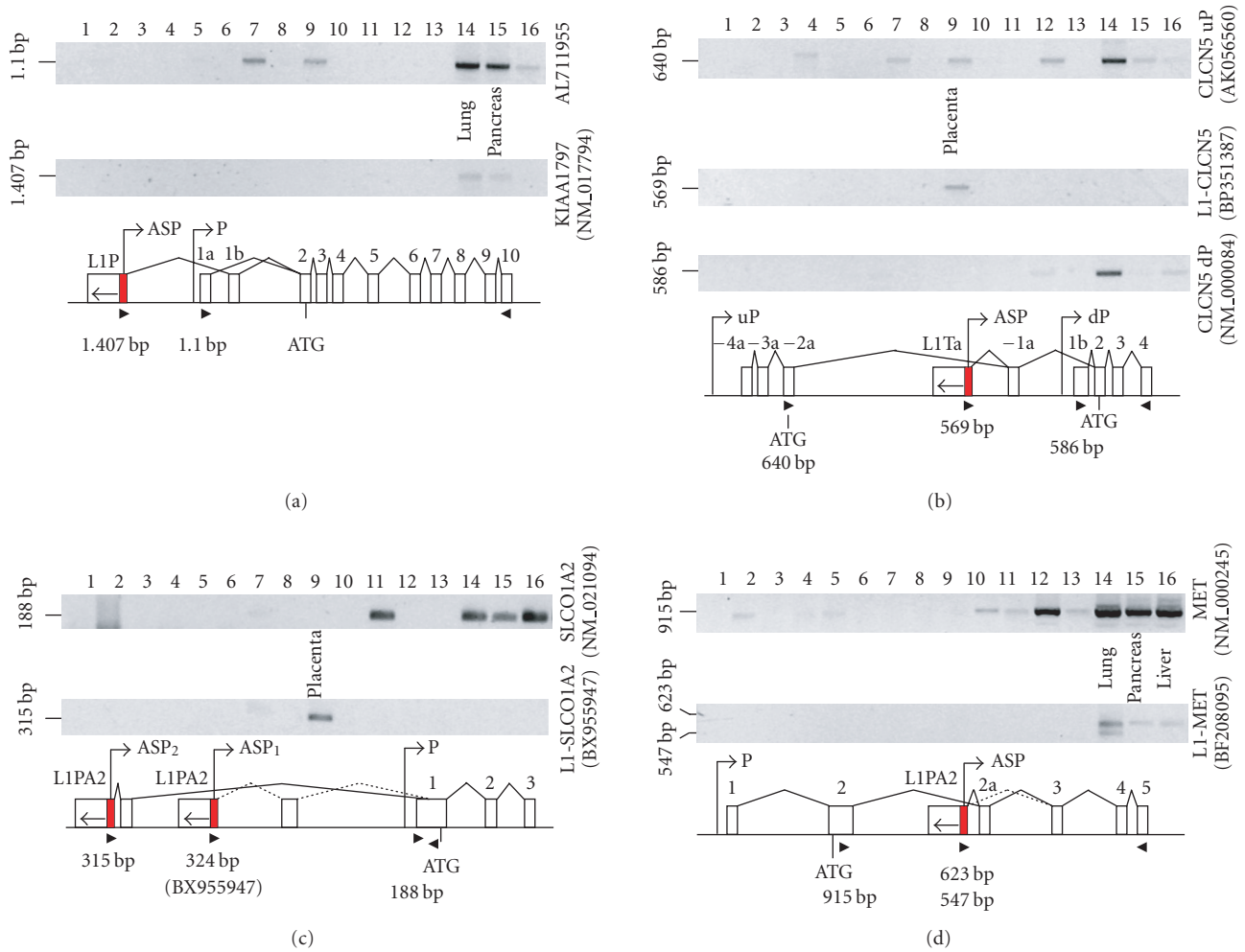


FIGURE 1: Distribution of chimeric mRNAs derived from the L1 ASP as an alternative promoter. The presence of native mRNAs derived from a gene predicted by (a) AL711955 and KIAA1797, (b) CLCN5 [25], (c) SLCO1A2, (d) MET proto-oncogene [26], and their corresponding chimeric transcripts is shown at the upper and lower RT-PCR panels. cDNAs were derived from the following human tissues: 1, thymus; 2, prostate; 3, spleen, 4, small intestine; 5, colon; 6, ovary; 7, testis; 8, peripheral blood leukocytes; 9, placenta; 10, skeletal muscle; 11, brain; 12, kidney; 13, heart; 14, lung; 15, pancreas, and 16, liver. GenBank accession numbers for each mRNA and chimeric L1 mRNA are shown. Product sizes are shown on the left of each panel and below the forward primer on the scheme. L1 (PA2 or Ta subfamily) is shown by a large box with the 5' UTR region indicated in red and its orientation is marked by an arrow. Exons are marked by open boxes (not in scale). Splicing schemes are shown by lines. The location of translation initiation, codon is marked by ATG. Primers used in PCR are shown by arrowheads below the exons. (b) Exons transcribed from the CLCN5 upstream promoter [uP] are designated with -1a to -4a. (c) A 315 bp RT-PCR product corresponds to L1-SLCO1A2 transcript derived from the upstream L1 ASP (L1 ASP₂), but not from the L1 ASP (L1 ASP₁) predicted by the EST (BX955947). (d) A minor L1-MET splice variant is shown by a broken line. P stands for promoter and dP stands for downstream promoter.

RESULTS

L1 ASP is predicted to function as an alternative promoter

We have previously characterized 9 out of 25 ESTs representing the L1 ASP-driven transcription of human genes [2]. Using the strategy described earlier [2] and an updated version of the dbEST (12 May 2004), we extended our search to reveal chimeric transcripts derived from an L1 ASP acting as a sole/alternative promoter or driving antisense transcription

of host gene. Our search revealed 81 ESTs containing the opposite strand of L1 5' UTR, followed by a region identical to a cellular mRNA or random genomic sequence. Of this large number of chimeric transcripts, 49 ESTs represented mRNAs derived from the genes annotated in RefSeq database [8] (see the supplementary table (Table 2)). The remaining 32 ESTs contained noncoding or repetitive DNA sequences (Alus, MIR, LTR, L1, etc) spliced to the L1 5' UTR. Since they contained only short ORFs (< 100 aa) and had no similarity to known proteins, as revealed by BLASTP analysis, they were not analyzed further.

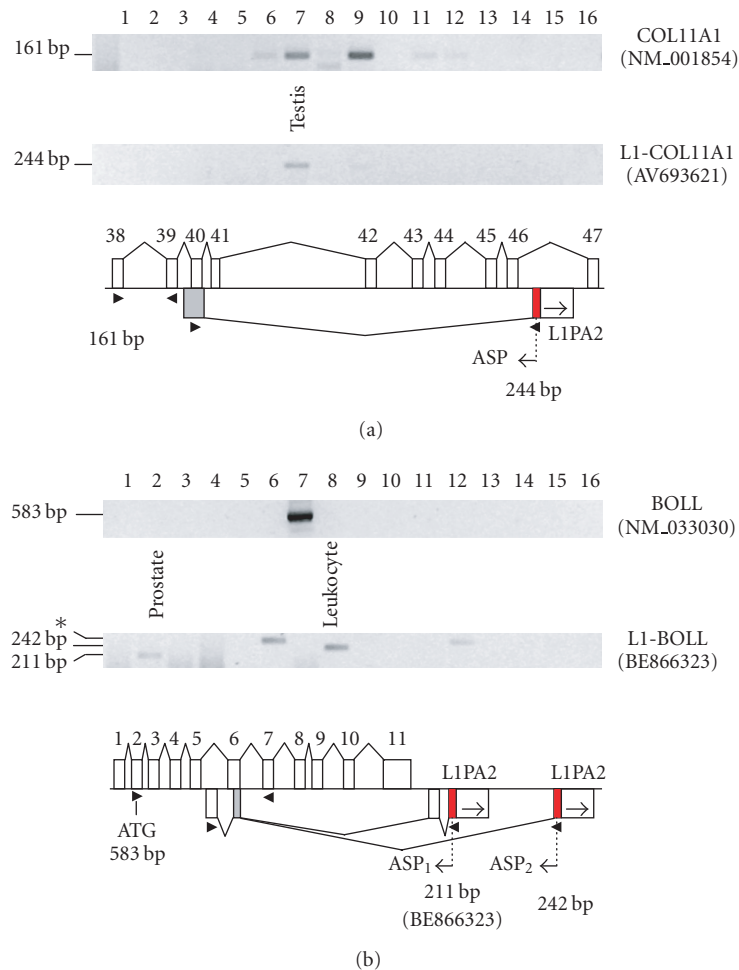


FIGURE 2: Distribution of antisense RNAs derived from L1 ASP. The presence of mRNAs derived from (a) *COL11A1* [27], (b) *BOLL* [28], and their antisense RNAs is shown at the upper and lower RT-PCR panels, respectively. The exons of antisense RNAs *L1-COL11A1* and *L1-BOLL* complementary to exon 40 of *COL11A1* and exon 6 of *BOLL* are shown as grey boxes. (a) *COL11A1* exons 38–47. (b) Two *L1-BOLL* splice variants and a nonspecific product, marked by an asterisk, are presented. A 211 bp product derived from the L1 ASP₁ is identical to EST BE866323 (splicing scheme III). A novel 242 bp product generated from the L1 ASP₂ corresponds to the splicing scheme V. For the remaining description details, see Figure 1 legend.

Because of our interest in the L1 ASP-driven transcription of human genes, we carried out a detailed analysis of the 49 chimeric ESTs (Table 2). While most of the ESTs (40 out of 49) corresponded to mRNAs generated from the L1 ASPs of full-length L1s located in introns, 7 ESTs/mRNAs (NM_017794, BP351387, BM557937, CF593264, BP358215, BX955947, and BU176833) were derived from L1 ASPs located upstream of genes. In these 7 cases, L1 ASP may function as an alternative promoter. Four of these cases (NM_017794, BP351387, BX955947, and BP358215) represented chimeric mRNAs that contained the first coding exon of the gene. Thus, their translation could produce proteins identical to those encoded by the respective gene (Table 3). These genes encoded hypothetical protein KIAA1797 (possibly involved in mitotic chromosome condensation), CLCN5 (chloride channel 5) [25], SLC01A2 (solute carrier organic anion transporter family member 1A2), and RGS6 (regulator

of G-protein signalling 6) [29]. For the remaining three ESTs, splicing occurred within the coding sequence, giving rise to the chimeric mRNA lacking bona fide translation initiation signals. Since translation initiation signals are commonly located in the second exon of mammalian mRNAs [30], an L1 ASP located in the first intron could also give rise to a translatable chimeric mRNA. Of the 3 ESTs (BM910612, BE735854, and BP352155) derived from such L1 ASPs, only one (BE735854) had translation initiation signals matching those of the bona fide mRNA.

Of the 49 ESTs/mRNAs analyzed, 45 chimeras matched the orientation of the respective gene, while 4 ESTs had regions complementary to the exons of known mRNAs and thus were derived from the opposite strand of the gene (Table 3).

Two of these ESTs (CB960713 and AV693621) were derived from the L1 ASPs located in the intron 25 of *ABCA9*

TABLE 2: Widespread L1 ASP-driven transcription of human genes revealed from ESTs/mRNAs.

EST ¹	Source ²	Similarity to L1 5'UTR opposite strand ³	Similarity to known mRNA ⁴	Location in the genome ⁵	Orientation ⁶
Type I splicing (1 EST)					
BU943355 ⁺⁺ (4 ex)	Pool of 40 cell line polyA+	4–59 ≡ 592–647 (96%) 60–289 ≡ 762–990 (96%) L1PA3 AC007780	Arylsulfatase G, NM_014960 331–649 ≡ 1342–1660 (99%)	NT_010641 (chr 17) 10/11	Sense
Type II splicing (2 ESTs)					
CD642260 (4 ex)	Embryonic stem cell line WA01/H1	12–117 ≡ 542–647 (97%) 118–230 ≡ 878–990 (96%) L1PA2 AC022762	Olfactory receptor, family 56, subfamily B, member 4, NM_001005181 373–728 ≡ 802–443 (98%)	NT_009237 (chr 11) 3'/1	Antisense
NM_017794 (46 ex)	RA-induced NT2 neuronal precursor cells	4–150 ≡ 501–647 (93%) 151–262 ≡ 878–990 (93%) L1P AL354879	Hypothetical protein KIAA1797, AL711955* 331–834 ≡ 60–563 (99%)	NT_008413 (chr 9) 5'/45	Sense
Type III splicing (22 ESTs)					
BM910612 (6 ex)	Brain, astrocytoma grade IV cell line	1–134 ≡ 514–647 (98%) L1Ta (Hs) AC011597	Fibronectin type III domain containing 6 (cytokine receptor), NM_144717 268–915 ≡ 336–982 (98%)	NT_086641 (chr 3) 1/7	Sense
BF676152 (3 ex)	Prostate	4–126 ≡ 520–647 (91%) L1PA2 AC097061	Hypothetical protein BC014608, NM_138796 127–713 ≡ 422–1005 (91%)	NT_021877 (chr 1) 5/11	Sense
AU123136 ⁺⁺ (7 ex)	Uninduced NT2 cell line	1–125 ≡ 523–647 (96%) L1PA2 AC079005	Breast carcinoma amplified sequence 3, NM_017679 126–623 ≡ 710–1208 (99%)	NT_010783 (chr 17) 9/24	Sense
AA226814 ⁺ (3 ex)	Ntera-2 neuroepithelial cells	1–111 ≡ 538–649 (93%) L1PA2 AC018470	Secernin 3 (dipeptidase), NM_024583 112–347 ≡ 843–1075 (96%)	NT_005403 (chr 2) 5/8	Sense
BU959632 (5 ex)	Pool of 40 cell line polyA+	4–45 ≡ 606–647 L1Ta (Hs) AC008496	Cardiomyopathy associated 5, NM_153610 46–559 ≡ 3235–3748 (97%)	NT_006713 (chr 5) 8/12	Sense

TABLE 2: Continued.

BF208095 ⁺ (6 ex)	Bladder carcinoma cell line	2–57 ≡ 592–647 (94%) L1PA2 AC002080	Hepatocyte growth factor receptor (MET proto-oncogene), NM_000245 132–456 ≡ 1387–1714 (99%) 462–663 ≡ 1805–2013 (92%)	NT_007927 (chr 7) 2/21	Sense
AA220950 ⁺ (3 ex)	Ntera-2 neuroepithelial cells	1–39 ≡ 609–647 (89%) L1PA3 AC022261	Dynein, cytoplasmic, intermediate polypeptide 1, NM_004411 40–247 ≡ 613–818 (95%)	NT_007910 (chr 7) 5/17	Sense
BM557937 (7 ex)	Brain astrocytoma grade IV cell line	1–110 ≡ 538–647 (93%) L1PA3 AC022748	Cholinergic receptor, nicotinic, beta poly-peptide 4, NM_000750 410–713 ≡ 168–471 (99%)	NT_024654 (chr 15) 5/6	Sense
BG335812 (> 6 ex)	Placenta choriocarcinoma cell line	2–105 ≡ 544–647 (93%) L1PA2 AC009949	Nuclear antigen Sp100, NM_003113 106–522 ≡ 139–556 (90%)	NT_005403 (chr 2) 2/25	Sense
BE865812 ⁺ (4 ex)	Bladder carcinoma cell line	1–43 ≡ 605–647 (97%) L1Ta (Hs) AL049838	Chromosome 14 open reading frame 37, NM_001001872 44–343 ≡ 933–1228 (96%)	NT_025892 (chr 14) 5/7	Sense
BE866323 ⁺ (4 ex)	Bladder carcinoma cell line	1–92 ≡ 556–647 (96%) L1PA2 AC073058 L1PA2 AC020550	Bol, boule-like (<i>Drosophila</i>), NM_033030 145–204 ≡ 790–731 (98%)	NT_005246 (chr 2) 3/11	Antisense
BP352155 (5 ex)	Well-differentiated squamous cell carcinoma cell line TE13	1–113 ≡ 535–647 (96%) L1PA2 AC004519	Hypothetical protein FLJ31340, BX346336* 114–490 ≡ 500–876 (98%)	NT_086723 (chr 7) 1/ > 5	Sense
BP351387 (5 ex)	Well-differentiated squamous cell carcinoma cell line TE13	1–67 = 581–647 L1Ta (Hs) AL663118	Chloride channel 5, NM_000084 213–583 = 243–613	NT_086939 (chr X) 5/12	Sense
BP351082 (> 4 ex)	Well-differentiated squamous cell carcinoma cell line TE13	1–71 ≡ 576–647 (95%) L1PA3 AC114734	Hypothetical protein MGC16169 (protein kinase) NM_033115 72–593 ≡ 1913–2433 (99%)	NT_086651 (chr 4) 17/24	Sense
BP369881 (6 ex)	Testis	1–65 ≡ 581–647 (92%) L1PA3 AL136525	WD repeat and FYVE domain containing 2, NM_052950 66–570 ≡ 460–963 (99%)	NT_086801 (chr 13) 3/12	Sense

TABLE 2: Continued.

AA226765 (3 ex)	Brain Ntera-2 neuroepithelial cells	1–67 ≡ 581–647 (92%) L1PA3 AC025170	Hypothetical protein FLJ35779, NM_152408 68–356 ≡ 480–767 (97%)	NT_086677 (chr 5) 4/11	Sense
CF593264 (> 5 ex)	Placenta	29–95 ≡ 581–647 (95%) L1PA3 AL050323	Phospholipase C, beta 1, NM_182734 174–769 ≡ 103–692 (98%)	NT_011387 (chr 20) 5'/33	Sense
BP873102 (5 ex)	Embryonal kidney cell line = “293”	1–67 ≡ 581–647 (95%) L1PA2 AL022400	RAB GTPase activating protein 1-like, NM_014857 68–583 ≡ 731–1244 (95%)	NT_086598 (chr 1) 4/21	Sense
CD110319 (2 ex)	Placenta “preeclamptic placenta”	25–92 ≡ 580–647 (97%) L1PA2 AC004452	FLJ16237 protein, NM_001004320 93–568 ≡ 428–900 (97%)	NT_086703 (chr 7) 2/13	Sense
BX476029 (5 ex)	Pooled from different tissues	2–77 ≡ 572–647 (93%) L1PA3 AL121946	Polycystic kidney and hepatic disease 1, NM_138694 78–567 ≡ 7273–7762 (99%)	NT_007592 (chr 6) 43/67	Sense
CB960713 (4 ex)	Placenta	30–107 ≡ 570–647 (96%) L1PA3 AC005922	ATP-binding cassette, subfamily A, NM_172386 108–208 = 3283–3183	NT_010641 (chr 17) 25/38	Antisense
CD644604 (3 ex)	Embryonic stem cells, cell line = “WA01”	14–115 ≡ 547–647 (94%) L1PA3 AC022029	Catenin (cadherin-associated protein), alpha 3, NM_013266 116–736 ≡ 755–1375 (98%)	NT_086771 (chr 10) 5/19	Sense
Type IV splicing (1 EST)					
CF594290 (9 ex)	Placenta	29–230 ≡ 531–732 (94%) 231–340 ≡ 878–988 (95%) L1PA2 AC022306	Hypothetical protein FLJ32800, NM_152647 354–451 = 1305–1402 452–780 ≡ 1642–1964 (97%)	NT_010194 (chr 15) 5/16	Sense
Type V splicing (19 ESTs)					
BE787024 ⁺⁺ (3 ex)	Lung large cell carcinoma cell line	17–215 ≡ 533–732 (98%) L1Ta (Hs) AC079750	Activin A receptor, type IC, NM_145259 216–752 ≡ 548–1086 (95%)	NT_005403 (chr 2) 2/9	Sense
BE568884 ⁺ (4 ex)	Bladder carcinoma cell line	1–178 ≡ 554–732 (97%)	CD96 antigen, NM_005816 179–627 ≡ 659–1113 (97%)	NT_086640 (chr 3) 2/15	Sense

TABLE 2: Continued.

BE617461 ⁺⁺ (6 ex)	Colon adenocarcinoma cell line	8–185 ≡ 553–732 (98%) L1PA2 AC092916	RAB3A interacting protein, NM_175625 186–738 ≡ 998–1556 (98%)	NT_086796 (chr 12) 3/10	Sense
BE568818 ⁺ (3 ex)	Bladder carcinoma cell line	1–163 ≡ 570–732 (93%) L1PA2 AC010585	Secretory carrier membrane protein 1, NM_052822 164–516 ≡ 717–1063 (97%)	NT_006713 (chr 5) 6/8	Sense
BU858570 (2 ex)	Pool of 40 cell line polyA+ RNAs	4–166 ≡ 571–732 (93%) L1PA2 AL691464	Guanylate binding protein 1, NM_002053 167–402 ≡ 259–494 (95%)	NT_004686 (chr 1) 2/11	Sense
BF028725 (3 ex)	Bladder carcinoma cell line	2–123 ≡ 612–732 (91%) L1PA2 AC004800	Hypothetical protein FLJ36166, NM_182634 124–264 ≡ 3282–3424 (95%)	NT_086704 (chr 7) 2/21	Sense
AA224229 ⁺ (4 ex)	6 week, differentiated, post-mitotic hNT, neurons	1–94 ≡ 640–732 (98%) L1Ta (Hs) AL365308	Chromosome 6 open reading frame 170, NM_152730 95–430 ≡ 2622–2957 (99%)	NT_086697 (chr 6) 22/30	Sense
BG542212 ⁺⁺ (> 3 ex)	Lung	2–187 ≡ 547–732 (97%) L1Ta (Hs) AC096569	Zinc finger protein 638, NM_014497 188–638 ≡ 3576–4013 (92%)	NT_022184 (chr 2) 18/28	Sense
AV693621 (2 ex)	Hepatocellular carcinoma	1–172 ≡ 559–732 (93%) L1PA2 AL627203	Collagen, type XI, alpha 1, variant A, NM_001854 187–279 = 3433–3341	NT_004623 (chr 1) 46/67	Antisense
BE735854 ⁺ (6 ex)	Pancreas adenocarcinoma cell line	1–95 ≡ 638–732 (93%) L1PA2 AC092903	Similar to beta-1, 4-mannosyltransferase, CD708577* 95–387 ≡ 174–466 (99%)	NT_005588 (chr 3) 1/ > 5	Sense
R64632 (4 ex)	Soares placenta Nb2HP	1–52 = 681–732 L1PA2 AL713859	Hypothetical protein FLJ10986, NM_018291 53–406 ≡ 1319–1671 (98%)	NT_029223 (chr 1) 11/14	Sense
BP352672 (4 ex)	Well-differentiated squamous cell carcinoma cell line TE13	1–126 ≡ 608–732 (94%) L1PA2 AL354711	Chromosome 9 open reading frame 39, NM_017738 127–603 = 152–631	NT_008413 (chr 9) 2/23	Sense
BP358215 (7 ex)	Mammary gland tumor cell line T47D	1–147 ≡ 586–732 (92%) L1PA2 AL391749	Regulator of G-protein signalling 6, NM_004296 148–581 ≡ 188–621 (99%)	NT_026437 (chr 14) 5'/17	Sense

TABLE 2: Continued.

H72033 (4 ex)	Soares breast 2NbHBst	1–107 ≡ 626–732 (97%) L1PA2 AC079005	Breast carcinoma amplified sequence 3, NM_017679 108–370 ≡ 710–967 (95%)	NT_010783 (chr 17) 9/24	Sense
CA488981 (3 ex)	Cell_line=ZR-75- 1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1, LNCaP	1–159 ≡ 574–732 (91%) L1PA2 AC034215	Monogenic, audiogenic seizure susceptibility 1 homolog, NM_032119 160–736 ≡ 17956–18532 (99%)	NT_086677 (chr 5) 83/98	Sense
BX955947 (3 ex)	Pooled from different tissues	1–116 ≡ 617–732 (89%) L1PA2 AC006559	Solute carrier organic anion transporter family, member 1A2, NM_021094 240–342 = 186–288	NT_009714 (chr 12) 5'/14	Sense
BX477512 ⁺⁺ (3 ex)	Pooled from different tissues	2–129 ≡ 605–732 (93%) L1PA2 AC024061	Hypothetical protein FLJ38736, NM_182758 130–551 = 3191–3216	NT_086827 (chr 15) 18/20	Sense
CN412489 ⁺⁺ (2 ex)	Embryonic stem cells, embryoid bodies from H1, 7 and H9 cell lines	1–151 ≡ 582–732 (98%) L1PA2 AL133299	FLJ46156 protein, NM_198499 152–348 = 1087–1283	NT_086806 (chr 14) 8/37	Sense
CN408255 (4 ex)	Embryonic stem cells, DMSO-treated H9 cell line	1–180 ≡ 553–732 (95%) L1PA2 AP00942	Baculoviral IAP repeat-containing 2, NM_001166 181–514 = 2766–3099	NT_033899 (chr 11) 6/9	Sense
Type VI splicing (4 ESTs)					
CD643062 (8 ex)	Embryonic stem cell line WA01/H1	10–220 ≡ 780–990 (97%) L1PA2 AC018741	Hypothetical LOC388927, XM_371478 237–744 ≡ 1–509 (99%)	NT_015926 (chr 2) ND	Sense
BU176833 (6 ex)	Eye retinoblastoma cell line	1–227 ≡ 763–989 (96%) L1PA3 AC105054	Rho GTPase activating protein 25, NM_014882 536–878 ≡ 419–757 (97%)	NT_022184 (chr 2) 5'/10	Sense
BE568192 (3 ex)	Bladder carcinoma cell line	1–60 ≡ 931–990 (98%) L1PA2 AP005264	Similar to hypothetical protein LOC375127, XM_496265 95–367 ≡ 213–490 (95%)	NT_010859 (chr 18) 3/5	Sense

TABLE 2: Continued.

BP245205 (3 ex)	Embryonal kidney cell line 293	6–135 ≡ 861–990 (95%) L1PA2 AC099512	Monogenic, audiogenic seizure susceptibility 1 homolog, NM_032119 138–574 ≡ 17953–18384 (98%)	NT_086677 (chr 5) 91/98	Sense
-----------------	--------------------------------------	--	--	----------------------------	-------

¹ EST/mRNA GenBank accession number and number of exons (ex) determined by SPIDEY [1]. ESTs are grouped according to 6 different splicing schemes [2]. Sixteen identical or similar ESTs described earlier by Nigumann et al [2] and Wheelan et al [44] are shown by + and ++, respectively.

² Source of the EST as annotated in EST division of GenBank.

³ EST similarity (≡) or identity (=) to a representative L1 genomic clone #11A [3]. Subfamily of L1 [4] and GenBank accession number were determined by genome browser [5]. For some ESTs the 5' nucleotides (< 28 nt) were derived either from vector/adaptor or represented as low quality sequence.

⁴ Similarity/identity to known mRNA as determined by BLASTN [6] and BLAST2 sequences [7] programs. mRNA description is based on the RefSeq database [8]. If the mRNA has not been described, an EST (marked by an asterisk) is shown. This EST contains a putative first exon transcribed from the non-L1 (native) promoter.

⁵ Genomic contig (accession no), chromosome (chr), and position of the L1 ASP in the intron, upstream (5') or downstream (3')/total number of exons, as determined with MegaBLAST and SPIDEY programs. ND stands for not determined.

⁶ Orientation with respect to the gene's transcription.

(ATP-binding cassette, subfamily A, member 9) [31] and intron 46 of *COL11A1* (collagen type XI alpha 1) [27], respectively (Table 3). The remaining two ESTs (CD642260 and BE866323) were derived from L1 ASPs located downstream of the gene. One of these L1 ASPs resided 77 Kb downstream of the single exon gene encoding olfactory receptor, family 56, subfamily B, member 4 (*OR56B4*) [32] and the other located 34 Kb downstream of *BOLL*, homologous to the *bol* or *boule*-like gene of *Drosophila* [28].

L1 ASP provides an alternative promoter for several human genes

To reveal the potential of L1 ASP to function as an alternative promoter, we determined the expression profile of the chimeric mRNAs (containing bona fide translation initiation signals) in 16 different human tissues. For comparison, we also determined transcription from the native promoters (genes' true promoters). Results for the three chimeric mRNAs (*KIAA1797*, *L1-CLCN5*, and *L1-SLCO1A2*) which were detected in the tissues studied are presented in the following section.

Figure 1(a) shows that both the chimeric *KIAA1797* mRNA, derived from the L1 ASP located about 26 Kb upstream of the first exon of gene, and the native mRNA (the 5' end of the mRNA was predicted from EST AL711955) are expressed in lung and pancreas. In addition, native mRNA is expressed in testis, placenta, and liver.

Figure 1(b) shows that the chimeric *L1-CLCN5* mRNA is expressed exclusively in placenta, while *CLCN5* mRNAs derived from the upstream and downstream promoters (located about 102 Kb and 44 Kb from the L1 ASP, resp) produce mRNAs expressed strongly in lung. Translation of the chimeric mRNA could yield a protein identical to the one

obtained from the *CLCN5* mRNA derived from the downstream promoter. However, the latter is inactive in placenta suggesting that the L1 ASP provides placenta-specific expression to one of the protein isoforms encoded by *CLCN5*. The other protein isoform has a 70 aa-long N-terminal extension and is derived from an mRNA generated from the *CLCN5* upstream promoter. This promoter is active in a number of tissues.

Figure 1(c) shows that the chimeric *L1-SLCO1A2* mRNA predicted from the EST (BX955947) is derived from the L1 ASP located 61 Kb upstream of the *SLCO1A2* first exon. Surprisingly, RT-PCR yielded a 315 bp product (instead of the expected 324 bp product) derived from another L1 ASP located about 24 Kb further upstream. This novel chimeric mRNA is expressed exclusively in placenta, while *SLCO1A2* mRNA is present in a number of tissues, but not in placenta. Therefore, similarl to *CLCN5*, L1 ASP is responsible for the placenta-specific expression of *SLCO1A2*.

Since the multiple tissue cDNA panel has been produced using different donors for different tissues (brain and lung pooled from 2 donors and other tissues pooled from 4–45 donors, except leukocytes which were pooled from 550 donors; the total number of donors was ~750), it is conceivable that an RT-PCR product represents a donor-specific L1 insertion rather than tissue specific activity of the L1 ASP in that chromosomal position. Sequence analysis showed that only one of the L1 elements (*L1-CLCN5*), for which the tissue-specificity of L1 ASP activity was examined (Figures 1 and 2), belongs to the highly polymorphic L1Ta subfamily [33]. The rest of the L1 elements, depicted in Figures 1 and 2, belong to the L1PA2 subfamily that expanded before the divergence of hominids [34], although some polymorphic insertions have been reported in humans [35]. It is unlikely that an L1 insertion is found in only one of the ~750 donors

TABLE 3: Examples of the L1 ASP functioning as an alternative promoter or driving antisense transcription of human genes.

EST ¹	Source ²	Similarity to L1 5'UTR opposite strand ³	Similarity to known mRNA ⁴	Location in the genome ⁵	Orientation ⁶
Type II splicing					
CD642260 (4 ex)	Embryonic stem cell line WA01/H1	12–117 ≡ 542–647 (97%) 118–230 ≡ 878–990 (96%) L1PA2 AC022762	Olfactory receptor, family 56, subfamily B, member 4, NM_001005181 373–728 ≡ 802–443 (98%)	NT_009237 (chr 11) 3'/1	Antisense
NM_017794 (46 ex)	RA-induced NT2 neuronal precursor cells	4–150 ≡ 501–647 (93%) 151–262 ≡ 878–990 (93%) L1P AL354879	Hypothetical protein KIAA1797, AL711955* 331–834 ≡ 60–563 (99%)	NT_008413 (chr 9) 5'/45	Sense
Type III splicing					
BE866323 ⁺ (4 ex)	Bladder carcinoma cell line	1–92 ≡ 556–647 (96%) L1PA2 AC073058 L1PA2 AC020550	Bol, boule-like (<i>Drosophila</i>), NM_033030 145–204 ≡ 790–731 (98%)	NT_005246 (chr 2) 3'/11	Antisense
BP351387 (5 ex)	Well-differentiated squamous cell carcinoma cell line TE13	1–67 = 581–647 L1Ta (Hs) AL663118	Chloride channel 5, NM_000084 213–583 = 243–613	NT_086939 (chr X) 5'/12	Sense
CB960713 (4 ex)	Placenta	30–107 ≡ 570–647 (96%) L1PA3 AC005922	ATP-binding cassette, subfamily A, NM_172386 108–208 = 3283–3183	NT_010641 (chr 17) 25/38	Antisense
Type V splicing					
AV693621 (2 ex)	Hepatocellular carcinoma	1–172 ≡ 559–732 (93%) L1PA2 AL627203	Collagen, type XI, alpha 1, variant A, NM_001854 187–279 = 3433–3341	NT_004623 (chr 1) 46/67	Antisense
BP358215 (7ex)	Mammary gland tumor cell line T47D	1–147 ≡ 586–732 (92%) L1PA2 AL391749	Regulator of G-protein signalling 6, NM_004296 148–581 ≡ 188–621 (99%)	NT_026437 (chr 14) 5'/17	Sense

TABLE 3: Continued.

BX955947 (3 ex)	Pooled from different tissues	1–116 ≡ 617–732 (89%) L1PA2 AC006559	Solute carrier organic anion transporter family, member 1A2, NM_021094 240–342 = 186–288	NT_009714 (chr 12) 5'/14	Sense
--------------------	----------------------------------	--	--	-----------------------------	-------

¹ EST/mRNA GenBank accession number and number of exons (ex) determined by SPIDEY [1]. ESTs are grouped according splicing schemes [2]. EST described earlier by Nigumann et al [2] is marked by +.

² Source of the EST as annotated in EST division of GenBank.

³ EST similarity (≡) or identity (=) to a representative L1 genomic clone #11A [3]. Subfamily of L1 [4] and GenBank accession number were determined by genome browser [5]. For some ESTs, the 5' nucleotides (< 28 nt) were either derived from vector/adaptor or represented as low quality sequence.

⁴ Similarity/identity to known mRNA as determined by BLASTN [6] and BLAST2 sequences [7] programs. mRNA description is based on the RefSeq database [8]. If the mRNA has not been described, an EST (marked by an asterisk) is shown. This EST contains a putative first exon transcribed from the non-L1 (native) promoter.

⁵ Genomic contig (accession no), chromosome (chr), and position of the L1 ASP in the intron, upstream (5'), or downstream (3')/total number of exons, as determined with MegaBLAST and SPIDEY programs. ND stands for not determined.

⁶ Orientation with respect to the gene's transcription.

represented in the MTC panel while it is present in GenBank (Table 3) and Ntera2D1 cell line (data not shown). Therefore we believe that the RT-PCR products obtained represent tissue-specific L1 ASP activity of fixed or high frequency L1 insertions.

In summary, the examples analyzed here provide evidence that L1 ASP can function as an alternative promoter in normal human tissues. Our results show that the L1 ASP-driven transcription correlates with that of the respective native promoter (Figure 1(a)) or expands the tissue-specific expression pattern of the respective gene (Figures 1(b) and 1(c)).

Although our primary goal was to reveal the potential of L1 ASP as an alternative promoter that generates translatable mRNAs, we also determined the distribution of the chimeric L1-MET mRNA derived from the L1 ASP located in the second intron of the MET proto-oncogene [26]. Figure 1(d) shows that the expression of the chimeric L1-MET mRNA correlates with that of the MET mRNA.

L1 ASP generates antisense transcripts complementary to different mRNAs

Of the 49 chimeric ESTs analyzed, only four corresponded to mRNAs that contained regions complementary to the exons of known mRNAs (see above). The expression data are presented for only those two so-called antisense RNAs which were detected in the human tissues examined.

Figure 2(a) shows that the chimeric L1-COL11A1 mRNA, derived from the L1 ASP located in the intron 46 of COL11A1, is expressed in testis and to a lesser extent in placenta. Similarly, COL11A1 mRNA is present in these tissues. It should be noted that L1-COL11A1 (EST: AV693621) contains a 90 nt region complementary to the entire exon 40 of COL11A1 (Table 3).

Figure 2(b) shows that two alternatively spliced variants of the chimeric L1-BOLL, derived from the L1 ASPs located about 34 Kb and 87 Kb downstream of BOLL, are expressed

in prostate and peripheral blood leukocytes, respectively. The 5' ends of these transcripts are spliced according to splicing schemes III and V [2]. BOLL mRNA is expressed exclusively in testis. L1-BOLL contains a 60 nt region complementary to the 3' part of exon 6 of BOLL (Table 3). These results suggest that L1 ASP-driven antisense transcription has no general correlation with the transcription of the host gene.

L1 ASP-derived transcripts are present in all human tissues examined

Our study revealed that chimeric transcripts derived from the six unique genomic regions are present only in a few tissues. To examine the tissue specificity of L1 ASP activity more generally, we studied tissue-specific distribution of L1 ASP-derived transcripts, in which splicing occurs within the L1 5' UTR (splice variants II and IV) [2]. The use of these splice variants allowed us to discriminate between the L1 ASP-derived spliced transcripts and transcripts passing through the whole L1 5' UTR. Figure 3 shows that the splice variant II is expressed in most human tissues, except in thymus, skeletal muscle, and brain. The variant IV shows a more uniform expression pattern with minimal expression in placenta, skeletal muscle, and brain. In summary, these results show that L1 ASP-derived transcripts are present in all human tissues examined.

L1 ASP-driven transcription is characterized by heterogeneous start site

The fact that the sequence corresponding to the opposite strand of L1 5' UTR is present in the EST or mRNA sequence (Table 2) does not necessarily mean that transcription is initiated in the L1 ASP region, that is, in the L1 5' UTR around positions +400 to +600 [3]. In order to find evidence that the L1 ASP region acts as a promoter in vivo, we analyzed the database of transcriptional start sites (DBTSS) [22] for the presence of transcriptional start sites (TSS) which

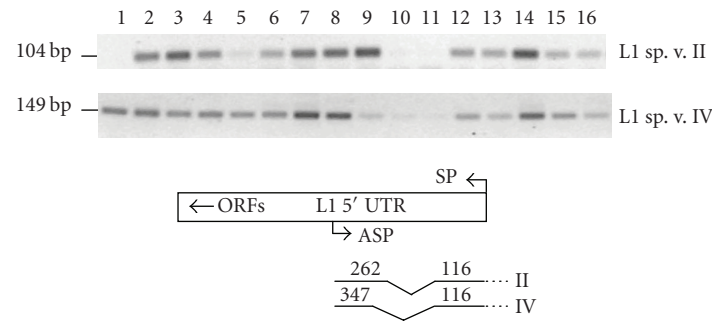


FIGURE 3: Distribution of L1 splice variants II and IV. The presence of splice variants was estimated by RT-PCR in 16 normal human tissues (numbered as in Figure 1 legend) using a reverse primer designed to hybridize to the junction of exons 1 and 2. The schematically represented splice variants II and IV use a common splicing acceptor site at position +116 and splicing donor sites located at positions +262 and +347, respectively [2]. SP stands for L1 sense promoter; sp v stands for splice variant.

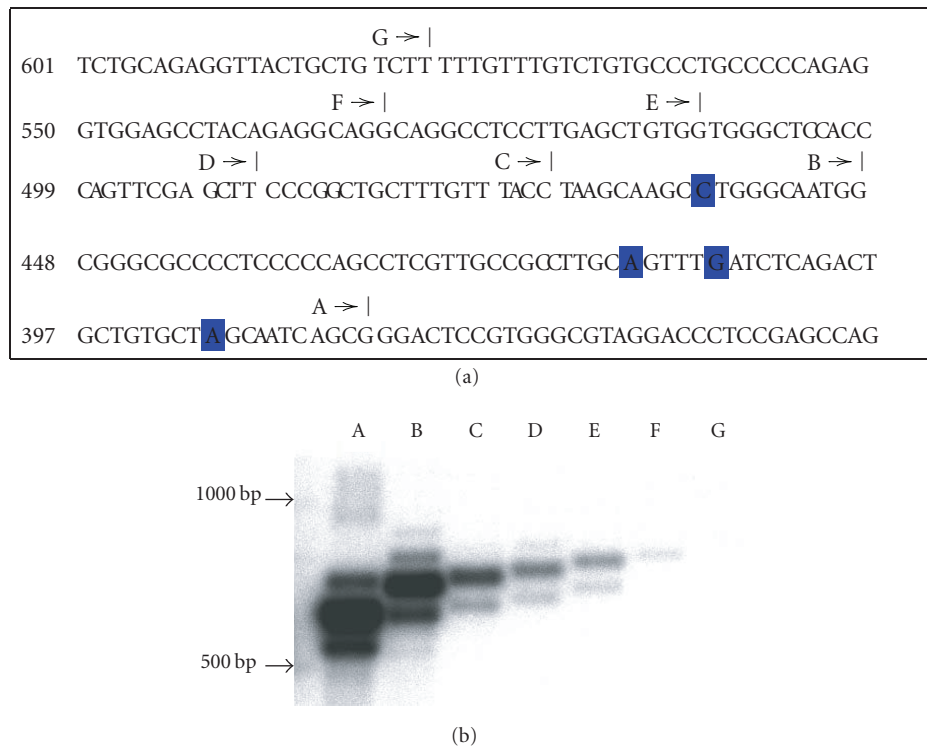


FIGURE 4: TSS mapped to the L1 ASP region. (a) The position of TSS present in the DBTSS is shown highlighted on the consensus sequence of L1Hs [4] between positions 347 and 601. TSS with single and multiple entries present in the database are represented by yellow and blue highlight, respectively. The letters above the sequence mark the 3' end of the oligonucleotide primers used in RT-PCR (see Table 1). (b) Southern blot RT-PCR analysis of the L1-MET transcripts. The lanes are marked according to the primers used in the PCR. Multiple bands on each lane represented the different splice variants of the L1-MET transcript, as confirmed by sequence analysis.

map to the opposite strand of L1 5' UTR. It has been estimated that more than 80% of the TSS in the DBTSS represent true sites of transcription initiation, that is, they correspond to the full-length cDNAs [36]. Twenty four of the 34 TSS, which mapped to the opposite strand of the L1 5' UTR, resided between positions +386 and +503 (Figure 4(a)). The observed nonuniform distribution of the TSS (~70% of TSS

within ~13% of the 5' UTR) clearly shows that the region from +386 to +503, overlapping with the L1 ASP region, must contain a promoter. These results also suggest that transcription initiates at various positions within the L1 ASP region (Figure 4(a)).

To confirm the transcription initiation in the L1 ASP region, we analyzed the distribution of L1-MET chimeric

transcripts (Figure 1(d)) by using RT-PCR and various oligonucleotide primers. Figure 4(b) shows that amplification of L1-*MET* cDNA can be carried out using primers A–F, but not by using primer G. This result indicates that the TSS is located in the L1 ASP region between the binding sites of primers A and F, while the region corresponding to primer G is absent from the L1-*MET* transcripts. Also, an in silico search for potential splicing signals [23, 24] did not reveal any acceptor sites in the region between primers G and E, lending support to the conclusion that transcription is initiated in the L1 ASP region rather than read through the L1 5' UTR. The difference in band intensities (Figure 4(b)) observed for different primer pairs is consistent with the predicted start site heterogeneity. In summary, our results show that the L1 ASP can act as a promoter in vivo and its activity is characterized by start site heterogeneity.

DISCUSSION

In this paper we show that L1 ASP can cause widespread transcription of human genes and its activity correlates with that of the native promoter in some cases, while in other cases it can expand the tissue-specific expression pattern of the respective gene. It is believed that two or more genes located in a single expression domain are coexpressed [37]. Accordingly, an L1 ASP located near or within a gene may behave like a “parasite” whose activity is dependent on the transcription of the gene. This is exemplified by the simultaneous transcription from the L1 ASP and native promoter (Figures 1(a), 1(d), and 2(a)). Surprisingly, in other cases the L1 ASP activity may be regulated independently, as observed here for L1-*CLCN5*, L1-*SLCO1A2*, and L1-*BOLL* mRNAs (Figures 1(b), 1(c), and 2(b)). Although the L1 ASP-driven transcripts were detected in all tissues examined (Figure 3), the results described suggest that the L1 ASPs at defined loci are not active in all tissues. The different tissue-specific activity of L1 ASPs can hardly be explained by their minimal sequence divergence, but could be explained with differences in their epigenetic state. In some cases, transcriptionally active epigenetic state could be stochastically confined to some L1s in certain tissues.

Our results show that L1 ASP acts as an alternative promoter of several human genes (Figures 1(a)–1(c)). Alternative promoters, giving rise to alternative first exons, generate variation in gene expression by increasing transcriptional flexibility and translational diversity. For example, the human *NOS1* gene, encoding neuronal isoform of nitric oxide synthase, has 9 alternative promoters, which determine its tissue-specific transcription and translational efficiency of the resulting *NOS1* mRNAs with different 5' UTRs [38]. Another striking example is the human *BDNF* gene, encoding brain-derived neurotrophic factor, which has 6 promoters and first noncoding exons differentially used in different parts of the brain (A Kazantseva and T Timmusk, personal communication). The L1 ASP, acting as an alternative promoter, generates a chimeric mRNA whose translation could produce a protein identical to the genuine protein. However, the translatability of this transcript depends on the length of the 5' UTR, the number of upstream ORFs, and the strength

of initiation signals [39]. Comparison between the 5' UTRs of the native and chimeric mRNA revealed no major differences in the above-mentioned factors that can abrogate the usage of the genuine ORF (data not shown). Therefore, it is likely that the chimeric L1 transcripts may be translated with efficiency comparable to that of the native transcripts.

Alternative promoters can also generate mRNAs with different 5' coding exons, which may be used in the generation of N-terminal variants of the same protein [40]. Similarly, most L1 ASPs located in introns may, in principle, produce chimeric mRNAs and their translation could yield N-terminally truncated proteins. However, transcription from an L1 ASP located in an intron (39 examples described in Table 2) may be strongly inhibited because of the readthrough transcription from the upstream native promoter [41, 42]. In addition, if transcripts from the intronic L1 ASPs are produced, they may not be readily translated because of the absence of proper initiation context. Although N-terminally truncated proteins with possible dominant negative effects have been shown to exist in normal and cancer cells [40] (references therein), additional experiments are required to prove the translation of chimeric L1 transcripts.

We have detected two L1 ASP-derived antisense RNAs complementary to the exons of *COL11A1* and *BOLL* mRNAs (Figure 2). The other two antisense RNAs predicted from the ESTs (Table 3) were not detected in the human tissues analyzed. Antisense RNAs and antisense transcription are known to cause downregulation of gene transcripts via RNAi-mediated mRNA degradation [43] and transcriptional collision [42], respectively. The possible regulatory interaction between sense and antisense RNAs or transcription may be revealed from the negative (or inverse) correlation of their expression. The partial positive correlation between *COL11A1* mRNA and its antisense counterpart and the negative correlation between *BOLL* and *L1-BOLL* suggest that there is no general correlation between the L1 ASP-driven antisense transcription and the transcription of the gene.

In summary, we have demonstrated that L1 ASP is active in a wide variety of normal human tissues and it is capable of functioning as an alternative promoter by providing the tissue-specific expression of several human genes.

ACKNOWLEDGMENTS

We thank Jaanika Riie for the help in computer analysis, Richard Tamme for critical reading of the manuscript, and Tõnis Timmusk for sharing unpublished data. This work was supported in part by the Grant no 5171 from Estonian Science Foundation.

REFERENCES

- [1] Wheelan SJ, Church DM, Ostell JM. Spidey: a tool for mRNA-to-genomic alignments. *Genome Research*. 2001;11(11):1952–1957.
- [2] Nigumann P, Redik K, Mätlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics*. 2002;79(5):628–634.

- [3] Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Molecular and Cellular Biology*. 2001;21(6):1973–1985.
- [4] Smit AF, Tóth G, Riggs AD, Jurka J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *Journal of Molecular Biology*. 1995;246(3):401–417.
- [5] Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Research*. 2002;12(6):996–1006.
- [6] Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997;25(17):3389–3402.
- [7] Tatusova TA, Madden TL. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiology Letters*. 1999;174(2):247–250.
- [8] Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Research*. 2001;29(1):137–140.
- [9] Deininger PL, Batzer MA. Mammalian retroelements. *Genome Research*. 2002;12(10):1455–1465.
- [10] Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science*. 2004;303(5664):1626–1632.
- [11] Rakyan VK, Blewitt ME, Druker R, Preis JL, Whitelaw E. Metastable epialleles in mammals. *Trends in Genetics*. 2002;18(7):348–351.
- [12] van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics*. 2003;19(10):530–536.
- [13] Kashkush K, Feldman M, Levy AA. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics*. 2003;33(1):102–106.
- [14] Ferrigno O, Virolle T, Djabari Z, Ortonne JB, White RJ, Aberdam D. Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nature Genetics*. 2001;28(1):77–81.
- [15] Whitelaw E, Martin DI. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nature Genetics*. 2001;27(4):361–365.
- [16] Duhl DM, Vrieling H, Miller KA, Wolff GL, Barsh GS. Neomorphic *agouti* mutations in obese yellow mice. *Nature Genetics*. 1994;8(1):59–65.
- [17] Schulte AM, Lai S, Kurtz A, Czubyayko F, Riegel AT, Wellstein A. Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(25):14759–14764.
- [18] Medstrand P, Landry J-R, Mager DL. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *The Journal of Biological Chemistry*. 2001;276(3):1896–1903.
- [19] Landry J-R, Rouhi A, Medstrand P, Mager DL. The Opitz syndrome gene *Mid1* is transcribed from a human endogenous retroviral promoter. *Molecular Biology and Evolution*. 2002;19(11):1934–1942.
- [20] Dunn CA, Medstrand P, Mager DL. An endogenous retroviral long terminal repeat is the dominant promoter for human β 1,3-galactosyltransferase 5 in the colon. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(22):12841–12846.
- [21] Dunn CA, Mager DL. Transcription of the human and rodent *SPAM1 / PH-20* genes initiates within an ancient endogenous retrovirus. *BMC Genomics*. 2005;6(1):47.
- [22] Suzuki Y, Yamashita R, Nakai K, Sugano S. DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. *Nucleic Acids Research*. 2002;30(1):328–331.
- [23] Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*. 1997;4(3):311–323.
- [24] Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *Journal of Molecular Biology*. 1991;220(1):49–65.
- [25] Fisher SE, van Bakel I, Lloyd SE, Pearce SH, Thakker RV, Craig IW. Cloning and characterization of CLCN5, the human kidney chloride channel gene implicated in Dent disease (an X-linked hereditary nephrolithiasis). *Genomics*. 1995;29(3):598–606.
- [26] Park M, Dean M, Kaul K, Braun MJ, Gonda MA, Vande Woude G. Sequence of MET protooncogene cDNA has features characteristic of the tyrosine kinase family of growth-factor receptors. *Proceedings of the National Academy of Sciences of the United States of America*. 1987;84(18):6379–6383.
- [27] Bernard M, Yoshioka H, Rodriguez E, et al. Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA demonstrates that type XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilagenous tissue. *The Journal of Biological Chemistry*. 1988;263(32):17159–17166.
- [28] Strausberg RL, Feingold EA, Grouse LH, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(26):16899–16903.
- [29] Chatterjee TK, Liu Z, Fisher RA. Human RGS6 gene structure, complex alternative splicing, and role of N terminus and G protein γ -subunit-like (GGL) domain in subcellular localization of RGS6 splice variants. *The Journal of Biological Chemistry*. 2003;278(32):30261–30271.
- [30] Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
- [31] Piehler A, Kaminski WE, Wenzel JJ, Langmann T, Schmitz G. Molecular structure of a novel cholesterol-responsive A subclass ABC transporter, ABCA9. *Biochemical and Biophysical Research Communications*. 2002;295(2):408–416.
- [32] Malnic B, Godfrey PA, Buck LB. The human olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(8):2584–2589.
- [33] Myers JS, Vincent BJ, Udall H, et al. A comprehensive analysis of recently integrated human Ta L1 elements. *The American Journal of Human Genetics*. 2002;71(2):312–326.
- [34] Furano AV, Duvernell DD, Boissinot S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends in Genetics*. 2004;20(1):9–14.
- [35] Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. Natural genetic variation caused by transposable elements in humans. *Genetics*. 2004;168(2):933–951.
- [36] Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*. 1994;138(1-2):171–174.
- [37] Spector DL. The dynamics of chromosome organization and gene regulation. *Annual Review of Biochemistry*. 2003;72:573–608.
- [38] Wang Y, Newton DC, Robb GB, et al. RNA diversity has profound effects on the translation of neuronal nitric oxide

- synthase. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(21):12150–12155.
- [39] Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene*. 2002;299(1-2):1–34.
- [40] Landry J-R, Mager DL, Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. *Trends in Genetics*. 2003;19(11):640–648.
- [41] Eszterhas SK, Bouhassira EE, Martin DI, Fiering S. Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Molecular and Cellular Biology*. 2002;22(2):469–479.
- [42] Prescott EM, Proudfoot NJ. Transcriptional collision between convergent genes in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(13):8796–8801.
- [43] McManus MT, Sharp PA. Gene silencing in mammals by small interfering RNAs. *Nature Reviews. Genetics*. 2002;3(10):737–747.
- [44] Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Research*. 2005;15(8):1073–1078.