Analytical Biochemistry 600 (2020) 113769

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Differential expressions of L1-chimeric transcripts in normal and matched-cancer tissues

Songmi Kim[a], Wonseok Shin[b], Yong-Moon Lee[c], Seyoung Mun[a,d,*], Kyudong Han[a,b,d,**]

[a] Department of Nanobiomedical Science & BK21 PLUS NBM Global Research Center for Regenerative Medicine, Dankook University, Cheonan, 31116, Republic of Korea
[b] NGS Clinical Laboratory, Dankook University Hospital, Cheonan, 31116, Republic of Korea
[c] Department of Pathology, Dankook University College of Medicine, Cheonan, 31116, Republic of Korea
[d] Center for Bio-Medical Engineering Core Facility, Dankook University, Cheonan, 31116, Republic of Korea

ARTICLE INFO

Keywords:
Alternative splicing
Antisense promoters (ASPs)
Gene expression
Long INterspersed elements (LINEs)
L1-chimeric transcripts

ABSTRACT

L1s are a *cis*-regulatory elements and contain bidirectional internal promoters within the 5' untranslated region (UTR). L1s provide bidirectional promoters that generate alternative transcripts and affect differential expressions in the human genome. In particular, L1 antisense promoters (L1ASPs) could produce aberrant transcripts in cancer tissues compared to normal tissues. In this study, we identified the L1-chimeric transcripts derived from L1ASPs and analyzed relative expression of L1-chimeric transcripts between normal and matched-cancer tissues. First, we collected 425 L1-chimeric transcripts by referring to previous studies. Through the manual inspection, we identified 144 L1-chimeric transcripts derived from 44 L1 antisense promoters, suggesting that the antisense promoter acted as an alternative promoter. We analyzed relative gene expression levels of 16 L1-chimeric transcripts between matched cancer-normal tissue pair (lung, liver, gastric, kidney, thyroid, breast, ovary, uterus, and prostate) using real-time quantitative PCR (RT-qPCR) and investigated putative transcription factor binding motifs to determine activity of L1ASPs. Taken together, we propose that L1ASPs could contribute to the differential gene expression between normal and cancer tissues.

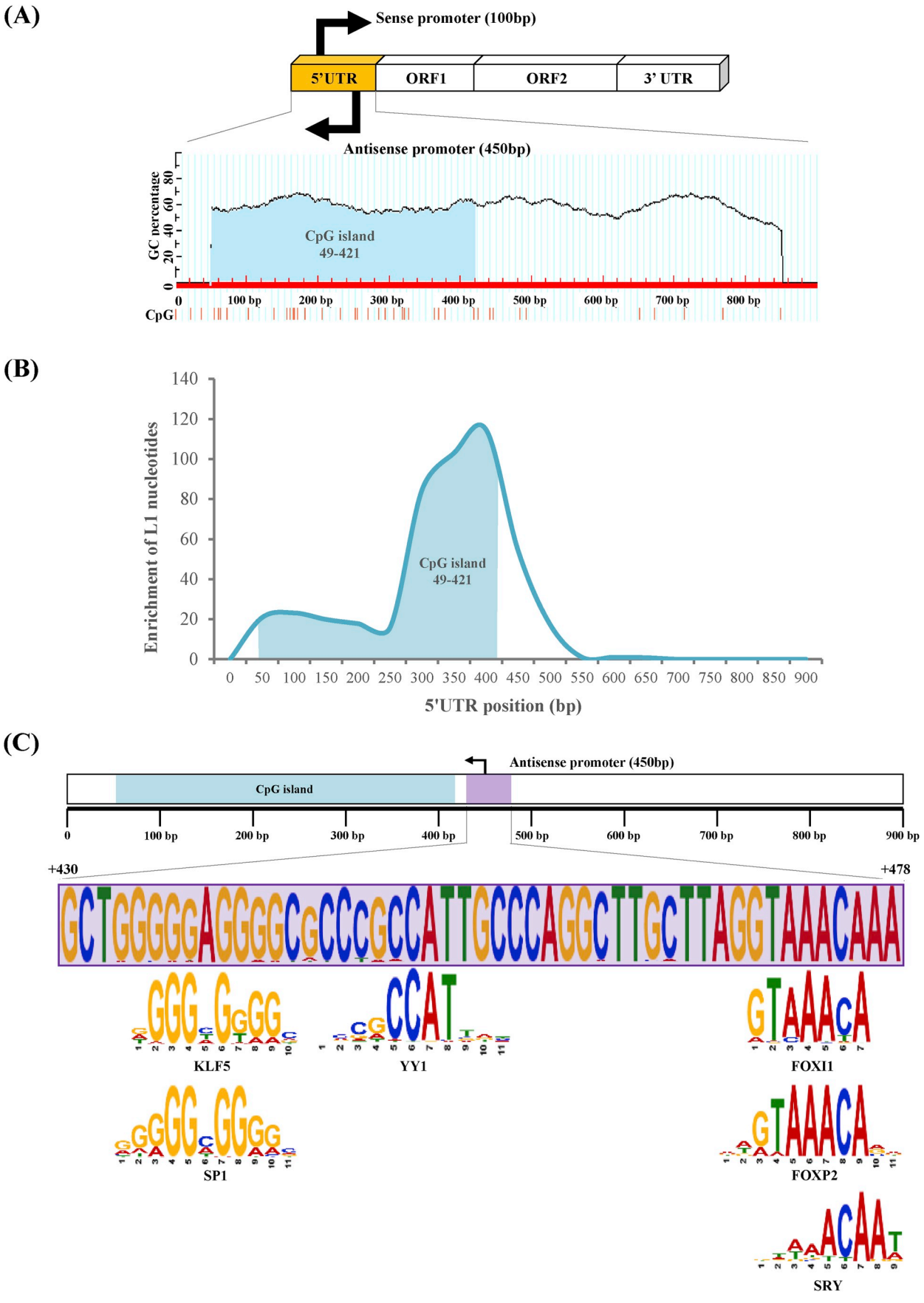
1. Introduction

Transposable elements (TEs) account for ~45% of the human genome [1]. TEs are continuously contributing to genomic structural variations which has played a major role in primate evolution. TEs are classified into two categories; DNA transposons and retrotransposons such as LINE, SINE, ERV, and SVA [2,3]. Among these elements, Long INterspersed Elements (LINE-1s, L1s) make up ~17% of the human genome, with a total of about 520,000 copies [1]. Approximately, 1800 copies of human-specific L1s are integrated into genome after the divergence of human and chimpanzee lineages [4]. L1s are autonomous elements, which can mobilize into another host genome by themselves. They contain two open reading frames (ORF1 and ORF2) flanked by 5′ and 3′ untranslated regions (UTRs). The 5' UTR of L1s has bidirectional internal promoters, including sense and antisense promoters. The ORF1 encodes an RNA binding protein, which binds to intermediate RNA. The ORF2 encodes endonuclease and reverse-transcriptase, which nicks the DNA and prepares target-primed reverse transcription (TPRT) [5–8].

Most of the L1s have lost their retrotransposition capability, due to acquisition of mutations or 5′ truncation resulting from the incomplete TPRT mechanism. Nevertheless, some L1s could contribute to regulation of gene expression at the transcriptional level as a *cis*-regulatory elements. The 5′ UTR of L1s contains bidirectional promoters, which can bind transcription factors and regulate nearby genes [9,10]. The 5′ UTR of L1s ranges from +1 to 909bp, including bidirectional promoters in the +100bp and +450bp regions (Fig. 1A) [10–14]. L1s contain a CpG island with a GC content of more than 50% at positions +49 to +421 in the 5′UTR (Fig. 1A). However, L1s do not contain a TATA-box called initiators in this region and they initiate transcription by other mechanisms [11]. Previous studies have reported that transcription of L1s is initiated at nucleotide +1 region and YY1, which is one of transcription factors, binds to nucleotides +13 to +21 [13,14]. Another transcription factor, SRY, binds to two regions, which is nucleotides 472–477 and 572–577 regions [15]. In addition, RUNX3

* Corresponding author. Department of Nanobiomedical Science & BK21 PLUS NBM Global Research Center for Regenerative Medicine, Dankook University, Cheonan, 31116, Republic of Korea.
** Corresponding author. Department of Nanobiomedical Science & BK21 PLUS NBM Global, Center for Regenerative Medicine, Dankook University, Cheonan, 31116, Republic of Korea.
E-mail addresses: 12200281@dankook.ac.kr (S. Mun), jim97@dankook.ac.kr (K. Han).

https://doi.org/10.1016/j.ab.2020.113769
Received 19 December 2019; Received in revised form 25 April 2020; Accepted 5 May 2020
Available online 08 May 2020
0003-2697/



(caption on next page)

Fig. 1. Schematic representation of position of L1ASPs in L1s. (A) LINE-1 consists of 5′UTR, two ORFs, and 3′UTR. They contain bidirectional promoters and CpG islands in the 5′UTR. The graph showed that GC percent (X-axis) and nucleotide position (Y-axis) within 5′UTR of L1s. Orange lines denote the CpG sites and blue boxes denote CpG islands. Black arrows indicate bidirectional promoters. (B) Transcription of L1-chimeric transcripts were initiated at various transcription start sites within 5′UTR of the L1s. The X- and Y-axis represent the position of the 5′UTR and the enrichment of the L1 sequences, respectively. (C) Transcription factor binding motifs were highly enriched at +430 to +478 within antisense promoter region of the L1s (Purple). Black arrows mean direction of the promoters. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Identification of L1-chimeric transcripts.

	L1-chimeric transcript	L1ASP
Computational analysis	425	122
Manual inspection	144	44
Primer design	61	38
PCR validation	37	22
qRT-PCR analysis	16	11

Table 2
Classification of 44 L1ASPs subfamilies.

L1 subfamily		L1 Ta subfamily	
L1PA2	28	non-Ta	28
L1HS	16	Pre-Ta (ACG/G)	9
		Ta-1	6
		Pre-Ta (ATG/G)	1

transcription factor (TF) binds to nucleotides 83–101 that increase transcription of L1s [16]. Although L1s have high mutation rate due to host defense mechanisms, the TF binding sites of the 5′UTR are conserved in the human genome [12].

Generally, the L1 sense promoter (+100bp) transcribes L1 itself, while the L1 antisense promoter (+450bp) could produce aberrant transcripts in the opposite direction. Recent studies have published that a new primate-specific L1 ORF (ORF0) located in nucleotides +452 to +236 of L1s includes two splice donor sites that could produce L1-chimeric transcripts, and thus generate fusion proteins with proximal exons [17,18]. Since L1-chimeric transcripts derived from L1ASPs were first described by Speek in 2001, several researches have reported that L1ASPs could affect diversity of transcripts in various cells and drive various transcripts in tissue-specific manner [10,19,20].

In addition, many researchers have focused on the association between L1-chimeric transcripts and cancer. Cruickshanks and Tufarelli have been reported eighteen cancer-specific L1-chimeric transcripts in breast and colon cancer cell lines using L1 chimera display [9]. L1-chimeric transcripts generated by hypomethylated L1ASPs within proto-oncogenes were identified in colorectal cancers [21]. In addition, L1ASP, located in intron 2 of *c-Met* gene, which is a receptor tyrosine kinase generates alternative transcripts in bladder cancer and breast cancers [18,22].

Cancer is the most common cause of mortality in South Korea [23]. In 2019, a total 221,347 Korean patients were newly diagnosed with cancers and 82,344 Koreans died of various cancers. The five major cancers of Korean men are lung cancer, gastric cancer, colorectal cancer, prostate cancer, and liver cancer. The five major cancers of Korean women are breast cancer, colorectal cancer, gastric cancer, thyroid cancer, and lung cancer [23]. Early diagnosis of cancer is very important for effective treatment and reducing mortality.

In this study, we collected 425 L1-chimeric transcripts generated by L1ASPs from previous studies [24,25]. Manual inspection allowed us to identify 144 L1-chimeric transcripts derived from 44 L1ASP. We also used RT-qPCR to validate their relative gene expressions of 16 L1-chimeric transcripts in both normal and matched-cancer tissues from the same individuals. Taken together, we suggest that L1ASPs could play an important role in the differences in transcript expression between normal and matched-cancer tissues.

2. Materials and methods

2.1. Computational analysis

We obtained 425 L1-chimeric transcripts derived from 122 L1 antisense promoters (ASPs), which are relatively young L1s, such as L1Hs and L1PA2 subfamilies [24,25]. From these data, we manually inspected 425 L1-chimeric transcripts and 122 L1s to verify subfamily and orientation of L1s in the human genome (hg38) using the UCSC genome browser (http://genome.ucsc.edu) with RepeatMasker track. We also used multiz alignment and conservation tracks to identify L1-chimeric transcripts derived from human-specific L1s. In addition, we confirmed structure of L1-chimeric transcripts and whether the transcripts were originated from L1s by using human mRNA and human EST tracks. Furthermore, we download DNA sequences of the L1-chimeric transcripts and 44 L1s using Table browser. We removed overlapped transcripts and non-verifiable transcripts by real-time qPCR (RT-qPCR). Finally, we used the RepeatMasker program (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) to identify 144 L1-chimeric transcripts containing L1-derived sequences in the first exon (Supplementary Table 1). TCGA SpliceSeq (http://projects.insilico.us.com/TCGASpliceSeq/index.jsp) and TSVdb (http://www.tsvdb.com/instruction.html), which provide transcript splicing patterns and splicing event details of The Cancer Genome Atlas project (TCGA) RNA-Seq data, were used for integration and visualization analysis of L1-chimeric transcripts [26,27].

2.2. Sample preparation

To verify L1-chimeric transcripts, we extracted RNAs from 9 cancer-normal tissue pairs (liver, lung, gastric, kidney, thyroid, breast, uterus, ovary, and prostate tissues). Among these, liver, lung and kidney tissues were derived from patient samples provided by Chonnam National University's Hwasun Hospital, ovary and breast tissues were from Korea University. Prostate tissues were provided from Pusan national university Hospital. Additionally, gastric, thyroid, and uterus tissues were from Keimyung Human Bio-Resource Bank (KHBB). All samples derived from the National Biobank of Korea were obtained with informed consent under the institutional review board (IRB)-approved protocols (IRB No. DKUH 2016-11-007). Total RNAs were extracted using Trizol reagent (ThermoFIsher Scientific; Waltham MA, USA) and RNeasy mini prep kit (Qiagen; Valencia CA, USA). In addition, we removed genomic DNA contamination by RNase-Free DNase set (Qiagen; Valencia CA, USA). The RNAs were measured concentrations using a microvolume spectrometer Colibri (Titertek-Berthold, Pforzheim, Germany) at the Center for Bio-medical Engineering Core Facility (Dankook University, South Korea).

2.3. cDNA synthesis and RT-PCR

We performed cDNA synthesis on 500 ng of RNAs to validate L1-chimeric transcripts using the ReverTra Ace qPCR RT Master Mix (Toyobo; Osaka, Japan). The RT master mix contains reverse transcriptase, RNase inhibitor, oligo dT primer, random primer, MgCl2, and dNTPs. The cDNA synthesis reactions were carried out for 15min incubation at 37 °C followed by incubation of 50 °C for 5min, and 98 °C for 5min using a thermocycler (Bio-Rad Laboratories; Berkeley CA, USA) at the Center for Biomedical Engineering Core Facility (Dankook

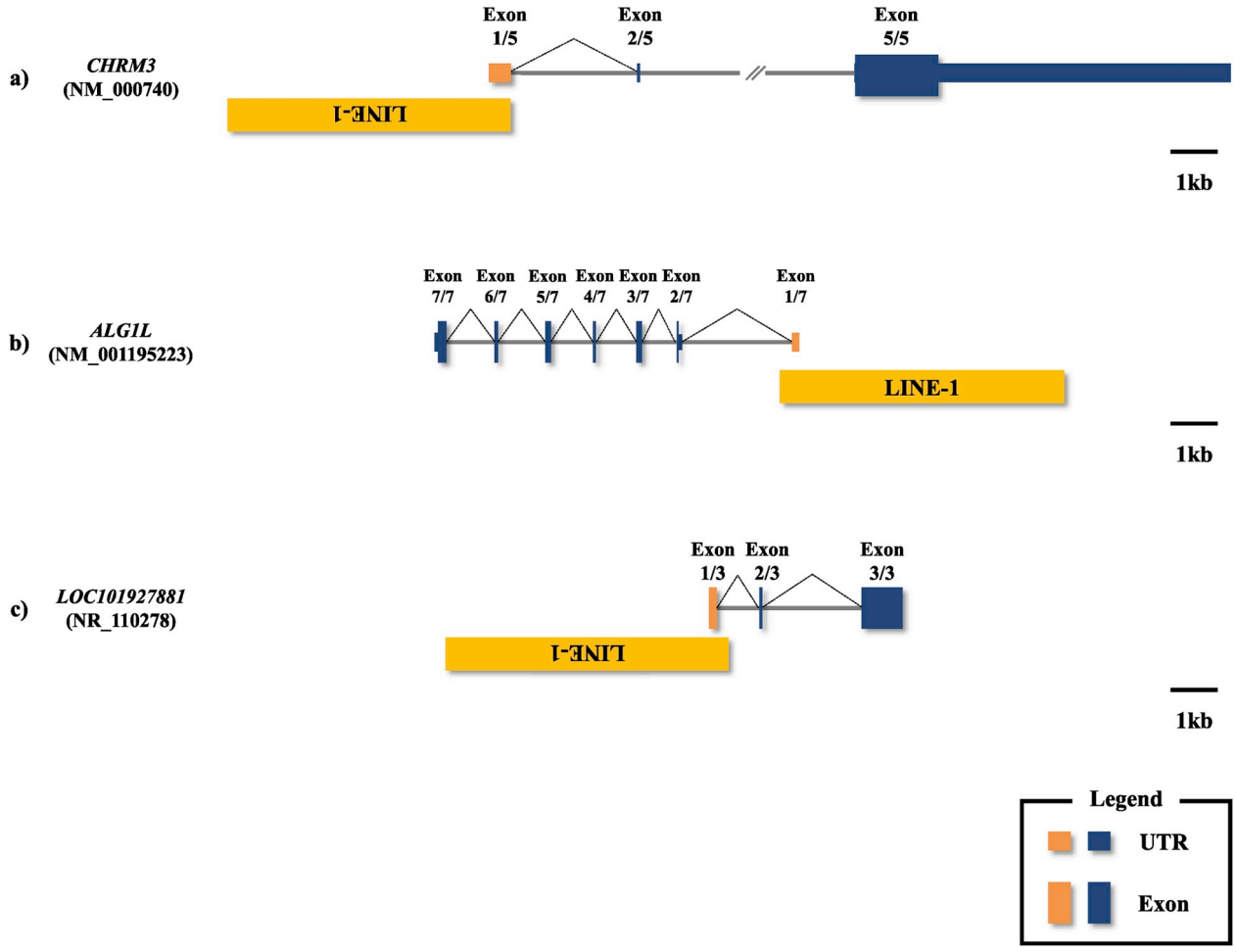


Fig. 2. Structural scheme of L1-chimeric transcripts derived from L1ASP in exonic region. These three L1s located in the exonic region of RefSeq provide the first exon to generate the L1-chimeric transcript (a) L1HS is located in the exonic region of *CHRM3* (NM_000740) with antisense direction, which provides first exon of *CHRM3*. (b) L1PA2 is in the antisense direction of the first exon of *ALG1L* (NM_001195223). (c) *LOC101927881* (NR_110278) is generated by L1PA2, which is located in first exon of *LOC101927881*. Yellow and Blue boxes denote partial L1 and exons, respectively. Gray and black solid lines indicate introns and scale bars, respectively. Splicing events are represented by black dot lines. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

University, South Korea). The synthesize cDNAs were diluted in 90 μl of RNase-free water and stored until use at −20 °C. The cDNAs were used as the template for the RT-PCR and RT-qPCR reaction. As a housekeeping gene, a primer pair of glycerol aldehyde-3-phosphate dehydrogenase (*GAPDH*) was designed in the human *GAPDH* sequences of GenBank (NM_002046) and the synthesized cDNA was verified. We designed oligonucleotide primer pairs for validation of L1-chimeric transcripts as previously described [25]. The list of primer pairs used in this study is shown in Supplementary Table 2. The PCR amplification of the human cDNAs was carried out using 2X AmpMaster™ Reaction Mix (GeneAll; Seoul, Korea). The PCR amplifications was conducted at an initial denaturation of 3 min at 95 °C followed by 30 cycles of 95 °C for 30 s, optimal annealing temperature for 15 s, and 72 °C for 1 min, and 1 min at 72 °C using a thermocycler (Bio-Rad Laboratories; Berkeley CA, USA) at the Center for Biomedical Engineering Core Facility (Dankook University, South Korea). The PCR products were sized on 1% agarose gels and visualized using Geldoc (Bio-Rad Laboratories; Berkeley CA, USA) at the Center for Bio-medical Engineering Core Facility (Dankook University, South Korea). The PCR products were purified using Favorprep Gel/PCR purification kit (Favorgen Biotech Corp; Ping-Tung, Taiwan) and performed cloning using TOP cloner TA Kit (Enzynomics; Daejeon, Korea). DNA analysis was carried out with ABI 3500 Genetic Analyzer (Applied Biosystems; Foster city CA, USA) at the Center for Bio-medical Engineering Core Facility (Dankook University, South Korea).

2.4. Real time-quantitative PCR (RT-qPCR)

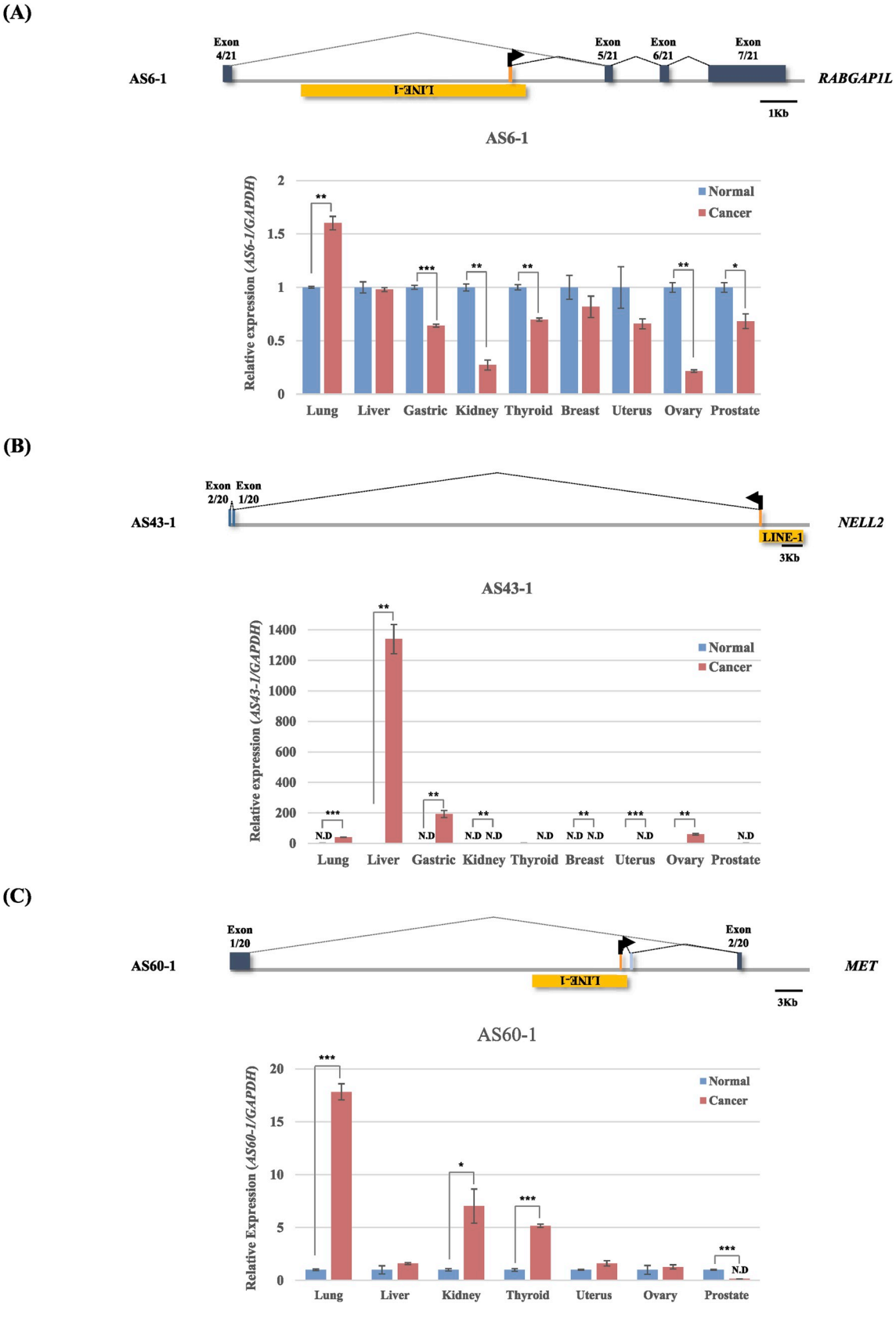
We used real-time quantitative PCR (RT-qPCR) to analyze the relative gene expressions of L1-chimeric transcripts between normal and matched-cancer tissues. RT-qPCR was performed using QuantiSpeed-SYBR green kit (PhileKorea; Seoul, Korea), Eco real-time PCR machine (Illumina; San Diego CA, USA) at the Center for Bio-medical Engineering Core Facility (Dankook University, South Korea), and the primers previously described in Supplementary Table 2. The PCR reactions were initiated with a 95 °C heating step for 10 min followed by denaturation at 95 °C, 10 s; annealing at optimal temperature, 10 s; extension at 72 °C, 15 s. After 40 cycles, the dissociation temperature range extends from 60 °C to 95 °C, with ramping at 0.3 °C per min. Following amplification, the melting curve analysis of the PCR products was performed to confirm specificity. The *GAPDH* as an endogenous control was normalized to expression level of L1-chimeric transcripts. The relative gene expression was calculated by using the $2^{-\Delta\Delta Ct}$ value for each sample [28,29]. For each cDNA sample, RT-qPCR was performed in triplicates.

Table 3
Information of 44 L1 antisense promoters.

L1ASP	Name	L1 subfamily	L1 position	Orientation	Size (bp)	Location	Referenec
AS1	CHRM3	L1HS	Chr1:239623499-239629523	–	6025	Exonic	[25] [24]
AS2	RAB3IP	L1HS	Chr12:69773411–69779441	–	6031	intronic	[25] [24]
AS3	FSTL4	L1HS	Chr5:133583289-133589299	+	6011	intronic	[25] [24]
AS4	FGGY	L1PA2	Chr1:59647748-59653779	–	6032	intronic	[25] [24]
AS5	COL24A1	L1HS	Chr1:85748520-85754548	+	6029	intronic	[25] [24]
AS6	RABGAP1L	L1PA2	Chr1:174233267-174239293	–	6027	intronic	[25] [24]
AS7	ZNF638	L1HS	Chr2:71411475-71417501	–	6027	intronic	[25] [24]
AS8	ACVR1C	L1HS	Chr2:157566354-157572377	+	6024	intronic	[25] [24]
AS9	CD96	L1PA2	Chr3:111556204-111562234	–	6031	intronic	[25] [24]
AS10	IGSF11	L1PA2	Chr3:118914244-118920271	+	6028	intronic	[25] [24]
AS11	USP53	L1PA2	Chr4:119274114-119280127	–	6014	intronic	[25] [24]
AS12	MARCH1	L1PA2	Chr4:164350750-164356776	+	6027	intronic	[25] [24]
AS13	EXOC4	L1PA2	Chr7:133889426-133895455	–	6030	intronic	[25] [24]
AS14	MAL2	L1PA2	Chr8:119159399-119165410	–	6012	promoter	[25] [24]
AS15	BIRC2	L1PA2	Chr11:102369167–102375192	–	6026	intronic	[25] [24]
AS16	SCFD1	L1HS	Chr14:30684810–30690837	–	6028	intronic	[25] [24]
AS17	C14orf37	L1PA2	Chr14:58032540–58038561	+	6022	intronic	[25] [24]
AS18	LRRC9	L1PA2	Chr14:59946505–59952523	–	6019	intronic	[25] [24]
AS19	RGS6	L1PA2	Chr14:71842965–71848996	–	6032	promoter	[25] [24]
AS21	ABCA9 ABCA9-AS1	L1HS	Chr17:69000001–69005148	–	5148	intronic	[25] [24]
AS24	CAMK4	L1HS	Chr5:111302239-111308262	+	6024	intronic	[25,32]
AS25	ALG1L	L1PA2	Chr3:125936639-125942696	+	6058	Exonic	[24]
AS27	BCAS3	L1PA2	Chr17:60881441–60887474	–	6034	intronic	[24]
AS28	C18orf61	L1PA2	Chr18:12202445–12208458	–	6014	intronic	[24]
AS29	CLCN5	L1HS	ChrX:50019457-50025505	–	6049	intronic	[24]
AS30	CMYA5	L1HS	Chr5:79778885-79784938	–	6054	intronic	[24]
AS31	CNTLN	L1PA2	Chr9:17285004-17291034	–	6031	intronic	[24]
AS33	FAM86JP	L1PA2	Chr3:125936639-125942696	+	6058	intergenic	[24]
AS34	FEZ1	L1HS	Chr11:125536610–125542640	+	6031	promoter	[24]
AS35	FEZF1-AS1	L1PA2	Chr7:122278916-122284945	–	6030	promoter	[24]
AS36	FMo4	L1PA2	Chr1:171373587-171379567	+	5981	intergenic	[24]
AS37	GBP1P1	L1PA2	Chr1:89412580-89418615	–	6036	intronic	[24]
AS38	HYDIN2	L1HS	Chr1:146717193-146723218	–	6026	intronic	[24]
AS39	IL20RB	L1HS	Chr3:136963694-136969736	–	6043	intronic	[24]
AS40	LINC01572	L1PA2	Chr16:72627275–72633298	+	6024	intronic	[24]
AS41	LOC100996634	L1PA2	Chr6:109245219-109251251	–	6033	intronic	[24]
AS42	LOC101927881	L1PA2	Chr2:128858985-128865016	–	6032	Exonic	[24]
AS43	NELL2	L1PA2	Chr12:44950959–44956988	+	6030	promoter	[24]
AS44	PLCB1	L1PA2	Chr20:8595102–8601127	–	6026	intronic	[24]
AS45	SPAG17	L1PA2	Chr1:118280888-118286905	+	6018	promoter	[24]
AS46	SPATA17 SPATA17-AS1	L1PA2	Chr1:217690493-217696523	–	6031	intronic	[24]
AS47	SVEP1	L1HS	Chr9:110417103-110423126	+	6024	intronic	[24]
AS48	TBC1D32	L1HS	Chr6:121162717-121168725	+	6009	intronic	[24]
AS60	MET	L1PA2	Chr7:116718498-116724489	–	5992	intronic	[24]

2.5. Prediction of L1 activity

To predict promoter activity of L1s, we conducted analysis of TF binding sites on 5′UTR of 44 L1s using MEME-chip web-based tools (http://meme-suite.org/tools/meme-chip). The significance of TF motifs were evaluated by using E-values ≤0.05 [30]. Furthermore, we used L1Xplorer (http://l1base.charite.de/) to confirm whether the ORFs of 44 L1s are structurally intact for L1 activity [31].



(caption on next page)

Fig. 3. Expression analysis of L1-chimeric transcripts in normal and matched-cancer tissues. There are three representative L1s with antisense direction in their introns or promoter regions that generated L1-chimeric transcripts. Structural scheme shows that 3 L1-chimeric transcripts derived from L1ASPs. Yellow and blue boxes indicate L1 and exons, respectively. Gray and black solid lines denote introns and scale bars, respectively. Splicing events are represented by dot lines (black and gray). Directions of transcription are represented by black arrows. The graphs represent the level of expression of L1-chimeric transcripts in 9 types of normal and matched-cancer tissue samples, determined by relative RT-qPCR. Relative expression levels of L1-chimeric transcripts were calculated by using the $2^{-\Delta\Delta Ct}$ method ($n = 1$ per tissue sample). *GAPDH* was used to normalize gene expressions as a reference gene. Error bars represent SDs. Asterisks indicate statistically significant differences by Student's t-test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). N.D, not detected. (A) AS6-1 transcript, (B) AS15-1 transcript, (C) AS60-1 transcript. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3. Results and discussion

3.1. Identification of L1-chimeric transcripts derived from L1ASPs

To investigate differential expressions in normal and matched-cancer tissue samples, we collected 425 L1-chimeric transcripts derived from 122 L1ASPs from previous studies [24,25]. A total of 144 L1-chimeric transcripts derived from human-specific L1s were manually inspected their subfamilies and structures of isoforms using UCSC genome browser (Table 1). The 144 L1-chimeric transcripts containing 140 EST transcripts, 2 mRNA sequences, and 2 RefSeqs were derived from 44 L1ASPs (Table 1; Supplementary Table 1). The 44 human-specific L1s are classified into two young subfamilies including 16 L1HS elements and 28 L1PA2 elements based on diagnostic mutations by using L1Xplorer (http://l1base.charite.de/l1xplorer.php). L1HS was divided into four subfamilies including Pre-Ta (ACG/G), Pre-Ta (ATG/G), and Ta-1, which had 9, 1, and 6 copies, respectively (Table 2). The majority of L1s provided 44 L1ASPs were full-length (~6 kb) except for one L1. Thirty-three L1s were located in the intronic region and three L1s were located in the exonic regions containing exon 1 of the genes (Fig. 2). In addition, six L1s were located in the promoter region of other adjacent genes, which seemed to influence the expression of the adjacent genes, and two L1s were located in the intergenic regions (Table 3). The L1s showed relatively high frequencies (>3) on chromosome 14 and 1, but not in five chromosome including chromosome 10, 13, 19, 21, and 22 (Supplementary Fig. 1). The chromosome 14 and 1 included 4 L1s and 9 L1s, respectively (Supplementary Table 1).

To investigate the transcription start sites, we analyzed the L1 sequences of 144 L1-chimeric transcripts (Fig. 1B). Most of L1-chimeric transcripts were initiated between +386 and + 503 in 5′UTR of L1 sequences. The average length of the L1 sequences within L1-chimeric transcripts was 152 nucleotides, which mainly included position between +300 and +425 (Fig. 1B). These results supported that the antisense promoter of L1 is located in this region (position at +450) and the transcription of L1 is initiated at various positions in two major regions (+378–431 and + 480–497) within the 5′UTR [11,12,33].

Furthermore, to examine putative TF binding sites in the 5′UTR of 44 L1 sequences, we used a MEME-chip (http://meme-suite.org/tools/meme-chip) web-based tools [30] to predict promoter activity. Several motifs, including families of Fox (Forkhead), KLF/SP, SRY and YY1, are involved in cell proliferation, differentiation, apoptosis, and development and are abundant in the 5′UTRs of 44 L1 sequences (position at +430 to +478) (Fig. 1C and Supplementary Fig. 2). These results suggest that transcription factors are not only able to recognize and bind TF binding sites in L1ASPs, but also to regulate gene expression of L1 and its neighboring genes [34].

3.2. Differential expression of L1-chimeric transcripts

Active L1s cause insertional mutations and could drive oncogenic process in a variety of ways. They could disrupt tumor suppressor genes or activate the oncogenes by altering gene expression [32,35]. From manual inspection, we designed 61 primer pairs to detect L1-chimeric transcripts, where the forward primer is designated within L1 and the junction of the exon and reverse primer is designated to exon. We excluded that the length of L1 was too short to design primers or variations of the transcripts could not be detected by real-time quantitative PCR (RT-qPCR). Thirty-seven L1-chimeric transcripts were validated in several normal and/or cancer tissues using reverse transcriptase (RT)-PCR, but some L1-chimeric transcripts were excluded from the analysis because of showing two or more alternative isoforms, or different band sizes. Seventeen L1-chimeric transcripts were validated by cloning and sanger sequencing (Supplementary Fig. 3). Among these, we analyzed 16 L1-chimeric transcripts derived from 11 L1ASPs in ten normal and matched cancer tissues (Table 1).

#AS6-1 transcript derived from L1PA2 is located in intron 4 of *RABGAP1L* (Fig. 3A). The *RABGAP1L* encodes RabGTPase-activating protein 1-like protein and catalyze an exchange of active GTP for GDP. The *RABGAP1* interacts with *TUFT1* (tuftelin 1) and regulates mTOR signaling pathway, resulting in intracellular compartment positioning and vesicular trafficking [36]. Our data showed that #AS6-1 transcript was down-regulated in most cancer tissues compared to normal tissues. Among them, in ovary and kidney cancers, the transcript showed a change in expression that was 5-fold lower in cancer samples that in normal samples, whereas in lung cancer, the cancer sample showed 1.5-fold higher expression than the normal sample (Fig. 3A). Kim et al. have been also reported that alternative transcript of *RABGAP1L* was broadly expressed in normal tissues and it was highly expressed in lung cancer sample [25].

#AS43-1 transcript generated by L1PA2 is located in approximately 37 kb upstream of *NELL2* (Fig. 3B). The *NELL2* encodes epidermal growth factor-like protein 2, involved in cell growth and differentiation. Our data showed that #AS43-1 transcript was rarely or not expressed in normal tissues, whereas it was highly expressed in several cancer tissues including lung, liver, gastric, and ovary cancers (Fig. 3B). In particular, the expression of liver was about 1300 times higher than normal liver. Previous studies have been reported that *NELL2* mRNA is mainly expressed in the brain and colorectal cancer cell lines (SW480) [37,38]. The *NELL2* activates mitogen-activated protein kinases (MAPKs) and phosphorylation of c-jun N-terminal kinase (JNK) in the brain [37,39]. This result supported that NELL2 protein contribute to cell growth, differentiations, and oncogenesis.

#AS60-1 transcript is a well-known L1-chimeric transcript generated by L1PA2 that is located in intron 1 of the *c-Met* gene, a proto-oncogene (Fig. 3C). Previous studies have shown that alternative transcripts of *c-Met* are highly expressed in various cancers, including bladder cancer, colorectal cancer, and liver cancer [21,22,40]. Our data also showed that the transcript was highly expressed in most cancers compared to matched-normal tissues (Fig. 3C). However, we could not detect Cq values in gastric and breast cancers and matched-normal tissues as well as prostate cancer.

Interestingly, L1ASP#27, located in intron 8 of the *BCAS3* gene, generates several L1-chimeric transcripts. Of these, we were able to detect six L1 chimeric transcripts. *BCAS3* gene is regulated by estrogen receptor alpha that is amplified and overexpressed in breast cancer cells [41]. The #AS27 transcripts were initiated at variable positions within 5′UTR (Fig. 4A). Our data showed that six L1-chimeric transcripts were differentially expressed in ten normal and matched-cancer tissues (Fig. 4B). Generally, #AS27-3 and #AS27-18 transcripts are highly expressed in several cancers including lung, liver, breast, uterus, and prostate cancers. However, all of transcripts were down-regulated in gastric, kidney, thyroid, and ovary cancers (Fig. 4B). In order to

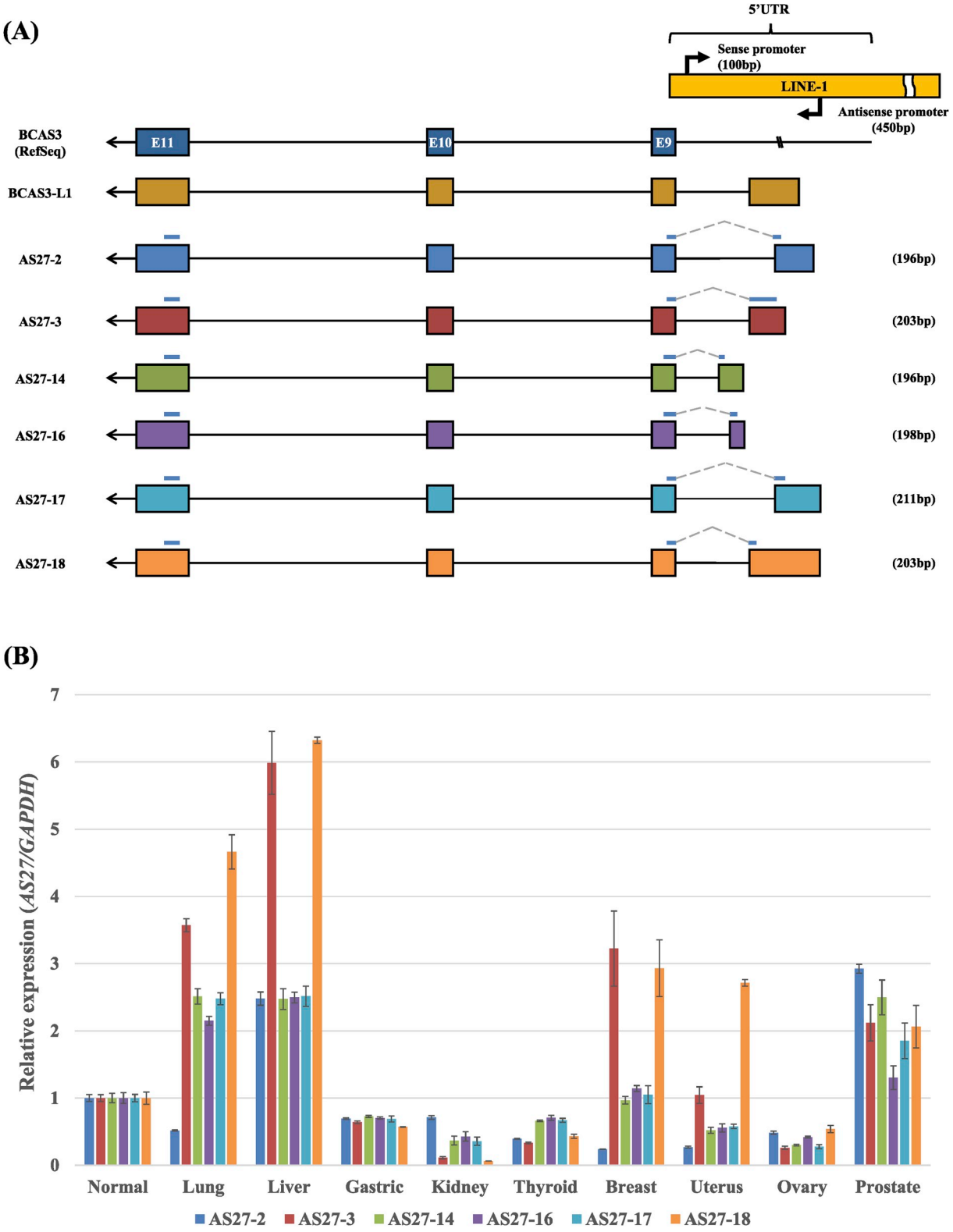


Fig. 4. Comparison of relative expression levels for six L1-chimeric transcripts derived from #AS27 between various cancers. (A) The structural scheme shows the each L1-chimeric transcript derived from #AS27. Boxes and solid lines are represented exons and introns, respectively. Primers and splicing events are represented by blue solid lines and gray dotted lines, respectively. (B) The relative expression in cancer tissues was analyzed by comparing L1-chimeric transcripts with matched-normal tissues with the RT-qPCR in triplicates ($n = 1$ per tissue sample). The relative expressions were calculated by using the $2^{-\Delta\Delta Ct}$ method. The expression data were normalized to the expression of *GAPDH* gene, as a house keeping gene. SD was shown as an error bar. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

investigate the contribution of alternative transcript derived from L1ASP and its expression, we explored a large cohort study of cancer, The Cancer Genome Atlas (TCGA) dataset. Most ASP transcripts which we examined could not be verified because they were located in promoter regions or were not considered alternative transcripts within TCGA data. Interestingly, the six L1-chimeric transcripts in the *BCAS3* gene, occurred by L1ASP#27 were found in all eight types of cancer except ovarian serous cystadenocarcinoma (OV), of which significant differences were observed in breast invasive carcinoma (BRCA), liver hepatocellular carcinoma (LIHC), and lung squamous cell carcinoma (LUSC) compared to normal (Supplementary Fig. 5). These results supported that L1ASPs could contribute to aberrant expression of genes and be significantly associated with cancers and tissue-specific expressions.

Furthermore, in order to confirm the amplification of false positive L1-chimeric transcripts, we additionally designed forward primers anchored to the L1 sequences of 9 L1-chimeric transcripts out of 16 L1-chimeric transcripts subjected to RT-qPCR (Supplementary Table 2). Of these, six L1-chimeric transcripts were validated by RT-PCR and gel electrophoresis. The results showed that the PCR product detected multiple band patterns in several normal and matched-cancer tissues. For Sanger sequencing of PCR products containing multiple bands, we performed cloning and sequencing analyses for each PCR product. Since the L1 sequences of forward primers have high similarity to L1 at other positions, the non-specific products were generated by PCR amplification. Consequently, we identified eight L1-chimeric transcripts, including five alternative L1-chimeric transcripts. Among newly identified five alternative L1-chimeric transcripts, four alternative L1-chimeric transcripts contain a longer L1 sequences in first exon and one alternative L1-chimeric transcript (AS34-1_2_AT.2) contains a partial *Alu*Sz sequence as a second exon (Supplementary Fig. 6). These results supported that TEs could generate transcript isoforms, providing alternative splicing sites in the human genome [42,43].

3.3. Prediction of L1ASP activity

To predict their retrotransposition competency using L1Xplorer database [31], we analyzed 44 L1 sequences that could generate L1-chimeric transcripts. Although the majority of L1s are full-length (>6 kb), they have non-intact ORFs. Only four L1s contain intact ORFs without gaps, frameshift mutations, and nonsense mutations (Supplementary Table 3). They could produce intact proteins, including RNA binding protein, endonucleases, and reverse transcriptase that are required to mobilize them into new genomic locations.

In addition, 44 L1s contain several TF binding sites that might regulate L1 transcription. Among them, 6 TF binding motifs, including two SRY sites, one YY1 BoxA + BoxA site, two Runx3 sites, and one Runx3ASP site, remained intact conserved forms in the 30 L1s (Supplementary Table 3 and Supplementary Fig. 4). Especially, in the 38 L1s, Runx3 antisense promoter binding motif is conserved at position +526 to +508. The Runx3 has been reported to play an important role in the regulation of L1 transcription and retrotranposition in the human genome [16].

L1ASP activity is also regulated by the DNA methylation. Generally, L1s are heavily methylated in the normal somatic cells. When L1 methylation status is hypomethylated, the transcription is increased in the sense and antisense direction of L1s. But methylation level of L1s are different in tissues and genomic loci [44,45]. Therefore, we analyzed TF binding sites to predict L1ASP activity and our results indicate that expression of transcripts derived from L1ASPs is controlled by several transcription factors.

4. Conclusions

We obtained 425 L1-chimeric transcripts, derived from 122 L1s, by referring to previous studies. We identified 144 L1-chimeric transcripts generated by 44 L1s. Forty-four L1s contain CpG island and TF binding sites in the 5′UTR. The transcripts derived from 44 L1s were initiated at variable position of the 5′UTR. We detected and analyzed 16 L1-chimeric transcripts by using RT-qPCR. The results showed that several L1-chimeric transcripts were differentially expressed between normal and matched-cancer tissues. In addition, some of the L1s showed a pattern of producing several L1-chimeric transcripts with different tissue-specific expression. Although we could not be generalized the data because we only used one sample for each type of cancer, our data supported that L1ASPs could regulate expression of their adjacent genes by providing splicing sites. L1ASPs also could contribute to differential expressions of genes between normal and matched-cancer tissues.

Ethical approval

The biospecimens for this study were provided by the Keimyung University Dongsan Hospital Korea Regional Biobank, the Biobank of Chonnam National University Hwasun Hospital, and Pusan National University Hospital, a member of the National Biobank of Korea, which is supported by the Ministry of Health and Welfare. All samples derived from the National Biobank of Korea were obtained with informed consent under institutional review board-approved protocols.

CRediT authorship contribution statement

Songmi Kim: Conceptualization, Methodology, Validation, Writing - original draft. **Wonseok Shin:** Validation, Visualization, Investigation. **Yong-Moon Lee:** Validation, Writing - review & editing. **Seyoung Mun:** Data curation, Writing - review & editing. **Kyudong Han:** Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationship that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by Basic Science Research Capacity Enhancement Project through Korea Basic Science Institute (National research Facilities and Equipment Center) grant funded by the Ministry of Education for Basic Science Research Capacity Enhancement Project (Grant No. 2019R1A6C1010033). The present research was supported by the research fund of Dankook university in 2019 for the University Innovation Support Program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ab.2020.113769.

References

[1] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[2] P.L. Deininger, J.V. Moran, M.A. Batzer, H.H. Kazazian Jr., Mobile elements and mammalian genome evolution, Curr. Opin. Genet. Dev. 13 (2003) 651–658.
[3] A.F. Smit, Interspersed repeats and other mementos of transposable elements in mammalian genomes, Curr. Opin. Genet. Dev. 9 (1999) 657–663.
[4] J. Lee, R. Cordaux, K. Han, J. Wang, D.J. Hedges, P. Liang, M.A. Batzer, Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons, Gene 390 (2007) 18–27.
[5] G.J. Cost, Q. Feng, A. Jacquier, J.D. Boeke, Human L1 element target-primed reverse transcription in vitro, EMBO J. 21 (2002) 5899–5910.
[6] D.C. Hancks, H.H. Kazazian Jr., Roles for retrotransposon insertions in human disease, Mobile DNA 7 (2016) 9.
[7] D.D. Luan, M.H. Korman, J.L. Jakubczak, T.H. Eickbush, Reverse transcription of

R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition, Cell 72 (1993) 595–605.
[8] W. Shin, S. Mun, J. Kim, W. Lee, D.G. Park, S. Choi, T.Y. Lee, S. Cha, K. Han, Novel discovery of LINE-1 in a Korean individual by a target enrichment method, Mol. Cell 42 (2019) 87–95.
[9] H.A. Cruickshanks, C. Tufarelli, Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter, Genomics 94 (2009) 397–406.
[10] M. Speek, Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes, Mol. Cell Biol. 21 (2001) 1973–1985.
[11] L. Lavie, E. Maldener, B. Brouha, E.U. Meese, J. Mayer, The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity, Genome Res. 14 (2004) 2253–2260.
[12] J. Lee, S. Mun, T.J. Meyer, K. Han, High levels of sequence diversity in the 5' UTRs of human-specific L1 elements, Comp. Funct. Genom. (2012) 129416 2012.
[13] R. Minakami, K. Kurose, K. Etoh, Y. Furuhata, M. Hattori, Y. Sakaki, Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element, Nucleic Acids Res. 20 (1992) 3139–3145.
[14] G.D. Swergold, Identification, characterization, and cell specificity of a human LINE-1 promoter, Mol. Cell Biol. 10 (1990) 6718–6729.
[15] T. Tchenio, J.F. Casella, T. Heidmann, Members of the SRY family regulate the human LINE retrotransposons, Nucleic Acids Res. 28 (2000) 411–415.
[16] N. Yang, L. Zhang, Y. Zhang, H.H. Kazazian Jr., An important role for RUNX3 in human L1 transcription and retrotransposition, Nucleic Acids Res. 31 (2003) 4929–4940.
[17] A.M. Denli, I. Narvaiza, B.E. Kerman, M. Pena, C. Benner, M.C. Marchetto, J.K. Diedrich, A. Aslanian, J. Ma, J.J. Moresco, et al., Primate-specific ORF0 contributes to retrotransposon-mediated diversity, Cell 163 (2015) 583–593.
[18] U. Miglio, E. Berrino, M. Panero, G. Ferrero, L. Coscujuela Tarrero, V. Miano, C. Dell'Aglio, I. Sarotto, L. Annaratone, C. Marchio, et al., The expression of LINE1-MET chimeric transcript identifies a subgroup of aggressive breast cancers, Int. J. Canc. 143 (2018) 2838–2848.
[19] K. Matlik, K. Redik, M. Speek, L1 antisense promoter drives tissue-specific transcription of human genes, J. Biomed. Biotechnol. (2006) 71753 2006.
[20] P. Nigumann, K. Redik, K. Matlik, M. Speek, Many human genes are transcribed from the antisense promoter of L1 retrotransposon, Genomics 79 (2002) 628–634.
[21] K. Hur, P. Cejas, J. Feliu, J. Moreno-Rubio, E. Burgos, C.R. Boland, A. Goel, Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis, Gut 63 (2014) 635–646.
[22] E.M. Wolff, H.M. Byun, H.F. Han, S. Sharma, P.W. Nichols, K.D. Siegmund, A.S. Yang, P.A. Jones, G. Liang, Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer, PLoS Genet. 6 (2010) e1000917.
[23] K.W. Jung, Y.J. Won, H.J. Kong, E.S. Lee, Prediction of cancer incidence and mortality in Korea, Canc. Res. Treat 51 (2019) 431–437 2019.
[24] S.W. Criscione, N. Theodosakis, G. Micevic, T.C. Cornish, K.H. Burns, N. Neretti, N. Rodic, Genome-wide characterization of human L1 antisense promoter-driven transcripts, BMC Genom. 17 (2016) 463.
[25] S. Kim, Y.J. Kim, K. Han, Analysis of L1-chimeric transcripts derived from bidirectional promoter of human-specific L1, Gene Genom. 38 (2016) 69–79.
[26] M. Ryan, W.C. Wong, R. Brown, R. Akbani, X. Su, B. Broom, J. Melott, J. Weinstein, TCGASpliceSeq a compendium of alternative mRNA splicing in cancer, Nucleic Acids Res. 44 (2016) D1018–D1022.
[27] W. Sun, T. Duan, P. Ye, K. Chen, G. Zhang, M. Lai, H. Zhang, TSVdb: a web-tool for TCGA splicing variants analysis, BMC Genom. 19 (2018) 405.
[28] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, Methods 25 (2001) 402–408.
[29] X. Rao, X. Huang, Z. Zhou, X. Lin, An improvement of the 2^(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis, Biostat. Bioinform. Biomath. 3 (2013) 71–85.
[30] P. Machanick, T.L. Bailey, MEME-ChIP: motif analysis of large DNA datasets, Bioinformatics 27 (2011) 1696–1697.
[31] T. Penzkofer, T. Dandekar, T. Zemojtel, L1Base: from functional annotation to prediction of active LINE-1 elements, Nucleic Acids Res. 33 (2005) D498–D500.
[32] S. Kim, C.S. Cho, K. Han, J. Lee, Structural variation of alu element and human disease, Genom. Inform. 14 (2016) 70–77.
[33] N. Yang, H.H. Kazazian Jr., L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells, Nat. Struct. Mol. Biol. 13 (2006) 763–771.
[34] S. Lin, L. Zhang, W. Luo, X. Zhang, Characteristics of antisense transcript promoters and the regulation of their activity, Int. J. Mol. Sci. 17 (2015).
[35] D.C. Hancks, H.H. Kazazian Jr., Active human retrotransposons: variation and disease, Curr. Opin. Genet. Dev. 22 (2012) 191–203.
[36] N. Kawasaki, K. Isogaya, S. Dan, T. Yamori, H. Takano, R. Yao, Y. Morishita, L. Taguchi, M. Morikawa, C.H. Heldin, et al., TUFT1 interacts with RABGAP1 and regulates mTORC1 signaling, Cell Discov. 4 (2018) 1.
[37] S. Kuroda, M. Oyasu, M. Kawakami, N. Kanayama, K. Tanizawa, N. Saito, T. Abe, S. Matsuhashi, K. Ting, Biochemical characterization and expression analysis of neural thrombospondin-1-like proteins NELL1 and NELL2, Biochem. Biophys. Res. Commun. 265 (1999) 79–86.
[38] T.K. Watanabe, T. Katagiri, M. Suzuki, F. Shimizu, T. Fujiwara, N. Kanemoto, Y. Nakamura, Y. Hirai, H. Maekawa, E. Takahashi, Cloning and characterization of two novel human cDNAs (NELL1 and NELL2) encoding proteins with six EGF-like repeats, Genomics 38 (1996) 273–276.
[39] K. Aihara, S. Kuroda, N. Kanayama, S. Matsuyama, K. Tanizawa, M. Horie, A neuron-specific EGF family protein, NELL2, promotes survival of neurons through mitogen-activated protein kinases, Brain Res. Mol. Res. 116 (2003) 86–93.
[40] C. Zhu, T. Utsunomiya, T. Ikemoto, S. Yamada, Y. Morine, S. Imura, Y. Arakawa, C. Takasu, D. Ishikawa, I. Imoto, et al., Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma, Ann. Surg Oncol. 21 (Suppl 4) (2014) S729–S735.
[41] A.E. Gururaj, S. Peng, R.K. Vadlamudi, R. Kumar, Estrogen induces expression of BCAS3, a novel estrogen receptor-alpha coactivator, through proline-, glutamic acid-, and leucine-rich protein-1 (PELP1), Mol. Endocrinol. 21 (2007) 1847–1860.
[42] G. Lev-Maor, R. Sorek, N. Shomron, G. Ast, The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons, Science 300 (2003) 1288–1291.
[43] R. Sorek, G. Ast, D. Graur, Alu-containing exons are alternatively spliced, Genome Res. 12 (2002) 1060–1067.
[44] K. Chalitchagorn, S. Shuangshoti, N. Hourpai, N. Kongruttanachok, P. Tangkijvanich, D. Thong-ngam, N. Voravud, V. Sriuranpong, A. Mutirangura, Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis, Oncogene 23 (2004) 8841–8846.
[45] C. Phokaew, S. Kowudtitham, K. Subbalekha, S. Shuangshoti, A. Mutirangura, LINE-1 methylation patterns of different loci in normal and cancerous cells, Nucleic Acids Res. 36 (2008) 5704–5712.