AD_____

Award Number: DAMD17-02-1-0597

TITLE: Retroelements and Genetic Instability in Breast Cancer

PRINCIPAL INVESTIGATOR: Victoria P. Perepelitsa

CONTRACTING ORGANIZATION: Tulane University Health Sciences Center
New Orleans, Louisiana 70112

REPORT DATE: April 2003

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2003	Annual Summary (1 Apr 02 - 31 Mar 03)

4. TITLE AND SUBTITLE
Retroelements and Genetic Instability in Breast Cancer

5. FUNDING NUMBERS
DAMD17-02-1-0597

6. AUTHOR(S):
Victoria P. Perepelitsa

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Tulane University Health Sciences Center
New Orleans, Louisiana 70112

E-Mail: vperepe@tulane.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*

LINE1 is the most prevalent human retroelement, and it contributes to genomic instability. The full extent of LINE-1 mobility in somatic tissues and tumors is not known. L1 expression is extremely low in differentiated cells except for testis, but it is significantly elevated in some breast cancer cells. Full-length LINE-1 RNA is not detected even when transiently expressed from the CMV promoter. These observations suggest that posttranscriptional mechanisms are involved in limitation of L1 expression.
By using a polyadq program, we identified 20 putative polyadenylation (polyA) sites located only in the sense strand of the L1.3 genome. We hypothesize that the use of the putative polyA sites located within the L1.3 genome limit the amount of full-length L1.3 mRNAs present in mammalian cells.
We developed a system that allows detection of the full-length L1 mRNA and any RNA species produced through internal polyadenylation. We determined that numerous putative polyadenylation sites found in L1 elements are functional and they attenuate full-length L1 RNA formation by about 50 fold. There is tremendous redundancy in these internal polyadenylation sites. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome.

14. SUBJECT TERMS:	15. NUMBER OF PAGES
regulation of gene expression, LINE1, polyadenylation, genomic instability, tissue specific expression	38
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Cover.....1

SF 298.....2

Table of Contents.....3

Introduction.....4

Body.....5

Key Research Accomplishments.....9

Reportable Outcomes.....10

Conclusions.....11

References.....12

Appendices.....13

Introduction.

LINE1 is the most prevalent human retroelement, and it contributes to genomic instability. The full extent of LINE-1 mobility in somatic tissues and tumors is not known. L1 expression is extremely low in differentiated cells except for testis, but it is significantly elevated in cancer cells including breast cancer cells. Full-length LINE-1 RNA is not detected even when transiently expressed from the CMV promoter. These observations suggest that posttranscriptional mechanisms are involved in limitation of L1 expression. By using a polyadq program, we identified 20 putative polyadenylation (polyA) sites located only in the sense strand of the L1.3 genome. We hypothesize that the use of the putative polyA sites located within the L1.3 genome limit the amount of full-length L1.3 mRNAs present in mammalian cells.We developed a system that allows detection of the full-length L1 mRNA and any RNA species produced through internal polyadenylation. We determined that numerous putative polyadenylation sites found in L1 elements are functional and they attenuate full-length L1 RNA formation by about 50 fold. There is tremendous redundancy in these internal polyadenylation sites. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome.

Body.

According with the approved Statement of Work we have made the following progress in the first year of funding:

Task 1. To identify functional polyA sites in L1.3 genome (Months 1-25)

A-B. The preliminary data included in the proposal were obtained by using a reporter gene system with the fragment of human L1.3 sequence, containing a cluster of the putative polyadenylation (polyA) signals, cloned in the intron immediately upstream of the reporter gene (opiluc system). This system detected significant decrease in the expression of the reporter gene from the construct containing the mentioned above L1.3 fragment compared to the control vector that did not have any L1.3 sequence. This result was consistent with the putative polyA signals being functional. At the time of submission we did not have an appropriate positive control, a construct containing a strong functional polyA signal, available. We constructed a vector that had SV40 late or SV40 early polyA signals cloned in the intron of the reporter construct. When tested, neither one of these vectors demonstrated any significant reduction in the reporter gene expression. Based on this result, we rendered the system unfit for our study. We have learned from the literature that there is a competition between splicing and polyadenylation, particularly, when both signals are located in the close vicinity of one another(2). It is possible, therefore, that the 300 bp. SV40 polyA signal could be more prone to this competition in our system then the 1700 bp. L1.3 fragment.

C-F. We next made a series of FRIRES constructs described in the submitted proposal. We cloned equal length (1kb.) L1.3 fragments, L1.3 fragments containing clusters of the putative polyA sites or individual putative polyA sites, and L1.3 fragments containing mutations of some of these putative polyA sites in between two reporter genes, Firefly (F) and Renilla (R) luciferase. Based on the ratios of the reporter genes of individual constructs and those of the first reporter gene between constructs, this system would allow detection of both functional polyA signals and potential RNA instability associated with the introduced fragments. The SV40 late and early polyA signals were used as positive control, and they both significantly reduced the expression of the second reporter gene (Fig.1, page 13 of appendices). This result was consistent with a functional polyA signal located downstream of the first reporter gene. Introduction of L1.3 fragments in this system resulted in different degree of reduction in the expression of the first reporter gene for some constructs. Unexpectedly, it also led to the overexpression of the second reporter gene (FR#2,4,5,6, figure 1, page *13* of appendices). This result could be partially explained by the presence of the IRES element that drives the expression of the Rluciferase. It has been described for some IRES containing systems that down regulation of translation of the upstream open reading frame (ORF) results in the more efficient IRES utilization. We

determined by Northern blotting assay of the polyA selected mRNAs from the cells transiently transfected with FRIRES series constructs that (i) our positive control was adequate, i. e. introduction of the SV40 polyA signal results in predominant production of mRNA species containing only the first reporter gene (Fig.2 page *14* of appendices), (ii) introduction of L1.3 fragments resulted in the decrease of the full-length mRNA produced (Fig.2, page 14 of appendices, and data not shown), which would be consistent with the reduction of the (F)luciferase activity mentioned previously. We were not able to unambiguously confirm by Northern blotting assay whether the putative polyA signals present in the L1.3 fragments were functional or not. The difficulty came from the fact that the expected products comigrated with the ribosomal RNA. We were able to almost completely eliminate rRNA by performing a double polyA selection of mRNAs from transient transfections, however, this additional step also resulted in the loss of the sensitivity of the assay. As an alternative, we deleted the IRES and the second reporter gene from the original series of FRIRES constructs, which resulted in a series of vectors that contained L1.3 fragments in their 3' untranslated region (UTR) called FRM/N. We performed a quantitative Northern blotting assay (by normalizing the amount of the Fluciferase mRNA produced from each construct to the amount of neomycin mRNA made from the same vector) of the polyA selected RNA species from the transient transfections with this series of constructs (Fig.3 page 15 of appendices). Our data demonstrated that each of the tested L1.3 fragments resulted in a significant reduction of the (F)luciferase mRNA. In this assay we have not detected any faster-migrating RNA species, which would be consistent with the utilization of the L1.3 putative polyA signals. This could be due to the competition between the relatively weak putative polyA signals located in the L1.3 fragments and a very strong SV40 polyA site present immediately downstream (3). These results demonstrate that L1.3 insertions that do not alter open reading frame of a gene can significantly decrease the total amount of mRNA produced by this gene potentially through their effect on mRNA stability.

Due to the inconclusive results from both of the systems, an alternative strategy has been employed.

To avoid any complications associated with artificial systems, we developed a sensitive northern blotting assay that detects human L1.3 RNA species produced from a vector containing a full-length L1.3 element capable of retrotransposition in transient transfections (this vector was described in the proposal for testing of the effect of mutations of the functional polyA signals on the rate of retrotransposition). This approach, even though, the most biologically relevant, highly innovative, and most useful for the future work on L1 elements was not proposed as a main method due to initial difficulties with the sensitivity of the assay. The development of this assay allowed us to detect for the first time the full-length L1.3 mRNA from transiently transfected cells. This opens a great opportunity to directly observe the effect of any genetic manipulations introduced into the L1.3 genome on the amount of the full-length mRNA, and the correlation of the latter with the rate of retrotransposition (Fig.2B of the appended manuscript). In addition to the full-length L1.3 mRNA we also detected multiple faster-migrating polyadenylated L1-specific RNA species (Fig.2B of the appended manuscript). The relative size of these RNAs roughly corresponded to the positions of the putative polyA signals identified in the L1.3 genome (Fig.1A of the appended manuscript). The size of the most abundant RNA

detected corresponded to the position of the strongest predicted putative polyA signal. These faster migrating RNA species were major products of L1.3 expression with the full-length L1.3 mRNA represented at roughly 2% of the total L1.3 RNA produced. We introduced two point mutations into the most conserved element of the polyA site, the AAUAAA hexamer. The major RNA species observed in the wild-type L1.3 completely disappeared in the mutant retroelement, demonstrating that the strongest predicted polyA signal is functional, and it is solely responsible for the 3'end formation of that RNA (Fig.2B of the appended manuscript). Interestingly, abolishment of this polyA signal resulted in a more effective utilization of the polyA sites located immediately up- or downstream of it. To further confirm the usage of this polyA site we performed a 3'RACE on the region predicted to produce major RNA species. We transiently transfected chicken fibroblasts (avian species do not have any endogenous L1 elements) to avoid any background. Consistent with the northern blotting results, the band corresponding to the strongest polyA site disappeared upon mutation (Fig.3 of the appended manuscript). Sequence analysis of this and other bands detected by 3'RACE demonstrated that both canonical (AAUAAA), second most common (AUUAAA) polyA signals and noncanonical (one base substitutions of AAUAAA) ones are utilized during L1.3 transcription. We are currently in the process of mutating canonical as well as strongest predicted noncanonical polyA signals present in the L1.3 genome. We have introduced mutations into the three most 5' canonical and the strongest predicted noncanonical polyA sites. The results obtained from these mutants demonstrate that disappearance of any given polyA signal in the L1.3 genome leads to a more efficient utilization of the other ones in the vicinity or downstream of it. Destruction of four polyA signals did not result in any significant increase in the amount of the full-length L1.3 mRNA. The presence of a great number of both classical and noncanonical functional polyA signals in the L1 elements (Fig.1A and 1B of the appended manuscript) ensures that a loss of even four of them (potentially due to the high mutation rate of the reverse transcriptase) would not be devastating to the host genome. This finding has a tremendous impact on our understanding of the process of L1 expression and its consequences on the human genome. It is clear that L1 elements evolved to have a very sophisticated and redundant mechanism to attenuate their expression. It is possible that the polyA signals can also play a regulatory role. Alternative polyadenylation is utilized fairly widely in a tissue- and development-specific manner, with testis been a hot spot for this process (1). If the polyA sites are "ignored" by the polyadenylation machinery, it can result in the increased amount of the full-length L1 mRNA, and, potentially, the rate of retrotransposition.

Task 3. To identify the effect of functional polyA sites on the rate of L1 retrotransposition (Months 29-36)

A. We introduced mutations into four functional polyA sites found in the full-length L1.3 expression vector. These mutations, however, did not significantly changes the amount of the full-length mRNA produced (based on the results of northern blotting). We will proceed with this task as soon as we have a mutant that produces higher amounts of the full-length L1.3 mRNA. We are currently having the L1.3 synthesized to eliminate all predicted canonical polyA signals.

Recommended changes.

Task 2. To identify the function and relative strength of functional polyA sites in normal and breast cancer cells.

We will perform this task with our sensitive and more direct northern blotting system instead of opiluc or FRIRES-based reporter systems.

Key Research Accomplishments.

Task 1. To identify functional polyA sites in L1.3 genome (Months 1-25)
Task 3a. To identify the effect of functional polyA sites on the rate of L1 retrotransposition (Months 29-36)

- We developed a system that allows detection of functional polyA signals in the L1 elements.
- We determined that putative polyA sites found in the coding region of the human L1.3 element are functional.
- We performed site-directed mutagenesis to of some of the polyA signals to determine the loss of function.
- We determined that both canonical and noncanonical polyA sites are utilized during L1.3 expression.
- We demonstrated by sequence analyses that the cleavage and polyadenylation at these polyA sites is in accordance with the distances and sequences described for the process.
- Our data demonstrates that there is tremendous redundancy of the usage of these polyA signals.
- We demonstrate that utilization of these internal polyA sites attenuates L1.3 expression by about 50 fold.

Reportable outcomes:

1. An appended manuscript "RNA truncation by premature polyadenylation attenuates human mobile element activity" has been submitted to the journal of *Nature Genetics.*

2. An appended abstract "LINE1 polyadenyletion sites or stability limit elements expression" for the poster that was presented at the Keystone Symposia meeting "Transposition and Other Genome Rearrangements" in Santa Fe, NM, February 8-14, 2003.

3. An appended abstract "RNA truncation by premature polyadenylation attenuates human mobile element activity " for the poster that was presented at the 15th Annual Tulane Research day on March 25, 2003.

Conclusions:

1. We developed a new approach for detection and analysis of the RNA species generated from a transiently transfected active human L1 element. This method allows for the first time to (i) directly detect the consequences of the genetic manipulations of the L1 genome on the produced mRNA; (ii) correlate the amount of the full-length L1 mRNA with the rate of retrotransposition *in vitro*.

2. We demonstrated that the putative polyA sites found in the coding region of the human L1.3 element are functional, and they attenuate the amount of the full-length mRNA by about 50 fold. This establishes a novel application of the conventional cellular machinery for transcriptional regulation of a gene expressed in a mammalian genome.

3. We demonstrated that both canonical and noncanonical polyA sites are utilized during L1.3 expression.

4. There is tremendous redundancy in this system, i.e. mutation of any or even multiple polyA signals leads to a more efficient usage of the multiple other ones distributed through out the genome.

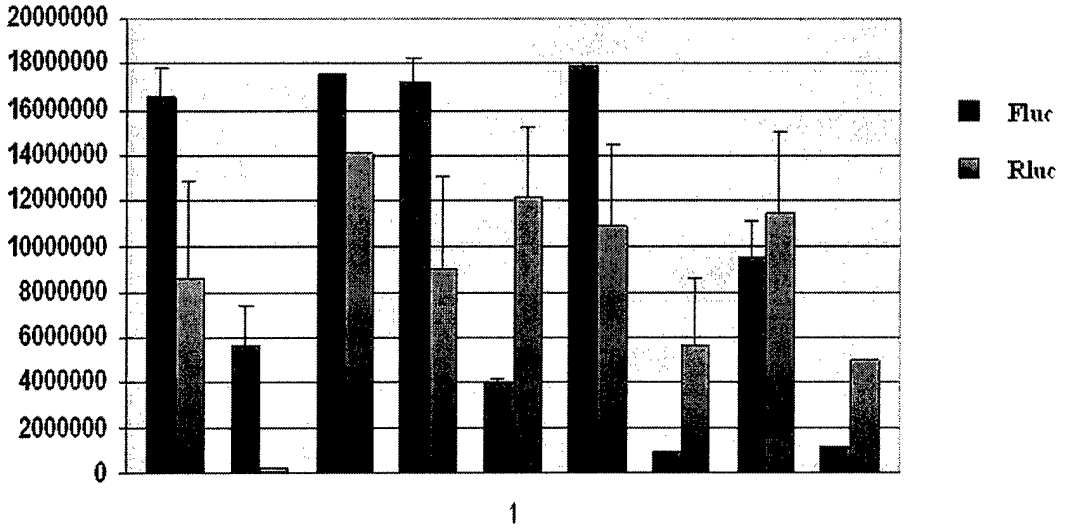
Reference List

1. **Branciforte, D. and S. L. Martin**. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol.Cell Biol. **14**:2584-2592.

2. **Peterson, M. L., M. B. Bryman, M. Peiter, and C. Cowan**. 1994. Exon size affects competition between splicing and cleavage- polyadenylation in the immunoglobulin mu gene. Mol.Cell Biol. **14**:77-86.

3. **Proudfoot, N. J., A. Furger, and M. J. Dye**. 2002. Integrating mRNA processing with transcription. Cell **108**:501-512.

Appendices.



	FR	SV40	5'LIP	LIP	FR#2	FR#3	FR#4	FR#5	FR#6
F/R	2.4	22	1.3	2.3	0.3	1.8	0.2	0.9	0.2
StDev	0.86	5.49		0.78	0.06	0.51	0.11	0.36	

Figure1. Reporter gene activity of pFRIRES-based constructs in 3T3 cells. Average ratios of Fluciferase to Rluciferase (F/R) and standard deviation for these ratios (StDev) obtained from at least tree independent experiments are shown. FR is a control vector, SV40 is a vector that has an SV40 late polyA site cloned into FR, 5'LIP and LIP are 5' and 3' fragments of the L1.3 promoter, #2-#6 are 1kb fragments of the L1.3 coding region.

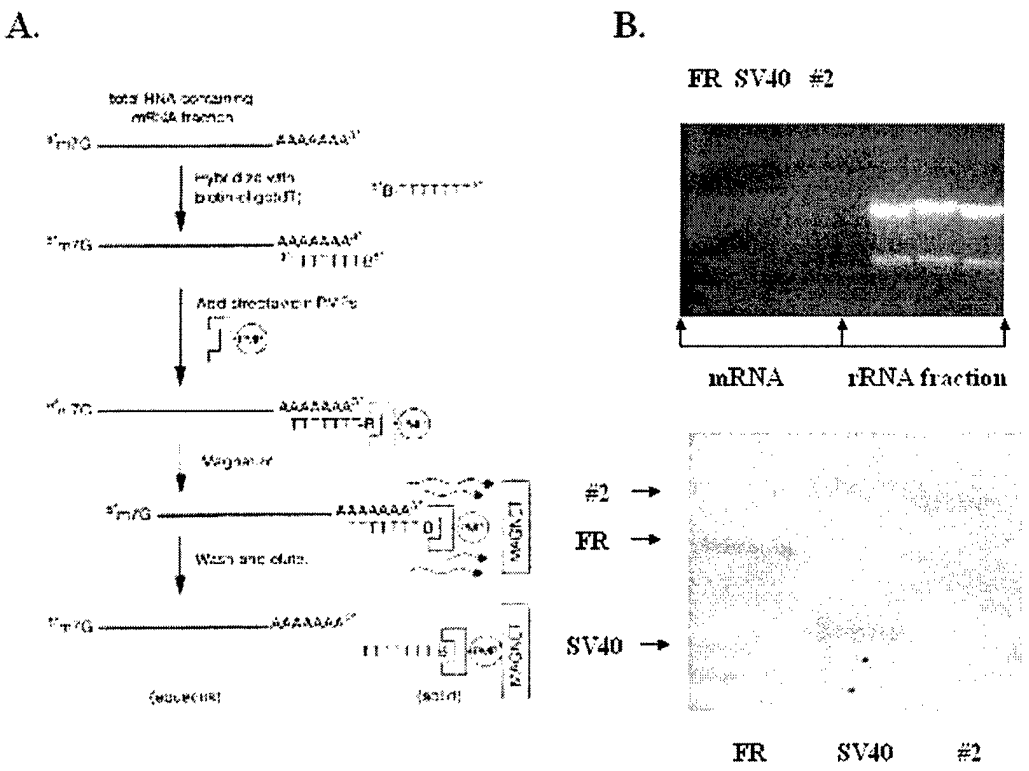


Figure 2. Northern blotting assay. A. A schematic representation of the polyA selection system (Promega) that was applied to NIH 3T3 cells transiently transfected with FR, SV40, and #2 constructs, B. 1% formaldehyde gel (top) showing the polyA selected fraction (mRNA) and rRNA fraction, autoradiograph of the mRNA fraction with randomly (^{32}P)-labeled (F)luciferase probe (bottom).

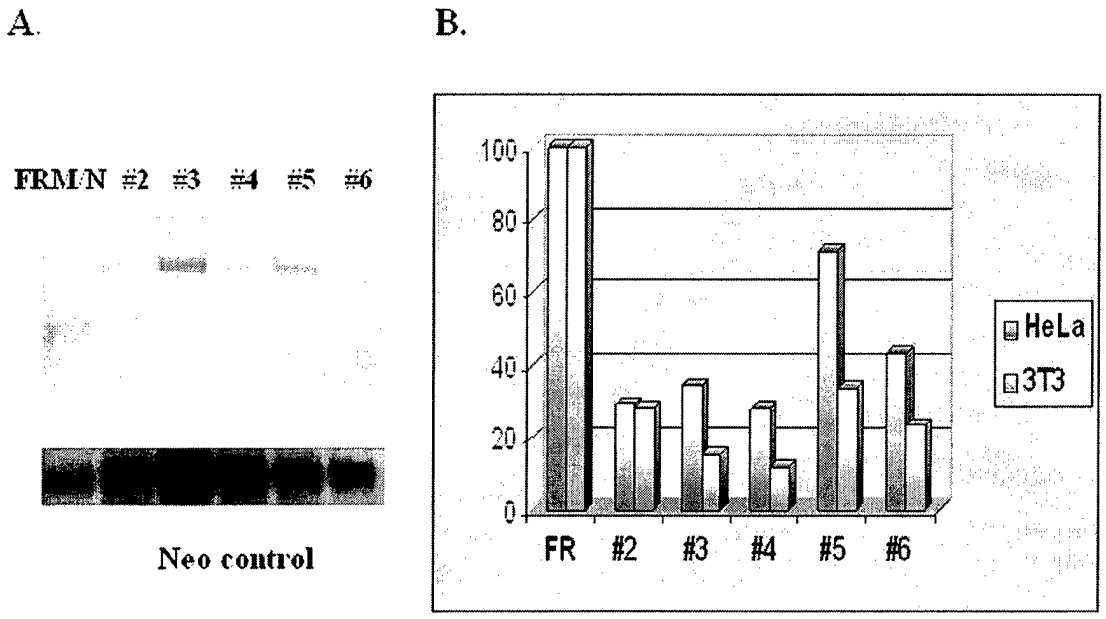


Figure 3. Effect of L1.3 fragments on the amount of mRNA produced in NIH 3T3 and HeLa cells. A. Northern blot of NIH 3T3 cells transiently transfected with FRM/N series of constructs, polyA selected, fractionated on a 1% formaldehyde gel, probed with randomly labeled probe to Fluciferase gene (top) and neo gene (bottom). B. Normalized data obtained from NIH 3T3 and HeLa cells that shows relative amount of mRNA produced from FRM/N series of constructs expressed as percent of the control vector FRM/N.

LINE1 polyadenyletion sites or stability limit elements expression.

Victoria P. Belancio and Dr. Prescott Deininger, Molecular and Cellular Biology,

Tulane University, New Orleans, LA, 70130

LINE1 is the most prevalent human retroelement, and it contributes to genomic instability. LINE1 elements comprise 17% of the human genome, which translates into about $5x10^5$ L1 copies, the majority of which are truncated at their 5' end. The full extent of LINE-1 mobility in somatic tissues and particularly in tumors is still not known.
During the course of evolution the activity of LINEs has tremendously decreased. LINE-1 is the only member of the autonomous non-LTR retrotransposones family that is currently active. It is believed that about 60 active copies of full-length LINE-1 are present in the human genome.
L1 expression is extremely low in all cell types of a mature organism except for testis. In contrast, significantly higher levels of LINE-1 expression were found in various cancer cells. The known factors involved in regulation of LINE1 expression, such as hypomethylation often associated with malignant transformation and promoter activity, cannot fully explain the observed pattern of expression.
In addition, full-length LINE-1 is not detected by Northern blotting in cell culture even when transiently expressed from the CMV promoter. Together, these observations suggest that posttranscriptional mechanisms might be involved in regulation or limitation of L1 expression.
By using a polyadq program, we identified 20 putative polyadenylation (polyA) sites located only in the sense strand of the L1.3 genome. We hypothesize that the use of the putative polyA sites located within the L1.3 genome and RNA instability limit the amount of full-length L1.3 mRNAs present in mammalian cells.

To address this question we employed two biologically relevant to LINE-1 genomic structure systems: (i) bicistronic and (ii) a conventional reporter systems that allow detection of both functional polyA sites and potential RNA instability. We determined by Northern blotting analysis that both of these factors are likely to contribute to a very limited amount of full-length LINE-1 mRNAs present in cells.

RNA truncation by premature polyadenylation attenuates human mobile element activity.

Victoria P. Belancio and Dr. Prescott Deininger, Molecular and Cellular Biology, Tulane University, New Orleans, LA, 70130

LINE1 is the most prevalent human retroelement, and it contributes to genomic instability. LINE1 elements comprise 17% of the human genome, which translates into about $5x10^5$ L1 copies, the majority of which are truncated at their 5' end. The full extent of LINE-1 mobility in somatic tissues and particularly in tumors is still not known.

During the course of evolution the activity of LINEs has tremendously decreased. LINE-1 is the only member of the autonomous non-LTR retrotransposones family that is currently active. It is believed that about 60 active copies of full-length LINE-1 are present in the human genome.

L1 expression is extremely low in all cell types of a mature organism except for testis. In contrast, significantly higher levels of LINE-1 expression were found in various cancer cells. The known factors involved in regulation of LINE1 expression, such as hypomethylation often associated with malignant transformation and promoter activity, cannot fully explain the observed pattern of expression.

In addition, full-length LINE-1 is not detected by Northern blotting in cell culture even when transiently expressed from the CMV promoter. Together, these observations suggest that posttranscriptional mechanisms might be involved in regulation or limitation of L1 expression.

By using a polyadq program, we identified 20 putative polyadenylation (polyA) sites located only in the sense strand of the L1.3 genome. We hypothesize that the use of the putative polyA sites located within the L1.3 genome and RNA instability limit the amount of full-length L1.3 mRNAs present in mammalian cells.

To address this question we developed a sensitive northern blot assay that allows detection of the full-length L1 mRNA as well as any RNA species produced through internal polyadenylation. We determined that the A-rich coding strand of the L1 elements contains numerous internal polyadenylation sites that attenuate full-length L1 RNA formation by about 50 fold. There is tremendous redundancy in these internal polyadenylation sites, and their presence is conserved throughout mammalian L1 elements. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome.

RNA truncation by premature polyadenylation attenuates human mobile element activity.

Victoria Perepelitsa-Belancio and Prescott Deininger

Tulane Cancer Center, SL66

and Dept. of Environmental Health Sciences

Tulane University Health Sciences Center

1430 Tulane Ave.

New Orleans, LA 70112

504-988-6385

pdeinin@tulane.edu

ABSTRACT

LINE-1 (L1, Long, INterspersed Element) expression is extremely limited in most mammalian somatic cells, but elevated in germ and cancer cells. Using northern blots and 3'-RACE, we show that the A-rich coding strand of the L1 elements contains numerous internal polyadenylation sites that attenuate full-length L1 RNA formation by about 50 fold. There is tremendous redundancy in these internal polyadenylation sites, and their presence is conserved throughout mammalian L1 elements. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome.

Introduction

LINE-1 elements (L1, Long, INterspersed Elements) are the only active member of the autonomous, non-LTR (long terminal repeat) retrotransposon family in mammals. L1 elements are present at greater than $5x10^5$ copies and represent 17% of the human genome [1]. Most L1 elements are severely truncated at their 5' end, and also accumulate other inactivating alterations [1-3]. Thus, there are only about 100 potentially active, full-length L1 copies per human genome [4]. They result in an estimated one new LINE insertion in every 100 human births, causing 0.1% of human germ-line disease [5,6]. Excessive amplification of L1 elements would have disastrous results for the stability of the human genome.

L1 expression is limited in most differentiated adult cells, but elevated in some cancer cells, testis, and during embryonic development [7-12]. This pattern of expression is

explained to some degree by the alterations in the methylation state of the L1 promoter [13,14] that occur upon malignant transformation and by regulation through the SRY family of transcription factors during embryogenesis [15]. However, even when expressed under the very strong CMV promoter in transient assays, very little full-length RNA is produced. Thus, it seems likely that L1 elements have inherent properties limiting their RNA levels, a trait that potentially evolved to limit damage to their host genome.

In order to minimize the negative impact on their host genomes, many mobile elements have evolved strategies to limit their amplification potential, particularly in somatic cells [16]. This includes differential transcription or splicing between germline and somatic cells [17], or poor codon usage [18]. We identified a number of putative polyadenylation signals within the coding region of the human L1 elements and hypothesized that these internal polyadenylation signals may represent a unique mechanism to cause truncation by premature polyadenylation of L1 transcripts. The classic AAUAAA sequence associated with a polyadenylation signal is generally required, but not sufficient to direct strong polyadenylation. This hexamer along with the downstream GU-, or U-, rich elements determines the appropriate cleavage sequence for efficient polyadenylation [19,20]. In this paper, we demonstrate that a number of the L1 internal polyadenylation sites are functional and result in a strong attenuation of expression from the L1 genome, as well as having implications for damaging expression of genes in which they insert.

Results

Mammalian LINE-1 elements contain putative polyadenylation signals.

The L1 genome has a large number of potential poly(A) signals, AATAAA and ATTAAA, in its sense strand (Fig. 1a), yet very few in the antisense direction. This is the opposite pattern seen in most cellular genes, as selection normally limits the number of potential poly(A) sites that could cause premature termination of transcription. The program POLYADQ [21] predicts that several of these sites are much stronger than the relatively weak poly(A) site normally found at the 3' end of the L1 element [22], and one of the internal sites was predicted to be stronger than the SV40 poly(A) site. Furthermore, the positions of these poly(A) sites were well conserved between active human and full-length gorilla elements, with similar numbers and approximate locations in active mouse elements (Fig. 1a). Additionally, there are multiple non-canonical poly(A) sites distributed throughout the L1.3 coding region (Fig. 1b) that may also be functional. Thus, the prevalence of poly(A) sites between different species suggests that these sites may perform a conserved function and limit the amount of full-length L1 RNA produced.

The strongest predicted L1.3 poly(A) site is functional.

We transiently expressed the active human L1 element, L1.3, that has been shown to undergo efficient retrotransposition [23], in mouse NIH 3T3 cells. This vector includes a CMV promoter for maximal transcription, an intron, and a selectable marker in the 3' non-coding region. Expression in mouse cells allowed us to use probes from the 5'end of the human L1 element to detect RNAs produced from this transfected L1 element without background from endogenous L1 RNAs [24]. We used the 5'-UTR strand-specific probe

(Fig. 2a), because it would permit identification of any RNAs truncated by polyadenylation at the internal poly(A) sites (Fig. 2b). This assay detected two high molecular weight, L1.3-specific RNAs enriched in the poly(A)-selected fraction that migrated between 7.4 and 9.4 kb (FL1.3, Fig. 2b). A doublet with the identical molecular weights was detected in the poly(A)-selected fraction with a strand-specific RNA probe to the neomycin selectable marker located at the 3'end of the L1.3 construct (Fig. 2a, and data not shown). The presence of the doublet is consistent with the inefficient splicing of this intron [25]. Multiple faster-migrating species were also present and much more abundant. The sizes of these RNAs roughly corresponded to the positions expected with usage of the putative poly(A) sites identified in the L1.3 sequence (bands #1 through #3, Fig. 2b). The bands were absent from the poly(A)-minus fraction, and they were not detected by the neomycin strand-specific probe located downstream of these termination sites (data not shown). The strongest band, # 3 (Fig. 2b), corresponds in size to an RNA that would be expected to terminate in the vicinity of the strongest predicted poly(A) site. We also observed abundant truncated bands using a northern blot with a transfected mouse L1 element, L1spa (data not shown), demonstrating the conservation of this premature polyadenylation in mammalian L1 elements.

Two inactivating point mutations were introduced in the most conserved AATAAA element of the strongest predicted poly(A) site [26]. Band #3 from the wild-type (wt) L1 element completely disappeared (Fig. 2b) in mRNAs produced from this mutant, indicating that this poly(A) site was responsible for the formation of this RNA species. This confirmed that at least one of the internal poly(A) sites is functional, and contributes to limiting the full-length L1.3 expression. Removal of the strongest poly(A) site

resulted in the more efficient utilization of the poly(A) sites located 5' and 3'relative to its position (bands #4 and 5, Fig. 2b). The amount of the full-length L1.3 mRNA (FL1.3 combined, Fig. 2b) remained similar in the wt and mutant at an average of 2% of the total L1.3-related transcripts. It has been previously observed that the relative activity of any given poly(A) site varies depending on competition with other nearby poly(A) sites. Thus, internal poly(A) sites present in the L1.3 genome may individually have a modest effect on limiting the L1.3 expression, but the sum of them results in a 50-fold decrease in the amount of the full-length mRNA that is produced in the NIH 3T3 cells.

We isolated total polyadenylated RNAs from human Ntera2 cells and HeLa cells to determine whether endogenous L1 RNAs showed a similar series of transcript lengths. Ntera2 cells have been demonstrated to express relatively high levels of full-length L1 RNA, while HeLa cells produce very little L1 RNA [11](Fig. 2b). The full-length RNA is of a different length because the endogenous L1 does not include the extra selectable marker region present in our transfected element. However, there is a similarity in the endogenous lower molecular weight bands detected in Ntera2 cells to those from expression of the L1.3 vector in NIH 3T3 cells. The differences are likely to represent variation in the poly(A) sites present within the full-length L1 elements being expressed from various endogenous loci. We also compared the band distribution in transiently transfected and non-transfected HeLa cells to those in the mouse cells (Fig. 2b). A similar pattern was observed in both cell lines with the exception of band #6 that was detected in HeLa but not in NIH 3T3 cells, suggesting possible species or cell type-specific differences in poly(A)-site utilization.

Both canonical and non-canonical internal poly(A) sites are utilized.

To further confirm the use of specific internal L1 poly(A) sites, we performed 3'RACE on poly(A)-selected mRNAs from L1.3 transfected cells. A chicken fibroblast cell line was used for this assay to prevent background from endogenous, mammalian, L1 RNAs. We first demonstrated that expression of the human L1 vector produced similar RNA species in chicken cells with only modest quantitative differences, using a northern blot (Fig. 2b). This suggests that, as with mouse cells, there may be species-, or tissue-specific differences in the relative utilization of the poly(A) sites. Employing the 3' RACE to the region expected to produce the major truncated L1 RNA products in the northern blots showed a pattern consistent with the northern blot results (Fig. 3). The band corresponding to the strongest poly(A) site in the wt disappeared upon mutagenesis. Sequence analysis of several of the bands revealed the usage of the canonical AATAAA, the next most commonly utilized variant ATTAAA, as well as other non-canonical, internal poly(A) sites with one base variation from the AATAAA (Fig. 3). All polyadenylations occurred at sites consistent with distance and sequence normally associated with 3' cleavage [19,20,27]. Further confirmation of the use of these internal poly(A) sites in the endogenous L1 elements comes from finding a number of transcripts in the EST database whose 3' end would correspond to the positions used by both the classic and non-canonical, internal poly(A) sites (data not shown).

Discussion

The finding that internal poly(A) sites have such a strong influence on the full-length L1 RNA levels has tremendous implications for the retrotransposition process in mammals and its impact on the stability of the genome. This process almost certainly limits

retroposition activity, resulting in a more acceptable level of damage produced by these elements both through their own retrotransposition, as well as through their potential to stimulate SINE and processed pseudogene retroposition [28-30]. It also limits general expression of the ORF products, potentially minimizing negative consequences of over-expressing domains, such as the endonuclease and reverse transcriptase, which may be detrimental to their host. Because most of the internal poly(A) sites are downstream of ORF1, it is also possible that these termination events are at least partly responsible for the very high levels of ORF1 protein found in cells relative to the extremely poorly expressed ORF2 product [31].

The abundance of internal poly(A) sites is likely to be directly related to the very high proportion of A residues on the coding strand of the ORFs (>40%). Although poly(A) sites may periodically be eliminated by random mutagenesis, the A-richness makes it likely that new poly(A) sites will be generated regularly in these elements. Because of the redundancy in the system, point mutations that destroy one or two poly(A) sites may have only a minimal effect on overall amplification capability. This includes not only a large number of classical poly(A) sites, but also many more non-canonical single-base substitution poly(A) sites in the sense strand of the L1.3 (Fig. 1). These non-canonical poly(A) sites function at reduced efficiency relative to the AATAAA poly(A) site *in vitro* [26]. Finding that some of the non-canonical poly(A) sites are also active in L1 (Fig. 3), suggests that the L1 sequence supports the use of a broad range of poly(A) sites.

The question arises as to whether the A-richness of L1 is caused by the need for the poly(A) sites, or whether the poly(A) sites are merely a by-product of this A-richness. It seems likely that some other process, such as preferential binding to L1 proteins or

limiting RNA secondary structures, plays a role in evolutionary selection for the A-richness. This A-richness appears to also contribute to lower L1 ORF expression by causing non-optimal codon utilization (P. Deininger, unpublished). Non-optimal codon utilization has previously been shown to limit expression in other transposon species [18]. Thus, the A-richness of the L1 RNA contributes to both truncations of the majority of its RNAs, as well as inhibiting the effectiveness of translation from the few full-length RNA molecules produced.

Quantitative variation in poly(A) site usage between NIH3T3, HeLa and chicken cells demonstrates that the sites may be differentially used. Alternative polyadenylation has been found to be involved in producing variant mRNAs in a tissue-, or development-specific manner [19,20], with testis being a common site of alternative polyadenylation [32]. Such regulation usually involves multiple weak poly(A) sites, with the strongest poly(A) site being located at the most 3' end. However, poly(A) sites utilized in this way are typically positioned in the introns or 3' UTRs of the regulated genes [32], rather than in their coding regions (Fig. 1a). The presence of so many internal poly(A) sites in L1 and their variable usage in different cell types suggest the possibility that this differential polyadenylation may be regulatory. It may not just attenuate expression, but could also be a strategy to limit its activity in somatic cells, while allowing more expression in cell types in which its retrotransposition could be passed on to progeny.

The few L1 elements found in genes are almost always in the reverse orientation, suggesting that insertion of an L1 element into a gene, particularly in the forward orientation, may have negative consequences. Because most L1 insertions are truncated [1], they do not usually carry any signals from their promoter, or 5' end. Therefore, the

most likely negative impacts of L1 elements would arise from the combined activity of the relatively weak poly(A) site at their 3' UTR [22] and the stronger internal poly(A) sites. Insertion of these elements in the forward orientation would be likely to cause at least a partial premature transcription termination in any gene in which they inserted. There are only a few potential poly(A) sites in the L1 antisense strand so that L1 insertions in the opposite orientation may be less damaging to the targeted gene.

L1 elements, and the genomes in which they reside, have evolved a highly sophisticated, and redundant, series of mechanisms to insure the scarceness of full-length L1 mRNA in any given somatic cell. Some suppression has been shown at the level of transcription through limitations in transcription factors [15] and by methylation [14]. Our observation of a novel application of the conventional mRNA 3'end formation extends the range of the known mechanisms for the control of mRNA levels in mammalian cells, and provides additional understanding of the selective pressures that act upon new retrotransposition events in reshaping the mammalian genome.

Methods

Cell culture. NIH 3T3 (ATCC CRL-1658), HeLa (ATCC CCL2), and Ntera2 (ATCC CRL-1973) cells were maintained at 37^0C and 5% CO_2 in DMEM (GIBCO) high glucose, 10% Colorado calf serum (CS, GIBCO); EMEM (GIBCO), 10% fetal bovine serum (FBS, GIBCO); and DMEM high glucose (GIBCO), 15% FBS, respectively.

Chicken fibroblasts (ATCC CRL-12203) were maintained at 39^{0}C and 5% CO_2 in DMEM high glucose (GIBCO) and 15% FBS.

Transient-transfection assay. 3.5 to 4.5x10^6 cells per 75 cm^2 cell culture flask (Corning Inc.) were transected with 3-6 μg of the L1.3 expression cassette by lipofectamine (6-12 μl of Plus reagent, 18-36 μl of lipofectamine, Invitrogen) 16 to 18 hr after plating. NIH 3T3 cells were incubated with the transfection cocktail in the serum-free media for 4 hr, chicken fibroblasts, HeLa, and Ntera2 cells were incubated with the transfection cocktail in the serum-free media for 3 hr. All cell types were incubated in their respective media for 24 to 26 hr. prior to harvesting RNA.

RNA extraction and poly(A) selection. Four 75 cm^2 cell culture flasks of each cell type were combined to extract total mRNA using TRIzol Reagent (Invitrogen). This was followed by chloroform extraction and isopropanol precipitation. The PolyATract mRNA isolation system III (Promega) was used to select polyadenylated RNA species as instructed by the manufacturer. Poly(A)-selected and precipitated RNA was resuspended in 30μl of RNAse-free water and fractionated in a single lane of an agarose-formaldehyde gel.

Northern blot assay. RNA was transferred to a Hybond-N nylon membrane (Amersham Pharmacia Biotech) by capillary transfer overnight at room temperature in a standard 5x sodium chloride/sodium citrate (SSC) solution. The RNA was cross-linked to the membrane with UV-light and pre-hybridized in 30% formamide, 1X Denhardt's solution, 1% SDS, 1 M NaCl, 100 μg/ml salmon sperm DNA, 100 μg/ml yeast t-RNA at 60 ^{0}C for at least 6 hr. Hybridization with a strand-specific probe (final concentration of 4-6x10^6 cpm/ml) was carried out overnight in the same solution at 60 ^{0}C. Multiple ten-minute

washes were performed at high stringency (0.1xSSC, 0.1% SDS) at 60 ^{0}C. The results of the northern blot assays were quantitated on a Fuji Phosphorimager.

The strand–specific probe used for the northern-blot assay was generated by MAXIscript T7 system (Ambion). DNA template for the probe was produced by PCR with the primers that amplified either L1.3 5'UTR (forward 5'-CAAGCTTGCGCGGC CGCGG, reverse 5"-TAATACGACTCACTATAGGCTTTGTGGTTTTA) or second exon of the neo^{R} cassette (forward 5'-GAAGAACTCGTCAAGAAGG , reverse 5'-TAATACGACTCACTA TAGGAGGACGAGGCAG). The T7 promoter sequence was included in the reverse primer of each pair. PCR products were fractionated on a 1% low-melting agarose gel, excised, and purified using a QIAquick gel extraction kit 50 (QIAGEN).

3' RACE analysis. 3' RACE (Clontech) was performed according to the manufacturer's protocol on poly(A)-selected total mRNA from chicken fibroblasts transiently transfected with L1.3 expression cassette as described above.

Site-directed mutagenesis. The QuikChange Site-Directed Mutagenesis kit (STRATAGENE) was used to change the AATAAA hexamer sequence of the strongest predicted poly(A) site in the L1.3 genome to GATCAA.

Accession Numbers. L1.3 - L19088

Acknowledgements: This work was supported by National Institutes of Health RO1 GM45668 (PLD) and DOD Breast Cancer Research Program, DAMD17-02-1-0597 (VP-B). We would like to thank Carl Schmid and Mark Batzer for helpful comments, John Moran, Haig Kazazian, Jr. and John Goodier for providing L1 vectors.

Literature Cited

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature* **409**, 860-921 (2001).

2. Weiner, A., Deininger, P., & Efstradiatis, A. The Reverse Flow of Genetic Information: pseudogenes and transposable elements derived from nonviral cellular RNA. *Annual Reviews of Biochemistry* **55**, 631-661 (1986).

3. Singer, M. F., Krek, V., McMillan, J. P., Swergold, G. D., & Thayer, R. E. LINE-1: a human transposable element. *Gene* **135**, 183-188 (1993).

4. Sassaman, D. M. *et al.* Many human L1 elements are capable of retrotransposition. *Nat.Genet.* **16**, 37-43 (1997).

5. Kazazian, H. H. J. Mobile elements and disease. *Curr.Opin.Genet.Dev.* **8**, 343-350 (1998).

6. Kazazian, H. H. J. & Moran, J. V. The impact of L1 retrotransposons on the human genome. *Nat.Genet.* **19**, 19-24 (1998).

7. Branciforte, D. & Martin, S. L. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol.* **14**, 2584-2592 (1994).

8. Ostertag, E. M. *et al.* A mouse model of human L1 retrotransposition. *Nat.Genet.* **32**, 655-660 (2002).

9. Bratthauer, G. L., Cardiff, R. D., & Fanning, T. G. Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* **73**, 2333-2336 (1994).

10. Asch, H. L. *et al.* Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol.Res.* **8**, 239-247 (1996).

11. Skowronski, J., Fanning, T. G., & Singer, M. F. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol.Cell Biol.* **8**, 1385-1397 (1988).

12. Trelogan, S. A. & Martin, S. L. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc.Natl.Acad.Sci.U.S.A* **92**, 1520-1524 (1995).

13. Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P., & Warren, W. D. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J.Biol.Chem.* **272**, 7810-7816 (1997).

14. Thayer, R. E., Singer, M. F., & Fanning, T. G. Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene* **133**, 273-277 (1993).

15. Tchenio, T., Casella, J. F., & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28**, 411-415 (2000).

16. Deininger, P. L. & Batzer, M. A. Mammalian retroelements. *Genome Res* **12**, 1455-1465 (2002).

17. Siebel, C. W. & Rio, D. C. Regulated splicing of the Drosophila P transposable element third intron in vitro: somatic repression. *Science* **248**, 1200-1208 (1990).

18. Lerat, E., Biemont, C., & Capy, P. Codon usage and the origin of P elements. *Mol.Biol.Evol.* **17**, 467-468 (2000).

19. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755-2766 (1997).

20. Zhao, J., Hyman, L., & Moore, C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol.Mol.Biol.Rev.* **63**, 405-445 (1999).

21. Tabaska, J. E. & Zhang, M. Q. Detection of polyadenylation signals in human DNA sequences. *Gene* **231**, 77-86 (1999).

22. Moran, J. V., DeBerardinis, R. J., & Kazazian, H. H., Jr. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530-1534 (1999).

23. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917-927 (1996).

24. Loeb, D. D. *et al.* The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol Cell Biol* **6**, 168-182 (1986).

25. Gilbert, N., Lutz-Prigge, S., & Moran, J. V. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**, 315-325 (2002).

26. Sheets, M. D., Ogg, S. C., & Wickens, M. P. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* **18**, 5799-5805 (1990).

27. Proudfoot, N. Poly(A) signals. *Cell* **64**, 671-674 (1991).

28. Deininger, P. L. & Batzer, M. A. Alu repeats and human disease. *Mol Genet Metab* **67**, 183-193 (1999).

29. Boeke, J. D. LINEs and Alus--the polyA connection. *Nat.Genet.* **16**, 6-7 (1997).

30. Esnault, C., Maestre, J., & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat.Genet.* **24**, 363-367 (2000).

31. Kolosha, V. O. & Martin, S. L. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc.Natl.Acad.Sci.U.S.A* **94**, 10155-10160 (1997).

32. Edwalds-Gilbert, G., Veraldi, K. L., & Milcarek, C. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* **25**, 2547-2561 (1997).

Figure 1. Putative poly(A) sites in the human, gorilla, and mouse LINE-1 elements. **(A)** Schematic representation of the putative poly(A) sites identified in the human L1.3, gorilla L1Gg-1A, and mouse L1spa elements. The approximate positions of the coding regions of these elements are shown. The genomes were aligned according to the beginning of the open reading frame1 (ORF1). **(B)** The positions of some of the more common single-base variants of the canonical polyadenylation signal, AAUAAA, in the human L1.3 genome are shown. An asterisk marks the strongest predicted poly(A) site.

Figure 2. Utilization of the internal poly(A) sites of L1.3. **(A)** Schematic representation of the L1.3 expression cassette [23]. The approximate positions of the endogenous L1 promoter, coding regions (ORF1 and ORF2), 3' UTR and polyadenylation site (L1.3pA), intron (IN) and inverted neomycin resistance gene are shown. The SV40 polyadenylation site is located immediately downstream of the L1 cassette. Positions of the putative internal poly(A) sites are marked by black (for AAUAAA) and shorter gray (for AUUAAA) vertical lines. The asterisk marks the strongest predicted polyA site. Horizontal lines represent the predicted mRNA species with the thickness of the line suggesting that different sites are likely to truncate higher proportions of the RNAs. The bottom dotted line is consistent with read-through transcripts [22]. Thick black horizontal lines labeled 5'UTR and neo probe reflect genomic positions of the strand-specific probes used for northern blot analysis. **(B)** Northern blot of L1.3 RNA species produced by the wild type (wt) element and one with the mutation in the strongest predicted polyA site (mut) in NIH 3T3, Ntera2, HeLa, and chicken cells probed with the strand-specific 5'UTR probe. FL1.3 is full-length L1.3 mRNA, enFL1.3 is the L1.3 full-length mRNA

from endogenous elements. Long horizontal arrows correspond to the positions of the molecular weight RNA marker (Invitrogen). RNA bands that are described specifically in the text are numbered.

Figure 3. 3' RACE analysis of the prematurely terminated L1.3 RNA species. 3'RACE was performed on poly(A)-selected total mRNAs from chicken fibroblasts transiently transfected with the wt or mut L1.3 expression vector described in figure 2. The upstream primer used in the PCR step corresponds to positions 1342-1359 of the L1.3 (accession # L19088) sense strand. Lanes represent: 1 kb DNA ladder (NEBiolabs, lane 1); 3'RACE reaction of the wt (lane 2); and mut (lane 3) L1.3 without reverse transcriptase (RT); 3'RACE reaction of the wt (lane 4) and mut (lane 5) L1.3 with the L1.3-specific primer. The band corresponding to the strongest predicted polyA site identified by the Northern blot is labeled 3. Genomic location and the hexanucleotide sequence of the identified functional poly(A) sites just upstream of individual bands, following cloning and sequencing, are shown next to their positions on the gel.

Fig. 1A

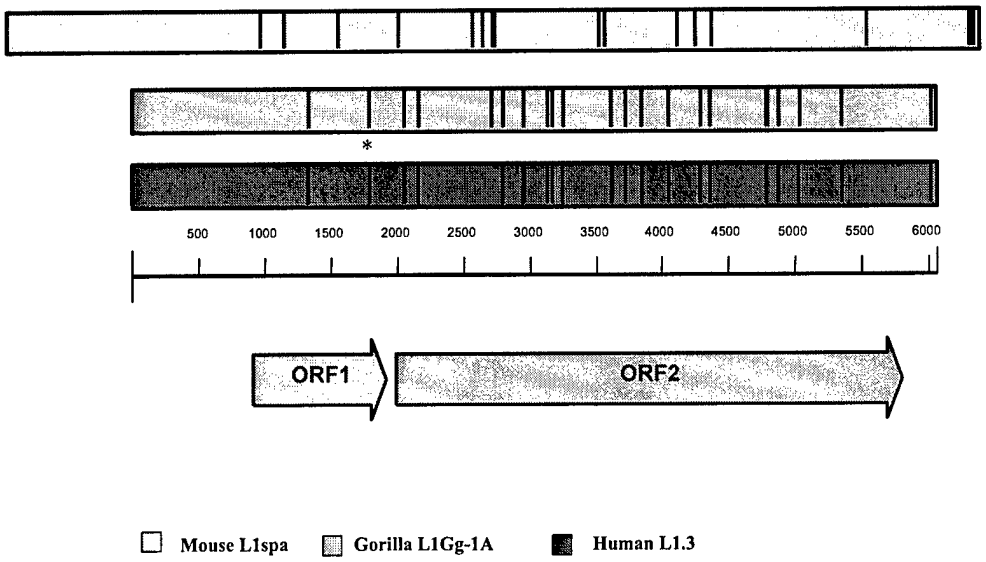

500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
ORF1
ORF2
Mouse L1spa
Gorilla L1Gg-1A
Human L1.3

Fig. 1B

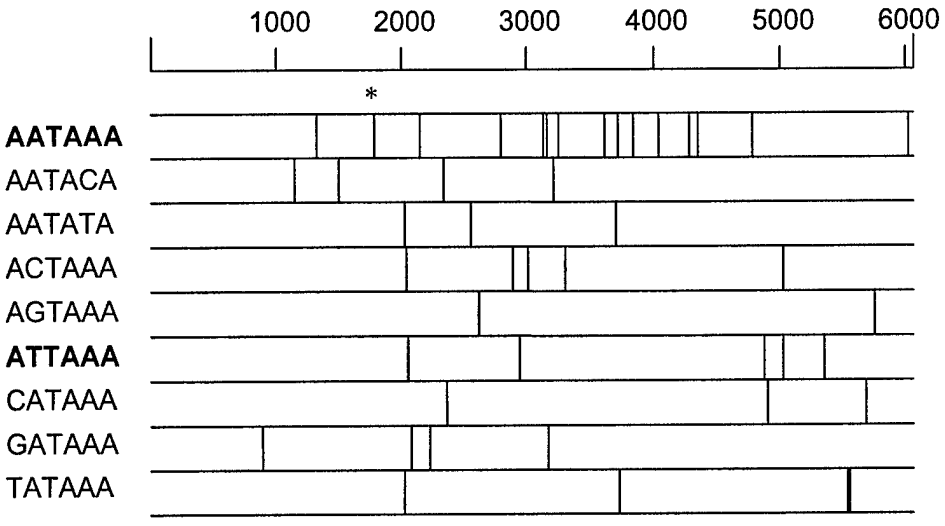

1000
2000
3000
4000
5000
6000
AATAAA
AATACA
AATATA
ACTAAA
AGTAAA
ATTAAA
CATAAA
GATAAA
TATAAA

Fig.2A

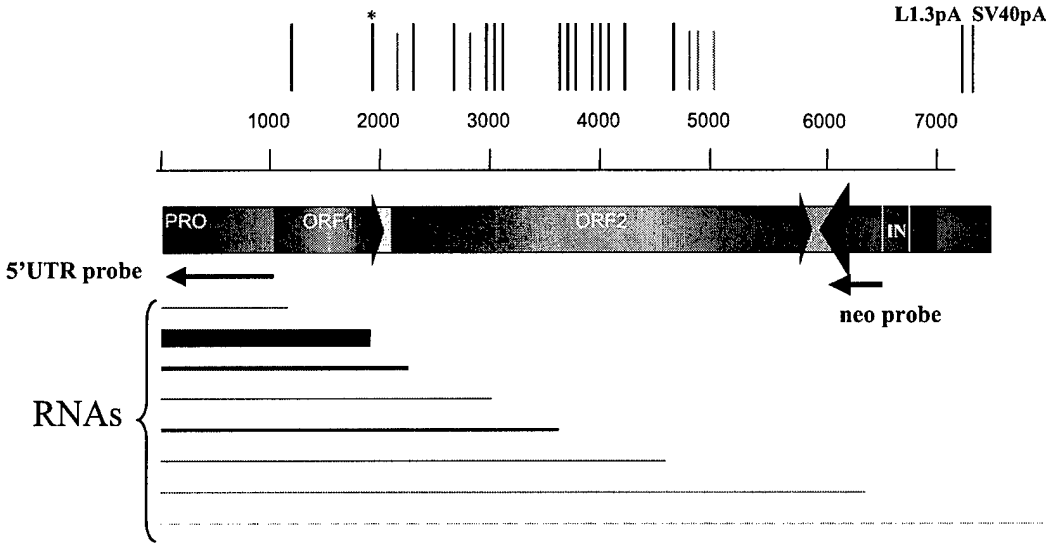

L1.3pA
SV40pA
1000
2000
3000
4000
5000
6000
7000
PRO
ORF1
ORF2
IN
5'UTR probe
neo probe
RNAs

Fig.2B

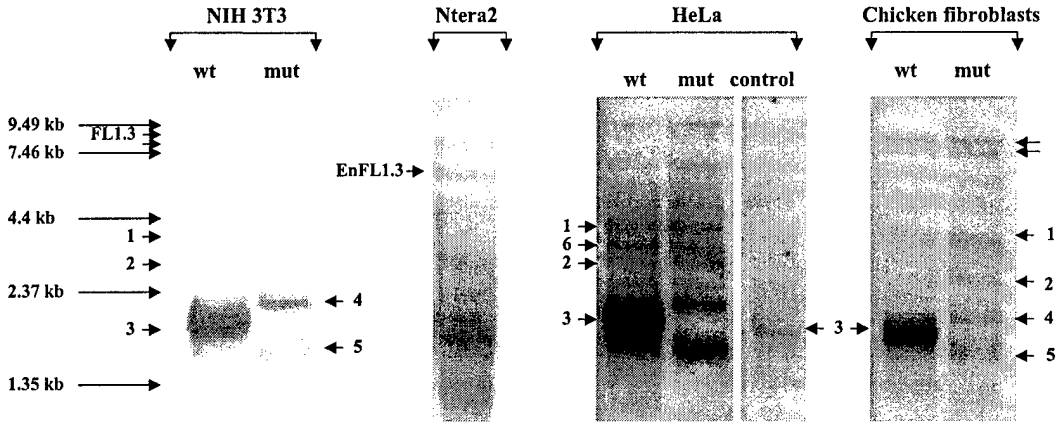
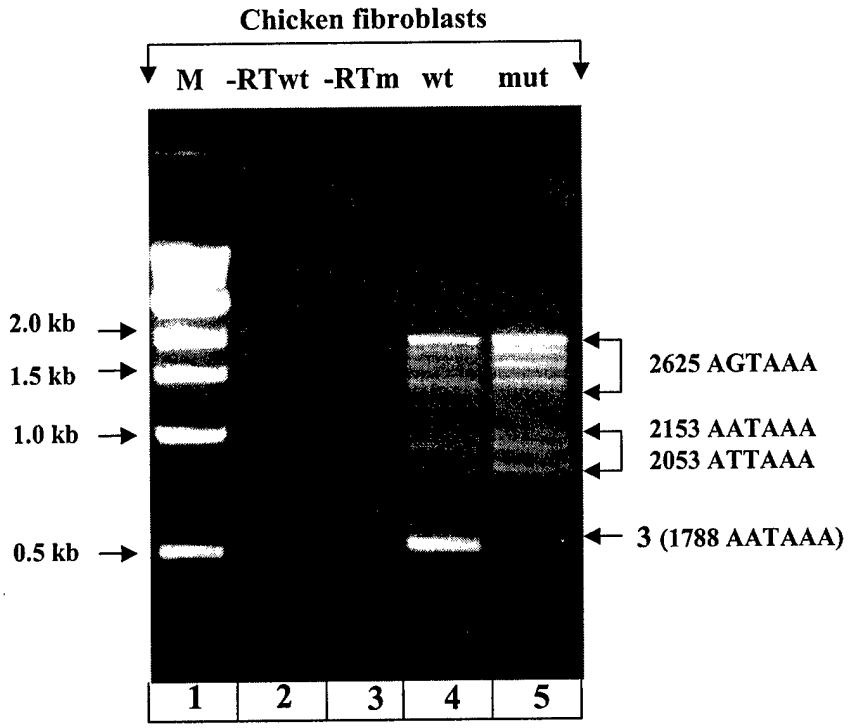

NIH 3T3
Ntera2
HeLa
Chicken fibroblasts
wt
mut
wt
mut
control
wt
mut
9.49 kb
7.46 kb
FL1.3
EnFL1.3
4.4 kb
1
2
2.37 kb
3
1.35 kb
4
5
1
6
2
3
3
1
2
4
5

Fig.3


Chicken fibroblasts
M
-RTwt
-RTm
wt
mut
2.0 kb
1.5 kb
1.0 kb
0.5 kb
2625 AGTAAA
2153 AATAAA
2053 ATTAAA
3 (1788 AATAAA)
1
2
3
4
5