Available online at www.sciencedirect.com
ScienceDirect
Gene 411 (2008) 38–45
GENE
www.elsevier.com/locate/gene

The impact of multiple splice sites in human L1 elements

V.P. Belancio, A.M. Roy-Engel, P. Deininger *

Tulane Cancer Center, SL66, Department of Epidemiology, Tulane University Health Sciences Center, 1430 Tulane Ave., New Orleans, LA 70112, United States

Received 23 October 2007; received in revised form 20 December 2007; accepted 28 December 2007
Available online 11 January 2008
Received by C. Feschotte

Abstract

LINE-1 elements represent a significant proportion of mammalian genomes. The impact of their activity on the structure and function of the host genomes has been recognized from the time of their discovery as an endogenous source of insertional mutagenesis. L1 elements contain numerous functional internal polyadenylation signals and splice sites that generate a variety of processed L1 transcripts. These sites are also reported to contribute to the generation of hybrid transcripts between L1 elements and host genes. Using northern blot analysis we demonstrate that L1 splicing, but not L1 polyadenylation, is delayed during the course of L1 expression. L1 splicing can also be negatively regulated by EBV SM protein known to alter this process. These results suggest a potential for L1 mRNA processing to be regulated in a tissue- and/or development-specific manner. The delay in L1 splicing may also serve to protect host genes from the excessive burden of L1 interference with their normal expression via aberrant splicing.

Keywords: LINE-1; Alternative splicing; Retroelement; Gene regulation; Mobile elements; RNA processing

1. Introduction

Long Interspersed Element-1, LINE-1 or L1, is a long-term resident of mammalian genomes that occupies up to 17% of the total DNA (Lander et al., 2001; Waterston et al., 2002). Insertional mutagenesis of L1 and its parasites, Small Interspersed Elements (SINEs) and SVA elements, has been recognized to contribute to human disease and cancer origins and progression (Deininger et al., 2003; Kazazian and Goodier, 2002; Ostertag et al., 2003).

The majority of the L1 inserts in mammalian genomes are 5′ truncated (Grimaldi et al., 1984; Lander et al., 2001; Waterston et al., 2002). Full-length L1 (FL1) elements contain an internal RNA polymerase II promoter that drives expression of a bicistronic mRNA terminating at the polyadenylation (polyA) site located at the 3′ end of the L1 sequence (Belancio et al., 2007; Kazazian et al., 1988; Swergold, 1990). The L1 mRNA encodes two open reading frames ORF1 and 2 that are essential for successful integration (Moran et al., 1996). Insertional mutagenesis associated with L1 activity relies on the production of the FL1 mRNA, ORF1 and ORF2 proteins for L1 integration (Moran et al., 1996), and only functional L1 ORF2 protein for mobilization of SINEs (Dewannieux et al., 2003).

Multiple mechanisms exist to control retrotransposition and expression of L1 elements. Cellular factors involved in retroviral defense, such as the APOBEC family of proteins are known to modulate L1 integration (Bogerd et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006). L1 promoter strength and methylation serve as a first limiting factor in the production of the functional full-length L1 mRNA (Hata and Sakaki, 1997; Matlik et al., 2006; Swergold, 1990). However, efficient transcription from the L1 promoter does not manufacture high levels of the FL1

Abbreviations: APOBEC, cytidine deaminases family of proteins; EBV SM protein, Epstein–Barr virus SM protein; FL1, Full-length L1; LINE-1, L1, Long interspersed element-1; L1.3, one of the active human L1 elements; L1spa, one of the active mouse L1 elements; ORF1, open reading frame 1; ORF2, open reading frame 2; PolyA, polyadenylation site; SINE, Short interspersed element; SC35 splicing factor; UTR, untranslated region.

* Corresponding author. Tel.: +1 504 988 6385.
E-mail address: pdeinin@tulane.edu (P. Deininger).

0378-1119/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.12.022

due to the complex and extensive processing of L1 RNA via premature polyadenylation at the internal polyA sites and splicing (Belancio et al., 2006; Perepelitsa-Belancio and Deininger, 2003). These processes restrain L1 expression even in cancer cells where L1 promoter methylation is lessened (Belancio et al., 2006; Ehrlich, 2002; Perepelitsa-Belancio and Deininger, 2003).

RNA processing by splicing and polyadenylation is often regulated in a developmental and tissue-specific manner, and altered upon malignant transformation (Kalnina et al., 2005; Wallace et al., 1999; Yeo et al., 2004). There is also a complex interplay between the two processes (Tian et al., 2007). Most of the splice and polyadenylation sites subjected to such regulation are weak (Batt et al., 1994; Garg and Green, 2007). The majority of the sites (functional and predicted) present within the L1 sequence fall into this category (Belancio et al., 2006; Perepelitsa-Belancio and Deininger, 2003). We have previously reported variation in the L1 processing in cancer cell lines of different origin (Belancio et al., 2006). This variation suggests a possibility that L1 mRNA splicing and polyadenylation could be regulated not only in normal cells but also in malignances.

Cis-signals for splicing and premature polyadenylation present in the L1 RNA sequence can also contribute to pre-mRNA processing of host genes that contain L1 elements. Multiple examples of the hybrid transcripts generated by splicing between L1 and human genes have been reported (Belancio et al., 2006; Matlik et al., 2006; Wheelan et al., 2005). We demonstrate that L1 RNA splicing is delayed, while premature polyadenylation of the L1-related transcripts occurs rapidly. This is likely to be due to the inefficient recognition/processing of the L1 splice sites. We propose a model in which this delay minimizes the negative impacts of L1 sequences residing within the introns of mammalian genes.

2. Materials and methods

2.1. Cell culture

NIH 3T3 (ATCC CRL-1658) and HeLa (ATCC CCL2) cells were maintained as described (Belancio et al., 2006; Perepelitsa-Belancio and Deininger, 2003).

2.2. Transient transfections

5×10^6 NIH 3T3 or HeLa cells were plated per T75 tissue culture flask (CORNING) and transiently transfected with 6 μg of L1.3 or L1Neo expression vectors 17–20 h after plating using Lipofectamine Plus Reagent (InVitrogen) as described before (Perepelitsa-Belancio and Deininger, 2003). For co-transfection experiments with EBV SM protein, the same number of cells was transfected with 5 μg of the L1.3 expression cassette and 3 μg of vector expressing EBV SM protein in either reverse or forward orientations using Lipofectamine Plus Reagent (InVitrogen). For the experiments that test the effect of the L1-related products on the onset of L1 splicing, 2.5×10^6 cells were transfected with 2 μg of L1 ORF2, L1spa, or L15′UTRluc expression vectors using Lipofectamine Plus Reagent (InVitrogen). 24 h later, cells were transfected again with 6 μg of the L1 expression cassette and 1 μg of the expression vectors mentioned above using Lipofectamine Plus Reagent (InVitrogen). RNAs were harvested 9 and 24 h post-transfection. Note that the post-transfection time is indicated as the time after the second transfection cocktail was added to the cells. The pBudCE4.1 basic expression vector was used as a negative control for transfections with L1 ORF2, and L1spa expression vectors. Firefly luciferase driven by SV40 promoter (pGL3-Promoter, Promega) was used as a negative control for the pGL3 vector expressing Firefly luciferase driven by the L1.3 5′ UTR.

2.3. Northern blot analysis

RNA extractions and northern blot analysis was performed as previously described (Belancio et al., 2006; Perepelitsa-Belancio and Deininger, 2003).

2.4. Plasmids

JM101/L1.3notag (L1.3) (Wei et al., 2001) and JM101/L1.3 (L1.3Neo) vectors are gifts from Dr. J. Moran (Sassaman et al., 1997). L1spa expression vector is a gift from Dr. Kazazian (Naas et al., 1998). EBV aSM and SM expression vectors are gifts from Dr. S. Swaminathan (Ruvolo et al., 1998). pBudORF2opt (Gasior et al., 2006) was created using the codon optimized $L1^{RP}$ as a source for the ORF2 coding sequences. The open reading frames were cloned into the expression vector pBudCE4.1 (InVitrogen), under control of the CMV promoter. The 5′ UTR expression vector was constructed by subcloning L1.3 5′ UTR sequence into pGL3-basic vector (Promega) to drive the expression of the Firefly luciferase gene (El Sawy et al., 2005).

3. Results

3.1. LINE-1 polyadenylation and splicing are differentially regulated

Multiple L1 loci likely undergo de-repression by the loss of promoter methylation during embryogenesis and carcinogenesis (Ehrlich, 2002; Lees-Murdock et al., 2003; Mays-Hoopes et al., 1986) or activation due to altered transcription factor availability (Tchenio et al., 2000; Yang et al., 2003). To analyze RNA species produced during progression of L1 expression, we performed a northern blot analysis of the time course of the transiently transfected wild-type L1.3 expression vector in HeLa cells (Fig. 1) with a strand-specific RNA probe. This probe is complementary to the first 100 bp of the L1 5′ UTR (5′ UTR 100 probe) and therefore detects both spliced and prematurely polyadenylated L1 transcripts (Fig. 1A) (Belancio et al., 2006). This northern blot analysis demonstrated a steady accumulation of the full-length L1 mRNA (1, 1.31, and 2.15 relative units at 6, 9, and 24 h post-transfection, respectively) and the previously characterized prematurely polyadenylated L1-related products (1, 1.26, and 2.3 relative units at 6, 9, and 24 h post-transfection, respectively) between 6 and 24 h after the addition of the transfection cocktail. Intriguingly, spliced L1-related mRNAs were almost undetectable in the early time points (less than 10% of the polyadenylated L1