The EMBO Journal Vol.17 No.2 pp.590–597, 1998

An actively retrotransposing, novel subfamily of mouse L1 elements

Thierry P.Naas[1,2], Ralph J.DeBerardinis[1], John V.Moran[1], Eric M.Ostertag[1], Stephen F.Kingsmore[3], Michael F.Seldin[4], Yoshihide Hayashizaki[5], Sandra L.Martin[6] and Haig H.Kazazian,Jr[1,7]

[1]Department of Genetics, University of Pennsylvania, School of Medicine, Philadelphia, PA 19104, [3]Department of Medicine, University of Florida, Gainesville, FL 32610, [4]Department of Biological Chemistry, University of California Davis, Sacramento, CA 95616, [5]RIKEN Tsukuba Life Science Center, 3-1-1, Tsukuba, Ibaraki 305, Japan, [6]Department of Cellular and Structural Biology, University of Colorado School of Medicine, Denver, CO 80262, USA

[2]Present address: Hôpital Antoine Béclère, 157 rue de la Porte de Trivaux, 92141 Clamart Cedex, France

[7]Corresponding author

T.P.Naas and R.J.DeBerardinis contributed equally to this work

Retrotransposition of LINEs and other retroelements increases repetition in mammalian genomes and can cause deleterious mutations. Recent insertions of two full-length L1s, $L1_{spa}$ and $L1_{Orl}$, caused the disease phenotypes of the spastic and Orleans reeler mice respectively. Here we show that these two recently retrotransposed L1s are nearly identical in sequence, have two open reading frames and belong to a novel subfamily related to the ancient F subfamily. We have named this new subfamily T_F (for transposable) and show that many full-length members of this family are present in the mouse genome. The T_F 5′ untranslated region has promoter activity, and T_F-type RNA is abundant in cytoplasmic ribonucleoprotein particles, which are likely intermediates in retrotransposition. Both $L1_{spa}$ and $L1_{Orl}$ have reverse transcriptase activity in a yeast-based assay and retrotranspose at high frequency in cultured cells. Together, our data indicate that the T_F subfamily of L1s contains a major class of mobile elements that is expanding in the mouse genome.

Keywords: L1/LINEs/pseudogene/retrotransposition/ reverse transcriptase

Introduction

LINEs, or L1s, are repeated sequences that pervade mammalian genomes (Burton *et al*., 1986). Although most L1s are inactive because they are truncated, rearranged or mutated, some are active and disperse by retrotransposition (i.e. transposition requiring reverse transcription and re-integration of the resulting cDNA). Full-length mouse and human LINEs contain shared features that include two protein-coding regions [open reading frames (ORFs) 1 and 2], a 3′ untranslated region (UTR) and a 3′ poly(A) tail. ORF1 encodes an RNA-binding protein (Martin, 1991; Hohjoh and Singer, 1996, 1997; Kolosha and Martin, 1997) required for retrotransposition (Moran *et al.*, 1996). ORF2 encodes an endonuclease (Feng *et al.*, 1996), a reverse transcriptase (RT; Mathias *et al.*, 1991) and a highly conserved cysteine-rich motif (Fanning and Singer, 1987), all of which are also required for retrotransposition (Moran *et al.*, 1996).

Mouse L1s, unlike human L1s, have 5′ UTRs with tandemly repeated units called monomers at their 5′-most end. Earlier studies revealed two types of monomers, A and F, which are about 200 bp long and unrelated in sequence (Loeb *et al.*, 1986; Padgett *et al.*, 1988). Although A- and F-type L1s are equally abundant in the genome, A-subfamily members display little divergence, often have intact ORFs and are transcribed (Schichman *et al.*, 1992, 1993; Severynse *et al.*, 1992). In contrast, F-subfamily members are divergent, lack intact ORFs and are not transcribed (Schichman *et al.*, 1992). Thus, while some A elements may be active, F elements appear ‘dead’ in that they dispersed millions of years ago and have accumulated deleterious mutations (Adey *et al.*, 1994).

New retrotranspositions of active human elements have been identified through systematic screening of disease genes (Kazazian *et al.*, 1988; Morse *et al.*, 1988; Miki *et al.*, 1992; Narita *et al.*, 1993; Holmes *et al.*, 1994). Although those insertions were truncated progeny of active L1s, isolation of their precursors (Dombroski *et al.*, 1991; Holmes *et al.*, 1994) led to demonstration of autonomous retrotransposition in cultured cells (Moran *et al.*, 1996).

Recently, four L1 insertions causing disease have been identified in mice (Kingsmore *et al*., 1994; Mulhardt *et al*., 1994; Kohrmann *et al.*, 1996; Takahara *et al*., 1996; Perou *et al.*, 1997). Surprisingly, two of these insertions ($L1_{spa}$ and $L1_{Orl}$) were full-length, and thus may retain the ability to retrotranspose (Kingsmore *et al*., 1994; Mulhardt *et al*., 1994; Takahara *et al*., 1996). A partial sequence of $L1_{spa}$ (Kingsmore *et al.*, 1994) and complete sequence of $L1_{Orl}$ (Takahara *et al.*, 1996) suggested that both were members of the inactive F subfamily. Here we show that $L1_{spa}$ and $L1_{Orl}$ are active autonomous retrotransposons that belong to a novel L1 subfamily related to the F subfamily. We name this new subfamily the T_F subfamily.

Results

$L1_{spa}$ and $L1_{Orl}$ are members of a novel L1 subfamily with many full-length members

We sequenced $L1_{spa}$ and showed that it is 7.5 kb long, has seven and a half monomers in its 5′ UTR and retains both ORFs with intact endonuclease and RT domains (Figure 1). Comparison of the $L1_{spa}$ sequence (DDBJ/EMBL/GenBank accession No. AF016099) with the reported sequence of $L1_{Orl}$ showed that they are closely related with an overall identity of 99.4% (Figure 1). Both

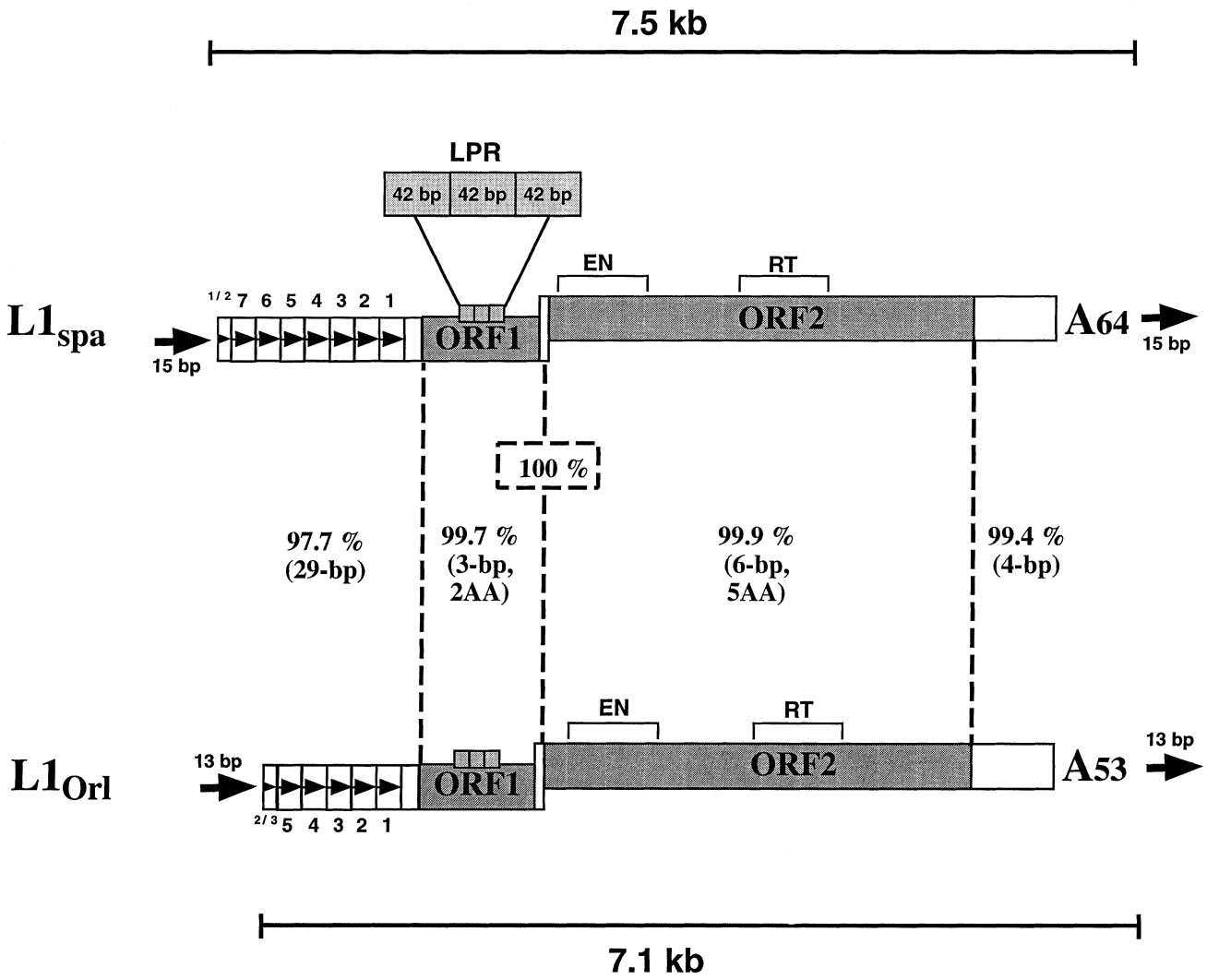


Fig. 1. $L1_{spa}$ and $L1_{Orl}$ are closely related. Each element contains a large 5′ UTR composed of monomers (boxed arrows), intact ORFs of 1113 bp and 3843 bp (shaded boxes) and a 667 bp 3′ UTR (open box) terminating in a stretch of adenines. $L1_{spa}$ contains seven and a half monomers, while $L1_{Orl}$ contains five and two-thirds. The length polymorphism region (LPR) of three 42 bp repeats within ORF1 is shown in both elements. The inter-ORF region consists of 40 bp. Functional domains of ORF2 are indicated (EN, endonuclease; RT, reverse transcriptase). Both elements are flanked by target-site duplications (large arrows). Percent sequence identity and number of nucleotide and amino acid substitutions between $L1_{spa}$ and $L1_{Orl}$ are presented for each region. The overall divergence over 7.1 kb is 42 bp (0.6%).

elements have three copies of a 42 bp repeat in the length polymorphism region (LPR) of ORF1. Reported expressed A-type L1s contain two copies of this sequence while genomic F-type elements have a single 42 bp block and a related 66 bp block; three 42 bp repeats have been observed in only a few genomic A-type elements (Schichman *et al.*, 1992).

The $L1_{spa}$ and $L1_{Orl}$ 5′ UTRs are 97.7% identical. Consensus sequences made by aligning the seven and a half $L1_{spa}$ monomers or the five and two-thirds $L1_{Orl}$ monomers showed 100% identity to each other, but only 73% identity to an F-monomer consensus (Figure 2a). Furthermore, $L1_{spa}$ and $L1_{Orl}$ are identical over the 3′-most 125 bp of monomer 1, which are divergent even among L1s from the same subfamily. Therefore, the two elements are distinct but closely related members of the previously uncharacterized T_F subfamily.

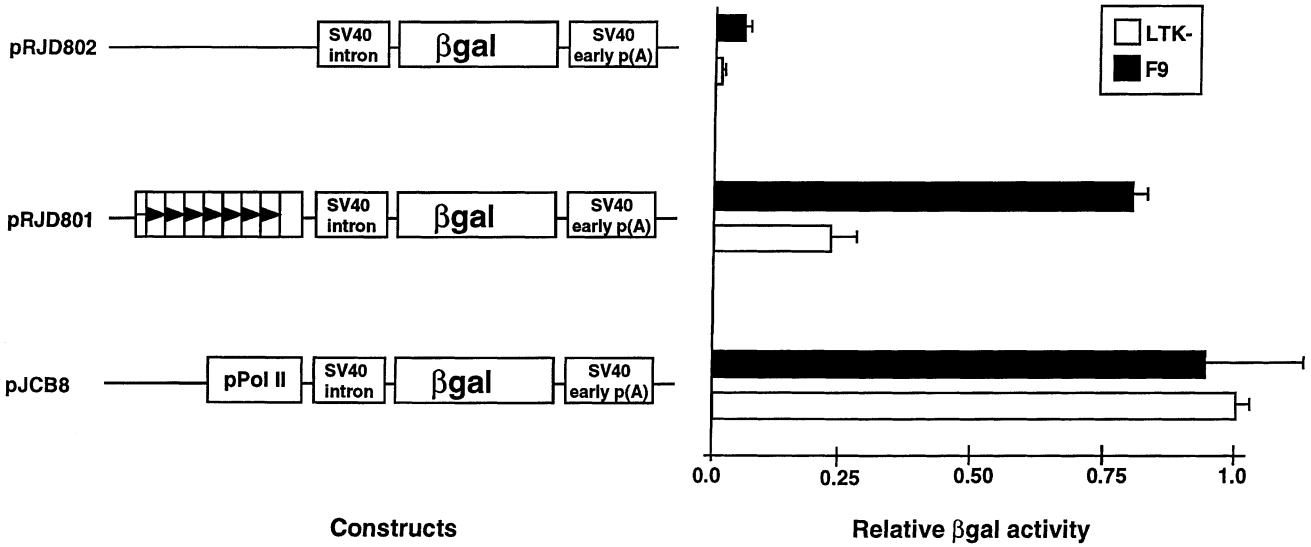
The human and mouse genomes each contain approximately 100 000 L1s, but the vast majority are 5′ truncated (Voliva *et al.*, 1983; Fanning and Singer, 1987; Hutchison *et al.*, 1989). As a result, the copy number of full-length L1s is much lower than that of 3′ UTR sequences. To estimate the number of full-length T_F L1s in the mouse genome, we screened two mouse genomic libraries with a probe containing T_F monomer sequence. We determined that the diploid mouse genome contains 2000–3000 copies of the T_F 5′ UTR.

To determine the percentage of T_F 5′ UTRs associated with downstream L1 sequence, we isolated 30 phage cores that hybridized with the T_F probe. Twenty-nine of these cores (97%) contained a phage that also hybridized to an ORF2 probe. We purified these 29 phage and PCR-amplified a 3.6 kb fragment encompassing most of ORFs 1 and 2 from 18 of them (62%). Thus, our data suggest that almost all of the 2000–3000 T_F sequences are associated with downstream L1 sequence, and that at least 60% are contained in

A

```
                .         .         .         .         .
F-recons.    1 GACATCCGGGCACCTTCCCTGCCAGAGGAGAGGTGTCTGCCCNGCCCGGG
               |||| |||| ||||||||   ||||||||  |||||  | ||||:||  |||
TF-cons.     1 GACAGCCGGCCACCTTCCGGACCAGAGGACAGGTGCCCGCCCGGCTGGGG
                .         .         .         .         .
F-recons.   51 A...GGGCTTTGCCTGAGCATCTGCGGGAGCCATCTTGGT......TCCC
               |   || ||  |||  |||| | ||||  ||||||||||        ||
TF-cons.    51 AGGCGGCCTAAGCCACAGCAGCAGCGGTCGCCATCTTGGTCCGGGACCCG
                .         .         .         .         .
F-recons.   92 GGATNCCACCGAGACTAGTCTGCACAGGTGAGAGTGTGGACTACAGAAGC
                :|    || | |||||||| |||||||||| |||  | | |||| |
TF-cons.   101 CCGAACTTAGGAAATTAGTCTGAACAGGTGAGAGGGTGCGCCAGAGAACC
                .         .         .         .         .
F-recons.  142 TACACAGCTTCTGGGACAGGCNGAAGCCACAGAGCTTCTGAGGCAGANCC
               |  |||||||||||  |||||||:|||| |||||||   |||||||||  :||
TF-cons.   151 T.GACAGCTTCTGGAACAGGCGGAAG.CACAGAGGCGCTGAGGCAGCACC
                .
F-recons.  192 TGTTTTGGGCCTTCA 206
                | ||||||
TF-cons.   199 CTGTGTGGGCCGGG. 212
```

B



C

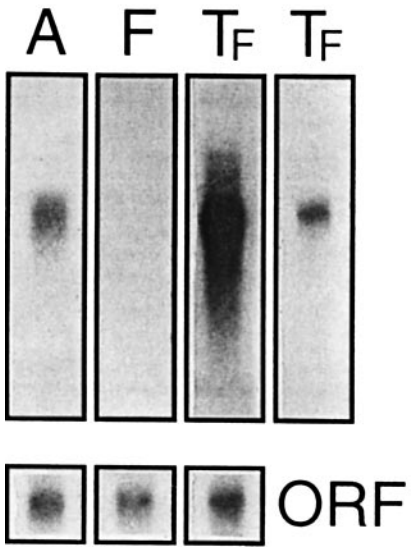


Fig. 2. (**A**) The T_F monomer is a novel variant of the F monomer. This alignment compares the 212 bp consensus sequence of the seven and a half $L1_{spa}$ monomers and the five and two-thirds $L1_{Orl}$ monomers ('T_F cons.') with a 206 bp reconstructed F-consensus monomer ('F-recons.') derived by Adey *et al.* (1994). Note that the identical T_F consensus sequence is derived from either $L1_{spa}$ or $L1_{Orl}$ monomers. The T_F and F consensus sequences are 27% divergent. (**B**) T_F 5′ UTR has promoter activity. pJCB8 contains the promoter for the large subunit of mouse RNA polymerase II (pPol II), pRJD801 contains the L1spa 5′ UTR and pRJD802 contains no promoter sequences. Each of these constructs was co-transfected into LTK– and F9 cells with a luciferase-containing construct (pJCB14). Luciferase activity was used to normalize β-gal activity in extracts prepared 48 h after co-transfection. Normalized β-gal activity of pJCB8 in LTK– cells was arbitrarily set at 1.0. β-Gal activity of other extracts is shown as a fraction of pJCB8 activity. Each bar represents the average β-gal activity of extracts from three independent co-transfections and the horizontal lines are maximum deviations from the means. (**C**) T_F is the most abundant L1 transcript in L1 RNPs. Lanes contain approximately equal volumes of RNA extracted from L1 RNPs which was hybridized to probes for A, F and T_F as indicated, then stripped and rehybridized to a probe (ORF) encompassing ORFs 1 and 2. The A, F and left T_F lanes were exposed for 120 h, showing that T_F RNA predominates in RNPs. The right T_F lane was exposed for 17 h in order to demonstrate that T_F RNA is in a single band of approximately equal intensity to that of A RNA after 120 h exposure.

full-length L1s. Considering that many of our PCR-negative phage probably contained T_F L1s truncated at internal restriction sites during construction of the library, the true number of full-length T_F L1s may approach the total number of T_F 5′ UTR sequences.

T_F 5′ UTR has promoter activity and T_F RNA predominates in ribonucleoprotein particles

To test the novel T_F 5′ UTR for promoter activity, we created an expression construct (pRJD801) in which the $L1_{spa}$ 5′ UTR directs transcription of the β-galactosidase (β-gal) gene (Figure 2B). We transiently transfected mouse LTK– and F9 cells with pRJD801 and assayed for β-gal activity after two days. In both cell lines, we observed substantial activity in both qualitative staining and quantitative enzyme assays (Figure 2B). In the quantitative enzyme assay, we compared the promoter activity of the $L1_{spa}$ 5′ UTR with the activity of a construct (pJCB8) containing the promoter for the large subunit of mouse RNA polymerase II (pPol II). We used pPol II as a reference promoter because its activity is similar in LTK– and F9 cells (Figure 2B). As another control, we transfected a construct (pRJD802) which lacked promoter sequences; cells transfected with this construct had very little β-gal activity (Figure 2B). These data prove that the T_F-type 5′ UTR of $L1_{spa}$ is an active promoter in mouse cell lines.

Since the $L1_{spa}$ 5′ UTR is an active promoter, we asked whether T_F-type L1s are expressed *in vivo*. Previous studies showed that L1 RNA and ORF1 protein co-localize in cytoplasmic ribonucleoprotein (RNP) particles isolated from mouse F9 cells (Martin, 1991). Northern blot analysis of RNA from RNP particles using F-, A- and T_F-monomer probes revealed that T_F-type RNA is much more abundant than A-type RNA (Figure 2C). We also used a probe that did not discriminate among L1 subfamilies (ORF) to ensure that equal amounts of L1 RNA were present in each lane (Figure 2C). Since $L1_{spa}$ and $L1_{Orl}$ both contain the longest known form of the ORF1 LPR (three 42 bp repeats), the abundance of T_F RNA in RNP particles is consistent with a previous demonstration that the larger form of ORF1 protein is most plentiful in particles (Kolosha and Martin, 1995).

$L1_{spa}$ and $L1_{Orl}$ encode RT and retrotranspose in cultured cells

We next determined whether the ORF2s of $L1_{spa}$ and $L1_{Orl}$ encode RT activity in a yeast-based assay. Both ORF2s demonstrated RT activity at levels comparable to the activity encoded by ORF2 of L1.2, the human L1 used as a positive control (Figure 3). Two mutations in the critical RT motif F/Y-X-DD, D709Y and D709N, showed no RT activity in this assay.

Retrotransposition in mammalian cells requires several functions in addition to RT activity and a functional promoter, including ORF1 protein and the endonuclease activity of ORF2 protein (Feng *et al.*, 1996; Moran *et al.*, 1996). To prove that $L1_{spa}$ and $L1_{Orl}$ are active retrotransposons, we used a cell culture assay that accurately detects authentic retrotransposition events (Moran *et al.*, 1996). In this assay, L1 retrotransposition does not require the 5′ UTR when an exogenous promoter is provided. We cloned $L1_{spa}$ and $L1_{Orl}$ lacking their 5′ UTRs downstream of the cytomegalovirus immediate early promoter (pCMV) to create constructs pTN202 and pTN207, respectively (Figure 4). Both constructs generated $G418^R$ foci in LTK– cells (40–300 events per 10^6 hygromycin-resistant cells, Figure 4). In contrast, an RT-defective allele of $L1_{spa}$ (pTN203) failed to generate $G418^R$ foci.

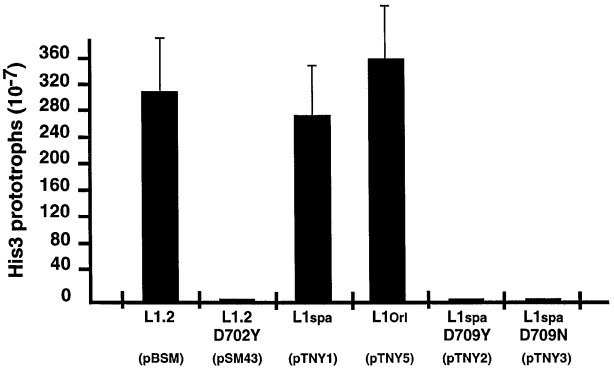


Fig. 3. $L1_{spa}$ ORF2 and $L1_{Orl}$ ORF2 contain reverse transcriptase activity in the *HIS3* pseudogene assay. His^+ prototroph formation requires the presence of a functional reverse transcriptase. Vertical bars represent frequency of prototroph formation, and vertical lines represent average deviation calculated for at least four independent experiments. The name of each tested ORF2 allele is shown beneath the corresponding frequency bar, and constructs bearing each allele are given in parentheses. Human L1.2 ORF2 serves as a positive control, and L1.2 containing the ORF2 D702Y mutation as a negative control. Note that residue 702 in human L1 ORF2 protein is equivalent to residue 709 in mouse L1 ORF2 protein.

PCR analysis on genomic DNA of individual $G418^R$ foci generated by pTN202 revealed that the *neo* insertions lacked the γ-globin intron, consistent with retrotransposition. Furthermore, Southern blots showed that *neo* sequences were located at different genomic positions (data not shown). These data verify that $L1_{spa}$ and $L1_{Orl}$ are capable of high-frequency retrotransposition in mouse cells.

Since $L1_{spa}$ and $L1_{Orl}$ are nearly identical, we concentrated further analyses on $L1_{spa}$. We cloned the 5′ UTR of $L1_{spa}$ into pTN202, creating pTN201, and found that the presence of both pCMV and the 5′ UTR greatly increased retrotransposition frequency compared with that of pTN202 (1400 events per 10^6 hygromycin-resistant cells, Figure 4). To test $L1_{spa}$ in a more biologically relevant manner, we then deleted pCMV from pTN201. This construct, pTN205, which contains the full $L1_{spa}$ sequence, generated retrotransposition events at frequencies similar to pTN202 (100–120 events per 10^6 hygromycin-resistant cells), while pTN206, which lacks both promoters, did not retrotranspose. Thus, retrotransposition of a T_F-subfamily L1 occurs at readily observable frequencies and does not require an exogenous promoter.

Discussion

We have used a cell culture assay to prove that L1s from humans (Moran *et al.*, 1996; Sassaman *et al.*, 1997) and mice (this work) can retrotranspose. Previously, we showed that a human full-length L1 containing a 5′ UTR retrotransposed more frequently than a human L1 lacking a 5′ UTR, even when both constructs contained pCMV (Moran *et al.*, 1996). Here we obtained a similar result with a

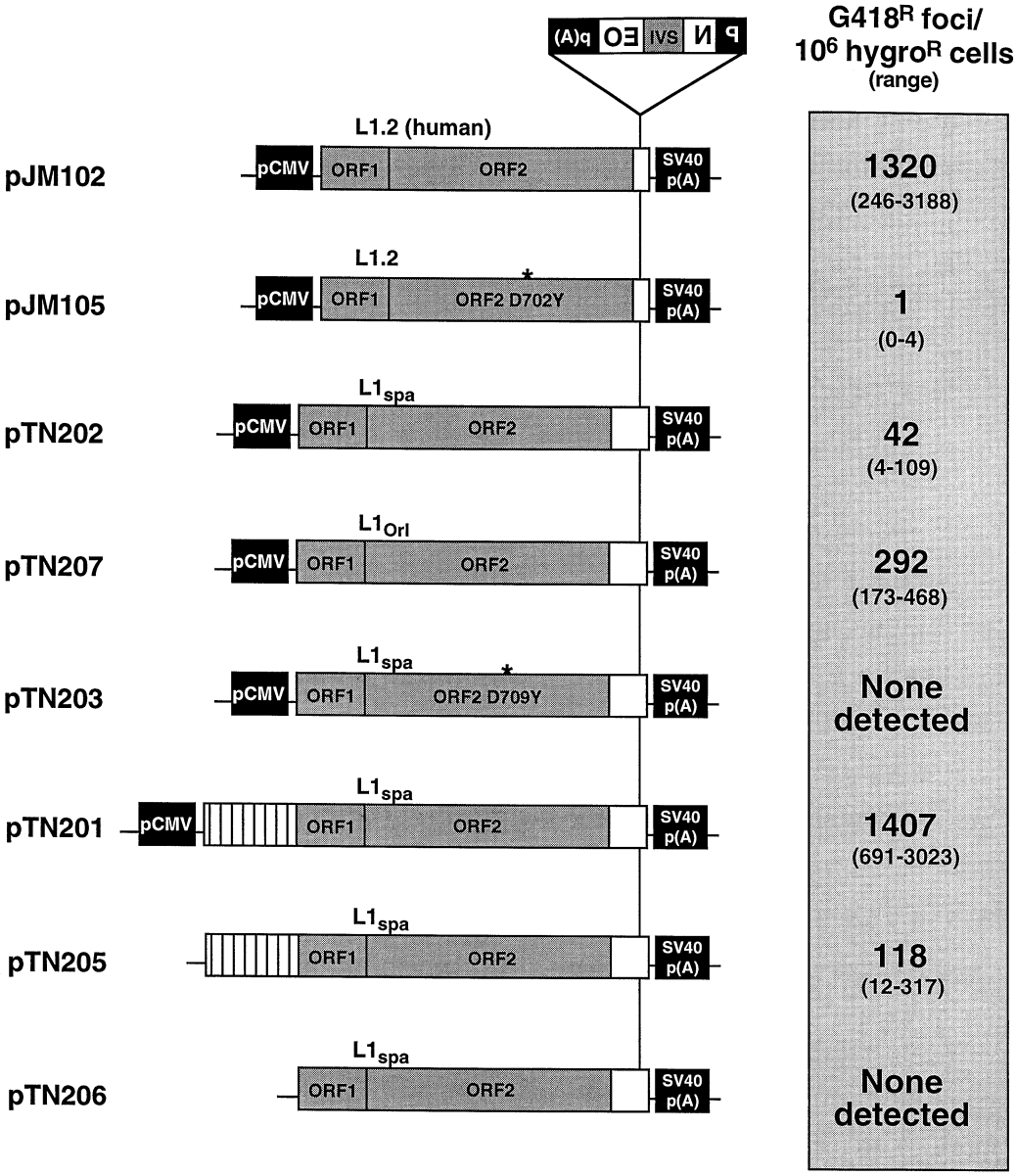


Fig. 4. $L1_{spa}$ and $L1_{Orl}$ retrotranspose at high frequency in LTK– cells. Each construct shown is in a pCEP4 backbone and contains a backward *neo* cassette in its L1 3′ UTR. In the *neo* cassette, 'P' is pSV40, 'NEO' is the *neo* gene, 'IVS' is intron 2 of the γ-globin gene, and 'p(A)' is the *neo* polyadenylation signal. pCMV and the SV40 late polyadenylation signal [SV40 p(A)] are represented by black boxes 5′ and 3′ to L1 sequences. Asterisks in ORF2 of pJM105 and pTN203 represent the location of active site mutations in the reverse transcriptase domains. Open rectangles 5′ to $L1_{spa}$ in pTN201 and pTN205 represent the seven and a half monomers and other 5′ UTR sequence of that L1 element. In the right column, retrotransposition frequency and range for at least six independent transfections are shown as $G418^R$ foci/10^6 hygromycin-resistant cells. For comparison, the previously reported retrotransposition frequency of pJM102 in HeLa cells is 335 events per 10^6 hygromycin-resistant cells (Moran *et al.*, 1996).

mouse L1 (compare pTN201 with pTN202 in Figure 4). This could result from increased L1 transcription on constructs containing both CMV and 5′ UTR promoters. Alternatively, presence of 5′ UTR sequence in an L1 RNA might make it a more efficient intermediate for retrotransposition than a transcript lacking 5′ UTR sequence. Perhaps the 5′ UTR of an active L1 has an additional function in retrotransposition independent of promoter activity, such as providing a protein docking site or improving RNA stability.

The sequences of $L1_{spa}$ and $L1_{Orl}$ suggest that they are members of a novel subfamily that contains active L1s. Two mouse mutations besides spastic and Orleans reeler have been caused by L1 insertions. One insertion into the sodium channel gene, *Scn8a*, is too short to be assigned to a subfamily (Kohrmann *et al.*, 1996). The other insertion, into the *beige* gene, is 1.1 kb long (Perou *et al.*, 1997); it matches the T_F subfamily at 27 of 37 polymorphic nucleotides, and the A subfamily at the other 10 sites. Thus, it is likely that the active progenitor of this L1 is also a member of the T_F subfamily.

All recent L1 retrotranspositions in humans and mice lack stop mutations within the inserted L1 DNA, while all randomly cloned human L1 cDNAs contain stop mutations (Skowronski *et al.*, 1988). It is therefore likely that the proteins encoded by active L1 progenitors function

preferentially to retrotranspose the transcript that encodes them, i.e. they function *in cis* (Dombroski *et al.*, 1991; Moran *et al.*, 1996; Boeke, 1997). Consistently, mutations in conserved domains of an active L1 progenitor markedly reduce its abililty to retrotranspose in cultured cells (Moran *et al.*, 1996). Therefore, it appears that *trans* complementation to retrotranspose transcripts from inactive L1s is a rare event. This *cis* preference would prevent efficient expansion of truncated, rearranged or otherwise mutated L1 transcripts, and curtail promiscuous retrotransposition of other cellular mRNAs. The two T_F L1s described here support *cis* preference because both have retrotransposed recently, contain intact ORFs and retrotranspose autonomously in cultured cells.

Why might many active mouse L1s belong to a particular subfamily? In the human genome, most active L1s are concentrated in a specific family called the Ta subset. There are only about 160 full-length members of this subset, about a quarter of which are active (Sassaman *et al.*, 1997). The mouse T_F subfamily probably presents a strikingly different picture. The first two characterized T_F-subfamily L1s, $L1_{spa}$ and $L1_{Orl}$, are the first examples of recent L1 insertions that have retained retrotransposition capacity. If T_F-subfamily L1s frequently retrotranspose as full-length insertions, they would expand among active L1s, leading to a preponderance of active L1s in the T_F subfamily. Determining the number of T_F elements capable of retrotransposition will provide important information concerning the evolution of the subfamily and its current role in shaping the mouse genome. Considering the large number of full-length T_F L1s, it is probable that this subfamily contains more active members than the ~40 in the human Ta subset, implying a greater potential for L1 mutagenesis in mice than in humans.

Many questions crucial to understanding the impact of L1 in shaping mammalian genomes have been impossible to address because no active L1s have been identified from an experimental organism. Characterization of an active mouse L1 subfamily now provides an opportunity to design whole-animal experiments using tagged active mouse L1s in transgenic mice.

Materials and methods

Bacterial and yeast strains

Bacterial strain DH5α (Stratagene) was used in cloning. Growth media, antibiotic selection and bacterial transformation were according to standard protocols (Sambrook *et al.*, 1989). *HIS3* pseudogene experiments were carried out in strain YH50 (*MATα his3Δ200 ura3-167 trp1Δ1 leu2Δ1 spt3-202*) (Dombroski *et al.*, 1994). Yeast transformation and media were according to standard protocols (Rose *et al.*, 1990).

Cloning techniques, DNA preparation and sequencing

Standard molecular biology techniques were performed as previously described (Sambrook *et al.*, 1989). Site directed mutagenesis was performed as described (Kunkel *et al.*, 1991). Plasmid DNAs were purified on Qiagen maxi columns (Qiagen). DNAs for transfection experiments were tested for superhelicity by electrophoresis on 0.8% agarose–ethidium bromide gels and only highly supercoiled preparations were used for transfections. DNA sequencing was perfomed on an Applied Biosystems DNA sequencer (ABI377). Oligonucleotide sequences used in this study are available upon request. All cloning-related PCR products, ORF2 mutations and relevant restriction fragments were sequenced in their entirety. Nucleotide sequences and deduced protein sequences were analyzed using the GCG software package (Biotechnology Center, University of Wisconsin–Madison, Madison, WI).

Filter hybridization of mouse genomic libraries and phage PCR

Partial *Sau*3AI λ genomic libraries derived from mouse strain 129 were obtained commercially (Stratagene) or from K.Kaestner. The libraries were screened with a probe containing $L1_{spa}$ monomers 5–$7\frac{1}{2}$ liberated by *Not*I and *Xba*I digestion of pTNC7. Secondary and tertiary screens also employed a 1.5 kb ORF2 probe liberated from pTNC7 by *Bst*XI digestion. Probes were labeled by random priming and hybridized under standard conditions (Sambrook *et al.*, 1989). The number of genomic T_F sequences was estimated from the number of positive plaques per filter, the number of phage per plate and the average insert size in the library. Phage purification was performed using standard protocols (Sambrook *et al.*, 1989). For PCR analysis on purified phage, we used $L1_{spa}$-derived oligomers to amplify a 3.6 kb fragment which included most of ORF1 and ORF2. This product contained nucleotides 2183–5814 of the $L1_{spa}$ sequence.

Clones used in yeast experiments

pBSM is a derivative of pSM42 (Dombroski *et al.*, 1994) in which the primer-binding site of Ty1 was mutated (J.V.Moran, Q.Feng, B.A.Dombroski, T.P.Naas, R.J.DeBerardinis, J.D.Boeke and H.H.Kazazian, manuscript in preparation). pSM43 is an active site mutant (ORF2 D702Y) of pSM42 (Dombroski *et al.*, 1994). pTNC1 was constructed by cloning a 3.9 kb *Pfu* polymerase PCR product containing $L1_{spa}$ ORF2 as a *Sal*I–*Sac*I fragment into pBluescript II (pBS, Stratagene). Site-directed mutagenesis of ORF2 was performed using this vector. pTNC2 is an ORF2 D709Y mutant of pTNC1, and pTNC3 is an ORF2 D709N mutant of pTNC1. *Sal*I–*Sac*I fragments were cloned from pTNC1, pTNC2 and pTNC3 into pBSM giving rise to pTNY1, pTNY2 and pTNY3, respectively. pTNC5 was constructed by cloning a 3.9 kb *Pfu* polymerase PCR product containing $L1_{Orl}$ ORF2 as a *Sal*I–*Sac*I fragment into pBS. The relevant *Sal*I–*Sac*I fragment was cloned into pBSM to make pTNY5.

Clones used in tissue culture experiments

Plasmid pNI contains a 15 kb *Not*I genomic DNA fragment with the entire $L1_{spa}$ insertion cloned into pBS (Kingsmore *et al.*, 1994). pA contains an 8 kb *Pvu*II genomic DNA containing the entire $L1_{Orl}$ insertion cloned into pBS (Takahara *et al.*, 1996). pJM102 and pJM105 were previously described (Moran *et al.*, 1996). $L1_{spa}$ was engineered to contain a unique *Not*I site either upstream of its 5′ UTR or immediately upstream of ORF1. A unique *Pac*I site was engineered at nucleotide 7430 in the 3′ UTR. *Not*I–*Pac*I fragments were cloned into pBS resulting in $L1_{spa}$ constructs either containing the 5′ UTR (pTNC201) or lacking it (pTNC202). The *Not*I–blunted *Pac*I fragments from these plasmids were cloned along with a *Bam*HI–blunted *Acc*I fragment containing the *neo* cassette from pJCC9 (Moran *et al.*, 1996) into *Not*I- and *Bam*HI-digested pCEP4 (Invitrogen). The resulting plasmids contained a tagged $L1_{spa}$ either containing the 5′ UTR (pTN201) or lacking it (pTN202). The D709Y mutant in ORF2 was subcloned from pTNC2 into pTNC202 producing pTNC203. A *Not*I–*Xho*I (*Xho*I cuts at one site in the γ-globin intron in the *neo* gene) fragment from pTNC203 was cloned into pTN202 digested with the same enzymes producing pTN203. $L1_{Orl}$ was engineered to contain a unique *Not*I restriction site immediately upstream of ORF1. A *Not*I–*Sfi*I fragment was cloned into *Not*I- and *Sfi*I-digested pTNC202 resulting in pTNC207. The *Not*I–blunted *Pac*I fragment from this plasmid was cloned along with a *Bam*HI–blunted *Acc*I fragment containing the *neo* cassette from pJCC9 (Moran *et al.*, 1996) into *Not*I- and *Bam*HI-digested pCEP4 (Invitrogen) as described above. The resulting plasmid, pTN207, contains a tagged $L1_{Orl}$ lacking the 5′ UTR . Plasmids pTN205 and pTN206 were constructed by cloning a *Not*I–*Xho*I fragment from pTN201 and pTN202, respectively, into pCEP4ΔCMV (Moran *et al.*, 1996).

$L1_{spa}$ 5′ UTR–β-gal constructs

pCMVβ (Clontech) was digested with *Eco*RI, blunted with T4 DNA polymerase and partially digested with *Not*I to yield a 6.4 kb fragment lacking pCMV and the SV40 intron. The 5′ UTR of $L1_{spa}$ was obtained from pTNC7, a pBS derivative containing full-length $L1_{spa}$ as a *Not*I–*Xho*I fragment. pTNC7 was digested with *Bst*XI, which cuts in the pBS polylinker 5′ to the $L1_{spa}$ sequence, blunted with T4 DNA polymerase and digested with *Sty*I, which cuts in monomer 1, resuting in a 1.5 kb fragment containing most of the 5′ UTR. Then a 300 bp *Sty*I–*Not*I PCR product of the remainder of the 5′ UTR was prepared with *Pfu*

polymerase. Ligation of the 1.5 kb blunted *Bst*XI–*Sty*I, 0.3 kb *StyI*–*Not*I and the 6.4 kb *Not*I–blunted *Eco*RI fragments resulted in pRJD800 which contained the entire $L1_{spa}$ 5′ UTR upstream of the β-gal gene. To add back the SV40 intron and create pRJD801, the 1.8 kb *Not*I fragment was liberated from pRJD800, blunted with T4 DNA polymerase and ligated with the blunted *Eco*RI–blunted *Xho*I fragment of pCMVβ. pRJD802 is a deletion derivative of pCMVβ created by removing pCMV by *Eco*RI and *Xho*I digestion, then blunting and recircularizing the backbone. pJCB8 contains the promoter for the large subunit of mouse RNA polymerase II (pPol II) in place of pCMV while pJCB14 contains the mouse phosphoglycerate kinase-1 promoter (pPGK) driving the luciferase gene (Bradford, 1997).

Promoter expression experiments

F9 or LTK– cells were grown as described by ATCC. Cells (4×10^5) were incubated in 7% CO_2 at 37°C for 18 h. Cells were transiently co-transfected with 1 μg of pJCB14 and 1 μg of either pJCB8, pRJD801 or pRJD802 using lipofectamine (BRL). Cell extracts were prepared after 48 h. β-gal and luciferase assays were performed using commercially available kits (Promega).

Cytoplamic RNP particles and Northern analysis

L1 RNP particles of mouse F9 cells were prepared by sedimentation through sucrose step gradients as described (Martin, 1991). RNA was extracted from 20 μl of the 30%/60% interface using TRIZOL (Life Technologies), resolved on formaldehyde–agarose gels and blotted on to nitrocellulose. Probes were made from gel-purified fragments by random-primed labeling with [α-^{32}P]dCTP. The A-monomer probe is from L1Md9 (Shehee *et al.*, 1988), the F-monomer probe is from Padgett *et al.* (1988) and the T_F probe was generated from $L1_{Orf}$. The ORF1–ORF2 probe spans positions 1490–7001 of L1Md-A2 (Loeb *et al.*, 1986). The specificity of the monomer probes was proven by filter hybridization to cloned fragments (data not shown).

HIS3 pseudogene experiments

The *HIS3* pseudogene assay was performed essentially as described (Derr *et al.*, 1991; Dombroski *et al.*, 1994). Yeast strain YH50 was co-transformed with Ty1/L1 ORF2 expression constructs and the indicator cassette plasmid (pSM50). Colonies were selected on SC medium (–Ura, –Trp). Transformants were purified and four colonies were grown as patches on SC medium (–Ura, –Trp) for 3 days at 30°C. In order to induce the expression of the Ty1/L1 construct, the patches were subsequently replica plated on to two different SC (–Ura, –Trp) plates containing 2% galactose and incubated for 5 days at 22°C. After induction, one plate was replica plated to SC medium (–His) to obtain a qualitative assessment of RT activity. Patches from the other plate were diluted in water, plated on SC (–His) and YPD medium, and grown for 4 days at 30°C. The relative RT activity was reported as the number of His^+ colonies/total number of colonies plated.

Retrotransposition assay

The retrotransposition assay was performed as described (Moran *et al.*, 1996). In this assay, an antisense neomycin resistance gene (*neo*) under the control of an SV40 promoter (pSV40) is interrupted by a sense γ-globin intron, and cloned into the 3′ UTR of L1 elements. $G418^R$ cells result only when an L1 message containing antisense *neo* is transcribed, the γ-globin intron is removed by splicing, the transcript is reverse transcribed and integrated into the genome, and the *neo* gene is expressed from pSV40. LTK– cells were grown as described (Moran *et al.*, 1996).

Acknowledgements

We thank Julie Saxton and Jeremy Bradford for technical assistance, Klaus Kaestner for providing a mouse genomic library, and Beth Dombroski for critical reading of the manuscript. We thank the DNA sequencing core facility at the University of Pennsylvania School of Medicine. This work was supported by NIH grants to H.H.K. and S.L.M.; T.P.N. was supported by an EMBO Long-Term Fellowship and J.V.M. by a Damon Runyon–Walter Winchell Cancer Research Fund Fellowship (DRG 1332).

References

Adey,N.B., Tollefsbol,T.O., Sparks,A.B, Edgell,M.H. and Hutchison,C.A.,III (1994) Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc. Natl Acad. Sci. USA*, **91**, 1569–1573.

Boeke,J.D. (1997) LINEs and Alus—the polyA connection. *Nature Genet.*, **16**, 6–7.

Bradford,J. (1997) Development of an L1-based transposon mutagenesis system in mice. Master's thesis, University of Pennsylvania, Philadelphia, PA.

Burton,F.H., Loeb,D.D., Voliva,C., Martin,S.L., Edgell,M.H. and Hutchison,C.A.,III (1986) Conservation throughout Mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. *J. Mol. Biol.*, **187**, 291–304.

Derr,L.K., Strathern,J.N. and Garfinkel,D.J. (1991) RNA-mediated recombination in *S. cerevisiae*. *Cell*, **67**, 355–364.

Dombroski,B.A., Mathias,S.L., Nanthakumar,E., Scott,A.F. and Kazazian,H.H.,Jr (1991) Isolation of an active human transposable element. *Science*, **254**, 1805–1808.

Dombroski,B.A., Feng,Q., Mathias,S.L., Sassaman,D.M., Scott,A.F., Kazazian,H.H.,Jr and Boeke,J.D. (1994) An *in vivo* assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **14**, 4485–4492.

Fanning,T.G. and Singer,M.F. (1987) LINE-1: a mammalian transposable element. *Biochim. Biophys. Acta*, **910**, 203–212.

Feng,Q., Moran,J.V., Kazazian,H.H.,Jr & Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.

Hohjoh,H. and Singer,M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **15**, 630–639.

Hohjoh,H. and Singer,M.F. (1997) Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.*, **16**, 6034–6043.

Holmes,S.E., Dombroski,B.A., Krebs,C.M., Boehm,C.D. and Kazazian,H.H.,Jr (1994) A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. *Nature Genet.*, **7**, 143–148.

Hutchison,C.A.,III, Hardies,S.C., Loeb,D.D., Shehee,W.R. and Edgell,M.H. (1989) LINES and related retrotransposons: long interspersed sequences in the eucaryotic genome. In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology Press, Washington, DC, pp. 593–617.

Kazazian,H.H.,Jr, Wong,C., Youssoufian,H., Scott,A.F., Phillips,D.G. and Antonarakis,S.E. (1988) Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, **332**, 164–166.

Kingsmore,S.F., Giros,B., Suh,D., Bieniarz,M., Caron,M.G. and Seldin,M.F. (1994) Glycine receptor β-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nature Genet.*, **7**, 136–142.

Kohrmann,D.C., Harris,J.B. and Meisler,M.H. (1996) Mutation detection in the *med* and *Med^J* alleles of the sodium channel *Scn8a*. *J. Biol. Chem.*, **271**, 17576–17581.

Kolosha,V.O. and Martin,S.L. (1995) Polymorphic sequences encoding the first open reading frame protein from LINE-1 ribonucleoprotein particles. *J. Biol. Chem.*, **270**, 2868–2873.

Kolosha,V.O. and Martin,S.L. (1997) *In vitro* properties of the first open reading frame protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl Acad. Sci. USA*, **94,** 10155–10160.

Kunkel,T.A., Bebenek,K. and McClary,J. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.*, **204**, 125–139.

Loeb,D.D., Padgett,R.W., Hardies,S.C., Shehee,W.R., Comer,M.B., Edgell,M.H. and Hutchison,C.A.,III (1986) 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.

Martin,S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.*, **11**, 4804–4807.

Mathias,S.L., Scott,A.F., Kazazian,H.H.,Jr, Boeke,J.D. and Gabriel,A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.

Miki,Y., Nishisho,I., Horii,A., Miyoshi,Y., Utsunomiya,J., Kinzler,K., Vogelstein,B. and Nakamura,Y. (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.*, **52**, 643–645.

Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H.,Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.

Morse,B., Rotherg,P.G., South,V.J., Spandorfer,J.M. and Astrin,S.M. (1988) Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. *Nature*, **333**, 87–90.

Mulhardt,C., Fischer,M., Gass,P., Simon-Chazzottes,D., Guenet,J.-L., Kuhse,J., Betz,H. and Becker,C.-M. (1994) The spastic mouse: aberrant splicing of glycine receptor β subunit mRNA caused by intronic insertion of L1 element. *Neuron*, **13**, 1003–1015.

Narita,N., Nishio,H., Kitoh,Y., Ishikawa,Y., Ishikawa,Y., Minami,R., Nakamura,H. and Matsuo,M. (1993) Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.*, **91**, 1862–1867.

Padgett,R.W., Hutchison,C.A.,III and Edgell,M.H. (1988) The F-type 5′ motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. *Nucleic Acids Res.*, **16**, 739–749.

Perou,C.M., Pryor,R.J., Naas,T.P. and Kaplan,J. (1997) The *bg* allele mutation due to a LINE 1 element retrotransposition. *Genomics*, **42**, 366–368.

Rose,M.D., Winston,F. and Hieter,P. (1990) *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sassaman,D.M., Dombroski,B.A., Moran,J.V., Kimberland,M.L., Naas,T.P., DeBerardinis,R.J., Gabriel,A., Swergold,G.D. and Kazazian,H.H.,Jr (1997) Many human L1 elements are capable of retrotransposition. *Nature Genet.*, **16**, 37–43.

Schichman,S.A., Severynse,D.M., Edgell,M.H. and Hutchison,C.A.,III (1992) Strand-specific LINE-1 transcription in mouse F9 cells originates from the youngest phylogenetic subgroup of LINE-1 elements. *J. Mol. Biol.*, **224**, 559–574.

Schichman,S.A., Adey,N.B., Edgell,M.H. and Hutchison,C.A.,III (1993) L1 A-monomer tandem arrays have expanded during the course of mouse L1 evolution. *Mol. Biol. Evol.*, **10**, 552–570.

Severynse,D.M., Hutchison,C.A.,III and Edgell,M.H. (1992) Identification of transcriptional regulatory activity within the 5′ A-type monomer sequence of the mouse LINE-1 retroposon. *Mammal. Genome*, **2**, 41–50.

Shehee,W.R., Chao,S., Loeb,D.D., Comer,M.R., Hutchison,C.A.,III and Edgell,M.H. (1988) Determination of a functional ancestral sequence and definition of the 5′ end of A-type mouse L1 elements. *J. Mol. Biol.*, **196**, 757–767.

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

Takahara,T. *et al.* (1996) Dysfunction of the *Orleans reeler* gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. *Human Mol. Genet.*, **5**, 989–993.

Voliva,C.F., Jahn,C.L., Comer,M.B., Hutchison,C.A.,III and Edgell,M.H. (1983) The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. *Nucleic Acids Res.*, **11**, 8847–8859.

Received August 18, 1997; revised October 14, 1997; accepted November 5, 1997