Expression of SINE and LINE loci in gonadal and somatic tissue of *P. maniculatus*

Saedeh Sepehri and David H Kass*

Department of Biology, Eastern Michigan University, Ypsilanti, MI 48197, USA

Received: January 7, 2015; accepted: February 9, 2015.

In this study, we isolated RT-PCR products generated from tissues of *Peromyscus maniculatus* to discern the expression of individual SINE (ID) and LINE (L1 and *mys*) loci in somatic and germ-line tissue, to determine if there is a limited subset of loci being expressed that would potentially exhibit features of a master gene. Though several L1 loci were expressed, a single L1 transcript based on RT-PCR of the ORF2 gene, exhibited over 100 identical matches to sequences within the *P. maniculatus* genomic database suggestive of a currently active master gene. In contrast, although *mys* expression also occurs in brain and testes, clones with complete reading frames, from the amplified region, at best yielded numerous matches of 98-99% similarity to genomic sequences. Additionally, no ID transcript matched the ID consensus sequence, albeit both a testes ID transcript and the BC1 transcript each exactly matched six genomic sequences, supporting the existence of distinct equivalently active source genes. However, 28 identical matches were observed when queried with the ID consensus sequence suggesting neither of these transcripts represent the predominant source for new copies. Overall, expression of various loci for each retrotransposon family occurs in the testes, with the consensus of expressed loci being identical to the genomic consensus indicating several source genes exist for each retrotransposon family that have undergone various modifications from their respective ancestral master genes.

Keywords: Retrotransposons; SINEs; LINEs; RT-PCR; *Peromyscus*.

.

Abbreviations: RT-PCR: reverse transcription polymerase chain reaction. LINE: long interspersed DNA element. SINE: short interspersed DNA element. RACE: rapid amplification of cDNA ends. LTR: long terminal repeats. ERV: endogenous retrovirus. ORF: open reading frame.

***Corresponding author:** David H. Kass, Department of Biology, 441 Mark Jefferson Science Complex, Eastern Michigan University, Ypsilanti, MI 48197, USA. E-mail: dkass@emich.edu. Phone: +1 734 487 2377. FAX: +1 734 487 2388

Introduction

Retrotransposons represent a group of transposable elements (TEs) that generate new copies via an RNA intermediate. Long interspersed DNA elements (LINEs) and short interspersed DNA elements (SINEs) represent the most abundant TE groups in both human and mouse genomes, comprising 34% and 27.4% respectively of their DNA sequence, with LTR endogenous retroviral sequences comprising 8% and 9.9% respectively [1, 2]. Both SINEs and LINEs remain active in these genomes continuing to generate genetic diversity [3, 4]. Although most retrotransposon integrations are apparently neutral; with recent insertions yielding presence/absence polymorphisms, integrations into certain locations

have yielded deleterious effects such as neurofibromatosis in an individual resulting from a *de novo* integration of *Alu* (SINE) in the NF1 gene [5] and hemophilia resulting from a *de novo* L1 (LINE) integration into the factor VIII clotting factor gene [6]. Numerous additional human disorders have been attributed to SINE and LINE integrations [3] as well as disorders in rodents [7]. Alternatively, retrotransposon integrations may contribute new functions in the genome, also referred to as exaptation [4, 8] possibly providing a network of *cis*-acting regulatory elements over time [9], as well as developing new proteins via exon shuffling by 3'-transduction of L1 [10]. LINEs may yield genetic heterogeneity proposed to be important in brain function based on *de novo* L1 integrations in neurons [11]. TEs have also been proposed to be involved in epigenetic regulation of the genome [12], as well as epigenetic regulators of phenotypic variation [13]. A well-established example of an exapted SINE is the ID-element containing BC1 RNA gene, which is highly transcribed in the brain and neurons and regulated by neuronal activity [8, 14].

There are limited retrotransposon loci capable of generating new copies [15]. Matches between the ID-consensus sequences of various rodents to their BC1 gene [16] led to the determination that BC1 serves as a master gene for the generation of new ID element integrations. However, BC1 could not account for the high copy number of ID elements in rat, of which many contain distinct variants [17] as well as in the deer mouse, in which all randomly isolated ID elements isolated from a genomic library screen were distinct from the deer mouse BC1 gene [18].

LTR-endogenous retroviruses (ERVs) are derived from ancient retroviruses and are retrotranspositionally incompetent in present-day humans [19]. They represent a very small portion of the human genome although there is growing speculation that demethylation of certain elements may reactivate them triggering malignant disorders and autoimmune responses [19]. However, various mechanisms have been proposed to keep ERVs stagnant [19, 20]. In contrast, the house mouse (*Mus musculus*) contains the intracistral A particle (IAP) LTR-retrotransposons that have demonstrated recent activity [21, 22]. Additionally, the white-footed mouse (*Peromyscus leucopus*) contains the retrotransposon *mys*, which has shown recent activity [23, 24]. Therefore, ERVs of different origins are active in various mammalian groups.

Non-autonomous SINEs are dependent on LINE proteins for their propagation [20, 25]. However, for integrations to be inherited, it is necessary for retrotransposition to occur in early development or in the germ line. There is evidence for both [26], although the use of transgenic rats and mice suggests that most L1 retrotransposition in early embryogenesis results more in somatic mosaicism than in heritable integrations [3, 27]. SINE and LINE expression have been identified in the germ lines of various mammals [28, 29]. The use of a reporter construct in transgenic mice supports IAP element expression essentially restricted in the male germ line [30]. Full-length L1 RNA and ORF1-encoded protein are expressed in the male germ line of the European house mouse (*M. musculus*) during meiosis [28]. The L1 ORF1 encodes an RNA-binding protein, whereas ORF2 of the L1 element encodes a multifunctional protein that includes endonuclease and reverse transcriptase activity [10]. Additionally Ostertag et al. [31] demonstrated retrotransposition of a human L1 in the male germ cells of a transgenic mouse. The expression of BC1 in testes is consistent with utilizing LINE machinery for heritable male germ-line integrations of ID elements [14].

Peromyscus is a choice model for gaining insights into the association of the heritable nature of SINE and LINE integrations as recent activity of L1 [32] and LTR retrotransposons [24, 25] have been exhibited, as well as the identification of young ID elements [18]. Additionally, the *P. maniculatus* BC1 gene

contains unique sequence variants in contrast to other rodents of the Muroidea superfamily, including a guanine at position 48 of the ID portion, and a TGTT sequence at the 3' end [18], whereas the consensus of individual genomic elements in *P. maniculatus* correspond to the *M. musculus* ID consensus sequence [18]. Therefore, in this study we examined germ-line and somatic tissue for expressed loci in the deer mouse based on current activity of these three types of retrotransposons, as well as based on the distinct variations between previously isolated genomic ID elements and the presumptive BC1 master gene.

Materials and Methods

Tissue samples:
Deer mouse (*Peromyscus maniculatus)* frozen preserved brain, liver, and testes tissue samples were purchased from the *Peromyscus* Stock Center at the University of South Carolina.

Isolation of RNA from animal tissue:
RNA was isolated from *Peromyscus maniculatus* brain, liver, and testes tissues, using the SV total RNA isolation system kit (Promega, Madison, WI) according to the manufacturer's protocol. Estimates of purity and concentration of the DNase-treated extracted RNA were assessed by absorbance readings using a spectrophotometer.

Isolation of L1 and mys cDNA using one- step reverse- transcriptase (RT) PCR
To amplify L1 and *mys* elements, the Access RT-PCR system (Promega Madison, WI) was used. The PCR primers used to amplify the L1 element were designed based on the consensus sequences of the ORF2 (open reading frame 2) region of deer mouse, and purchased from BIONEER (Alameda, CA). The primers used to amplify the *mys* elements were derived from the ORF1 region of the *Peromyscus leucopous* full-length *mys-1* element, and purchased from Integrated DNA Technologies (IDT) Inc. (Coralville, Iowa). The primer sets were as follows: LINE 72F: 5'-ATACAAGATCAACTA-3'; LINE 365R: 5'-TTGAATCTGTAGATTGCTT-3'; Mys 620F: 5'-GCAAGCAACTTCGGGAATCT-'3; Mys 1100R: 5'-AAGTCTTCTAGGTATACATA-3'. The condition for first strand cDNA synthesis was 1 cycle for the reverse transcription activity at 45^{o}C for 45 minutes, and 1 cycle for AMV RT enzyme inactivation and RNA/primer denaturation at 94^{o}C for 2 minutes. Conditions for the second strand cDNA synthesis and PCR amplification were as follows: 40 cycles of 94^{o}C for 30 seconds, 41^{o}C for 1 minute, and 68^{o}C for 2 minutes with a final extension at 68^{o}C for 7 minutes.

Isolation of ID cDNA using two-step RT-PCR
To detect the expression of ID elements a two-step RT-PCR was performed. First strand synthesis of cDNA was generated utilizing a 3' RACE primer (5'-GCCTTCGAATTCAAGTTTTTTTTT TTTTT-3'). To ensure that RT-PCR products were generated from RNA polymerase III-derived transcripts, as opposed to ID elements within RNA polymerase II-transcribed genes, we size fractionated the cDNA from first strand synthesis using a Microcon YM-100 column (Millipore) with a cutoff of 300 nt, and collected the flow-through. We used 10 µl of this cDNA as a template for the second step RT-PCR in a 50 µl volume using 1X GoTaq buffer, 200 nM dNTPs, 250 nM Pm ID-F primer (5'-GGGGTTGGGGATTTAG-3'), 250 nM RACE primer (5'-GCCTTCGAATTCAAG3') and 1 unit GoTaq enzyme (Promega, Madison, WI) under the following conditions: 95^{o}C for 2 min, followed by 32 cycles of 94^{o}C for 1 minute, 50^{o}C for 1 minute and 72^{o}C for 1 minute, with a final 5-minute extension at 72^{o}C. As a control for possible contaminating genomic DNA in the samples, RT-PCR was performed using RNA isolated from the testes tissue without the reverse transcriptase enzyme. The RT-PCR was analyzed on 1.5% agarose gels stained with GelRed (Biotium, Hayward,CA).

In order to analyze individual transcripts, RT-PCR products were shot-gun cloned into the pGEMT-Easy vector (Promega, Madison, WI)

and transformed into JM109 cells according to the manufacturer's protocols. Using blue/white colony selection, plasmid DNA was isolated from white colonies using the Wizard Plus SV miniprep DNA Purification System (Promega, Madison, WI) and analyzed for inserts by digestion with *Eco*RI, and separation by electrophoresis using a 1.5% agarose stained with ethidium bromide.

Cloned RT-PCR products were submitted to Functional Biosciences, Inc. (Madison, WI) for DNA sequencing. The sequences were analyzed for similarity to SINEs and LINEs by screening the GenBank database using BLAST [33]. Sequence alignments were performed using the Clustal W program with molecular phylogenies generated by neighbor joining (Tamura-Nei distance) using default settings in MacVector (MacVector, Inc., Cary, NC). To assess the potential of the RT-PCR products as source genes, Trace BLASTs scanning the developing *P. maniculatus* genomic (WGS) DNA library (http://www.hgsc.bcm.edu/other-mammals/peromyscus-genome-project) were performed with both coverage (%C) representing the proportion of the length of the sequence that matches, and identity (%I) representing the similarity of the sequences.

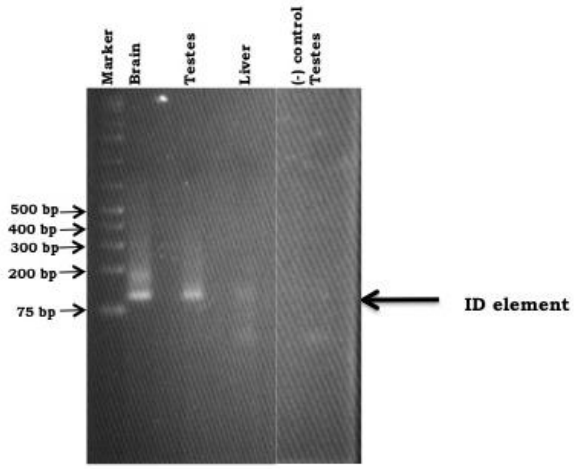


Figure 1. RT-PCR and controls (no reverse transcriptase) of ID elements from fractionated RNA (cDNA) isolated from deer mouse tissue, and analyzed on a 2% agarose gel. Lanes are labeled in the figure. The negative control contains no reverse transcriptase. Marker used was GeneRuler 1kb Plus DNA ladder from Thermo Scientific, (Waltham, MA).

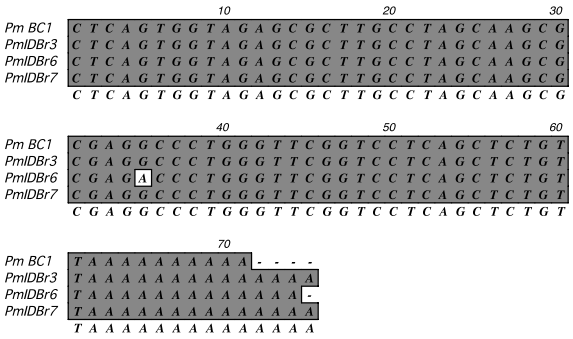


Figure 2. DNA sequence alignment of ID-generated RT-PCR clones from deer mouse brain with the known BC1 gene.

Results

Analysis of ID element transcription in *Peromyscus maniculatus* somatic and germ-line tissues

ID elements isolated from a *P. maniculatus* genomic DNA library demonstrate variation from the BC1 master gene [18] and therefore we analyzed transcripts from the brain and germ-line tissues to depict if ID elements predominantly integrated prior to changes to BC1 or if alternative loci are responsible for their mobility. RT-PCR generated relatively intense bands at the anticipated size, which would be roughly 100 bp considering the size of the ID element, A-tail, and RACE primer, for brain and testes tissues, and a less intense band in the liver (figure 1) using size fractionated cDNA supporting the isolated transcripts are not contained within RNA polymerase II-based transcripts. In contrast, we performed RT-PCR of B2 elements and observed greater intensity bands in liver, and testes with less intensity in brain (data not shown), suggesting variable levels of expression of different TEs in different tissues. We shot-gun cloned the products from brain and testes, and analyzed the sequences of individual clones. Three clones generated from the brain RNA contained the unique deer mouse diagnostic BC1 nucleotides, and were all virtually identical except for a single variant within one clone (figure 2). Among eleven clones isolated from the testes, three were identical to each other and BC1, with a slightly

smaller A-tail in one. None of the other eight clones contained the diagnostic BC1 nucleotides of deer mouse, nor exhibited an exact match to each other (figure 3a). Therefore, more than one ID-containing locus is transcribed in the testes. An alignment of the consensus sequence of the eight non-BC1 testes ID transcripts with the ID consensus sequence (IDCS) demonstrated an exact match, albeit one nucleotide (position 63 of ID or 47 when deleting the primer) was polymorphic for a purine (figure 3b), suggesting these are independently active loci descended from a common ID master locus, which is not BC1. Each clone may represent a potentially retrotranspositionally-competent source gene.

Table 1. Number of BLAST matches of 100% coverage and 100% identity resulting from screening the *P. maniculatus* WGS database queried with testes RT-PCR (non-BC1) product sequences, as well as the ID consensus sequence and the BC1 sequence.

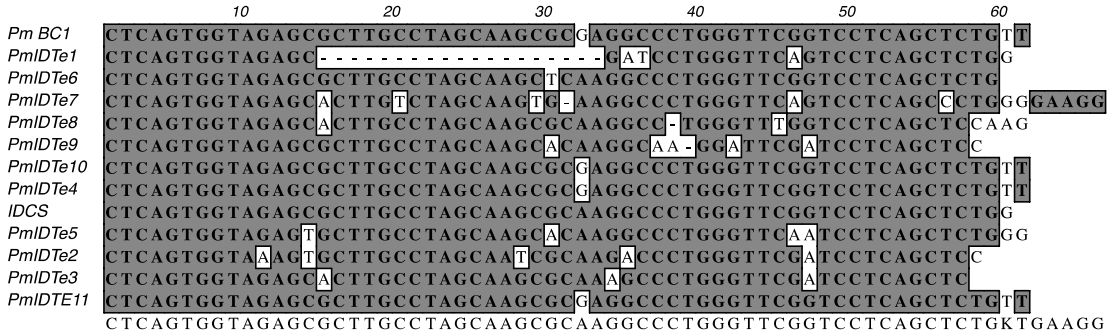
Query Sequence	*P. maniculatus* WGS Trace BLAST matches to 100% coverage (C) and 100% identity (I)
PmIDTe1	2
PmIDTe2	1
PmIDTe3	3
PmIDTe5	0
PmIDTe6	7
PmIDTe7	2
PmIDTe8	0
PmIDTe9	0
IDCS	28
PmBC1	6

To further analyze the possibility of multiple ID source genes, RT-PCR sequences were used as query sequences to scan the developing *P. maniculatus* genomic (WGS) DNA library (http://www.hgsc.bcm.edu/other-mammals/peromyscus-genome-project) via NCBI Trace BLAST to identify the number of hits containing 100% coverage (C) and 100% identity (I) (table 1), omitting the sequences corresponding to the primer. Since the Te1 clone contained a large deletion, we analyzed the two Te1 hits further and found them to be the same locus. Additionally, among the three hits from Te3, two were from the same locus. Therefore, the numbers are generally overestimates. However, the number of exact matches for Te6 and the PmBC1 ID portion were both higher and equivalent (primer sequences were omitted for all cases), hence both potentially active source genes. We also queried the database with the ID consensus sequence, which is the same sequence for both the mouse and the consensus of randomly isolated clones from a genomic *P. maniculatus* library [18], and obtained 28 hits of 100% C and 100% I (table 1). In general, there may be a master gene that was not picked up by our analysis that matches IDCS, or there are less active source genes derived by an ancestral master gene, which would combine to resemble the consensus sequence. The combination of a higher proportion of non-BC1 transcripts from testes and BLAST hits to IDCS, in comparison to BC1, support BC1 as not being the primary master gene, and perhaps the Te6 clone represents a currently active master gene.

Analysis of expression of *P. maniculatus* LINEs in germ-line and somatic tissue

RT-PCR was performed from RNA in brain, liver, and testes to assess expression of L1 loci in these tissues (figure 4). Although not fully quantitative, there appears to be less transcription in the brain compared to the other tissues, particularly considering the contrast to ID elements which appeared highly expressed in brain (figure 1). Individual L1 intra-ORF2 clones were isolated from brain and testes RT-PCR products. Although a doublet was observed in PCR (figure 4), all the shot-gun cloned inserts were the expected 300 bp. Additionally, we isolated the 400 bp fragment from the gel and cloned it, and also obtained a 300 bp insert (data not shown) suggesting the possibility of a mobility shift. The sequences (with removal of primers) from three clones derived from brain and four from testes were aligned (figure 5) with that of human, mouse (*M. musculus*), and previously determined *P. maniculatus* ORF2 sequences representing the two distinct *Peromyscus* L1 lineages [32]. A molecular phylogeny of the nucleotide sequences (figure 6) supports the two distinct lineages [32], with transcripts clustering with one or the other. Along with the finding that no cloned L1

A.



B.

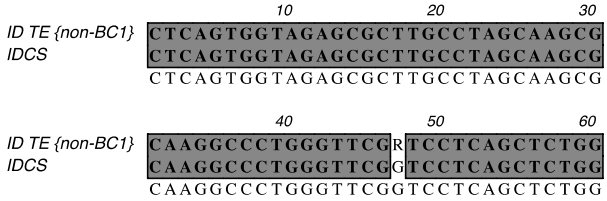


Figure 3. Sequence alignment of ID elements. a) Elements isolated by RT-PCR from deer mouse testes, aligned with the deer mouse (and rodent) ID consensus sequence (IDCS) and deer mouse BC1. A tails were omitted. b) Alignment of the non-BC1 ID consensus sequence of testes transcripts to the ID genomic consensus sequence.

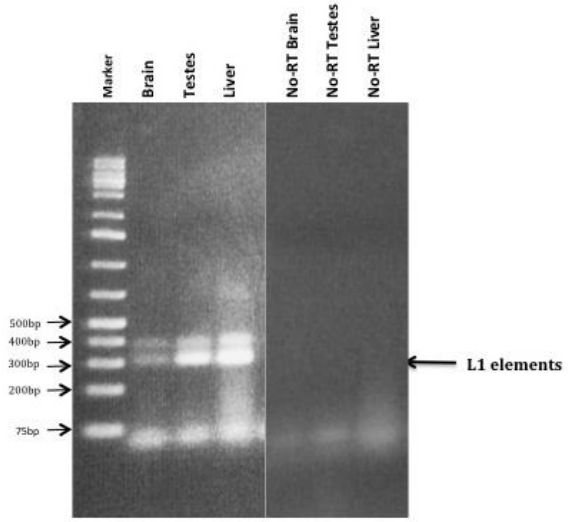


Figure 4. RT-PCR and controls (no reverse transcriptase) of LINE L1 products detected on a 1.5% agarose gel. Marker used was GeneRuler 1kb Plus DNA ladder from Thermo Scientific, (Waltham, MA).

sequences were identical, the data indicate several loci are expressed, consistent with the possibility of several source genes. Additionally, all but one clone, which was derived from testes, displayed an open reading frame for the complete cloned region further consistent with several potential source genes. When translated to amino acid sequences and aligned (figure 7), two clones, one from brain (Br7) and one from testes (Te8) were identical to each other and to the consensus amino acid sequence providing an additional attribute as potential source genes.

To further address the possibility of Br7 and Te8 representing source genes, albeit an expressed L1 in brain would not generate a heritable integration, we performed a Trace BLAST queried with cloned nucleotide sequences to the *P. maniculatus* WGS database. Br7 generated numerous hits of 100% coverage (100% C) and 99% identity (99% I). Manually examining the first several hits, the same nucleotide differed. When using Te8 as the query sequence, there were 112 hits of 100% C and 100% I, followed by numerous hits of 100% C and 99% I. Manual inspection of the first several 99% matches exhibited the same nucleotide variation. These findings are consistent with the possibility of Te8 representing a source gene. We also utilized the one sequence (Te10) that contained several stop codons, and retrieved only 1 hit of 100% C and 100% I, with many 100% C but only 98% I. Both these features are inconsistent with Te 10

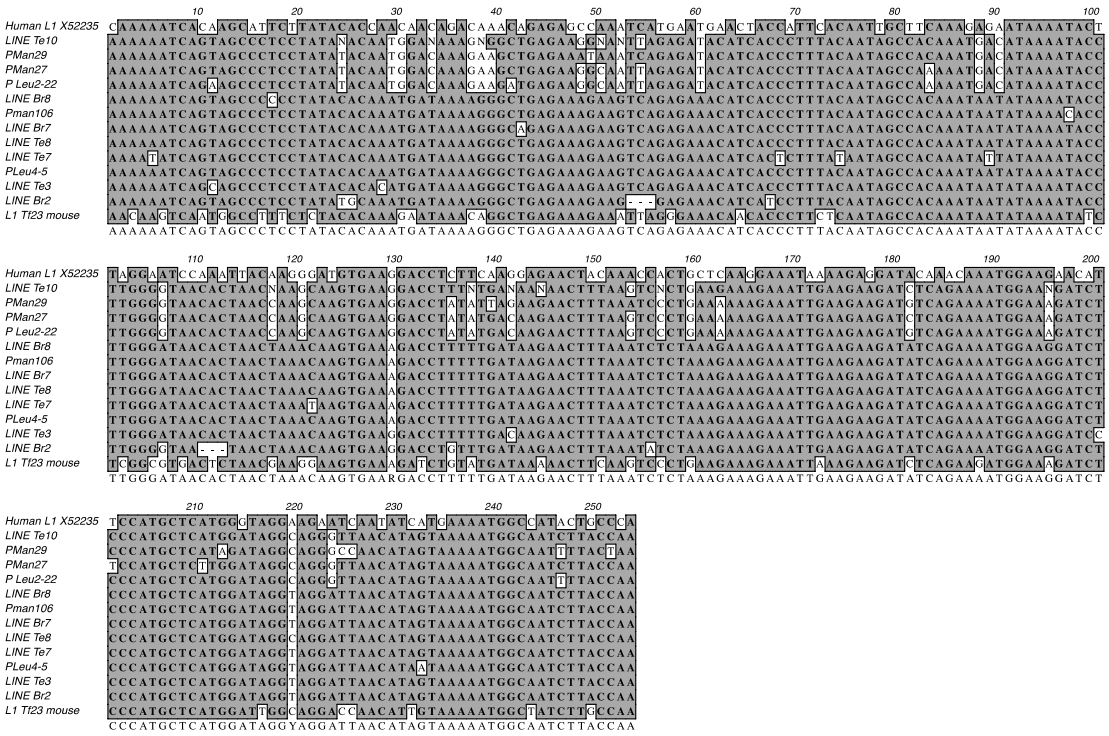


Figure 5. Alignment of LINE ORF2 deer mouse RT-PCR cloned products to a genomic sequence, and to mouse and human segments of full-length elements. Te refers to testis derived and Br to brain derived. Genomic L1 elements representing two L1 lineages in *Peromyscus* [32] are included.

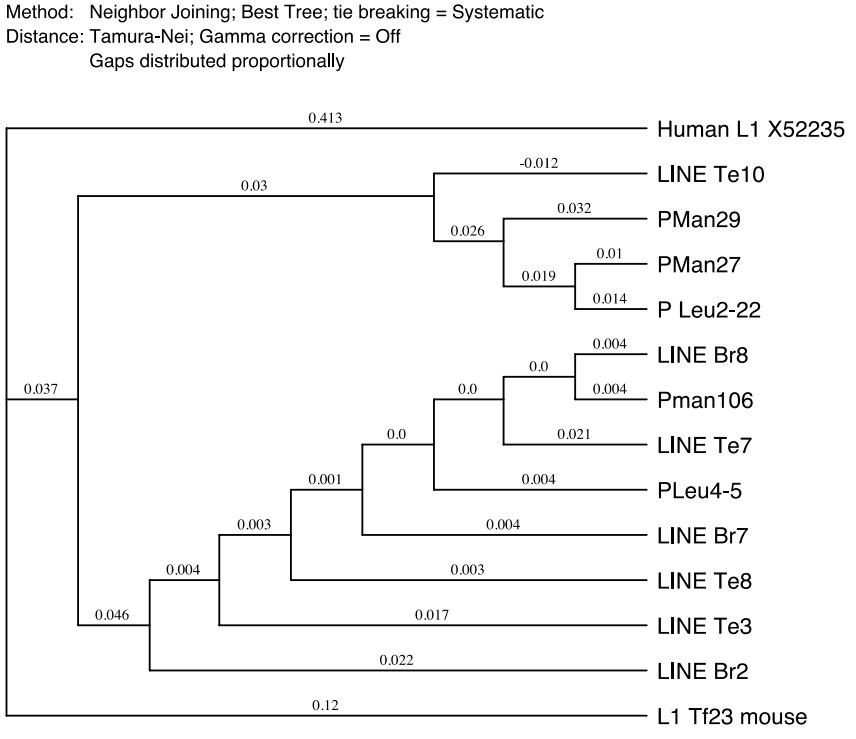


Figure 6. Molecular phylogeny of nucleotide sequences of RT-PCR products with human (accession # X52235), mouse (accession #AF08110), and genomic L1 elements representing two L1 lineages in *Peromyscus* [32]

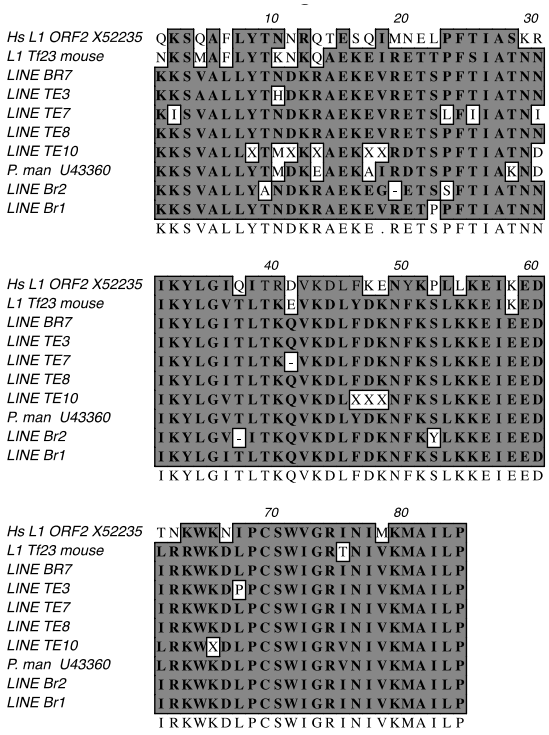


Figure 7. Alignment of amino acid sequences from translated L1 RT-PCR products with previously determined mouse, human, and *Peromyscus* sequences.

representing a potential source gene. Considering the higher level of sequence identity and the greater number of transcripts of the lineage (lineage 1 [32]) containing LINE Te8, there appears to be more recent activity of this L1 lineage in relation to the other. Since Te3 was more divergent than the other elements of the more active lineage and derived from testes, we assessed its potential as a source gene by querying the *P. maniculatus* genomic database, obtaining no matches of 100% C and 100% I, six matches of 100% C and 99% I (all the same variant) and numerous 100% C and 98% I with one variant being consistent and the second variant differing. This finding reduces the likelihood of Te3 representing an active source gene. Overall, these data are consistent with Te8 representing a portion of an active master gene. Although Br7 is identical to the amino acid sequence of Te8 (figure 7), the lack of 100% hits and not being from the germ line, is consistent with the possibility of being active, possibly generating somatic integrations, but not serving as a master gene for heritable integrations.

To gain insights into important amino acid sequences associated with L1 propagation, we incorporated ORF2 amino acid sequences of L1 elements from human (accession X52235) and mouse (accession AF08110) representing young subfamilies [34, 35], plus a translated *P. maniculatus* genomic sequence (accession PNU43360) into an alignment of translated RT-PCR clones (figure 7) to develop a molecular phylogeny (figure 8). It is evident that certain amino acids are highly conserved among mammalian LINEs (figure 7). As anticipated the human L1 ORF2 is most distantly related, as well as equivalently related to both the mouse and deer mouse amino acid sequence (figure 8). Additionally, the mouse ORF2 forms an outgroup from the *Peromyscus* sequences (figure 8).

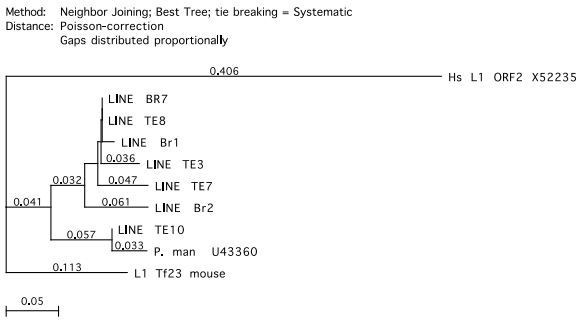


Figure 8. Molecular phylogeny of amino acid sequences of translated deer mouse RT-PCR products, with a genomic deer mouse element and ORF2 segments of full-length human and mouse L1 elements.

Analysis of *mys* expression in somatic and germ-line tissue.

We utilized RT-PCR with ORF1-based primers to analyze expressed *mys* elements in somatic and germ-line tissue (figure 9), verifying the lack of contaminating genomic DNA by performing the same reaction without reverse transcriptase (data not shown). Though the testes tissues demonstrated an amplified product of the expected 480 bp, the product within the brain was much less intense than the other tissues, which parallels the L1 results, but a reverse

finding to the ID results. This is consistent with BC1 being highly expressed in the brain.

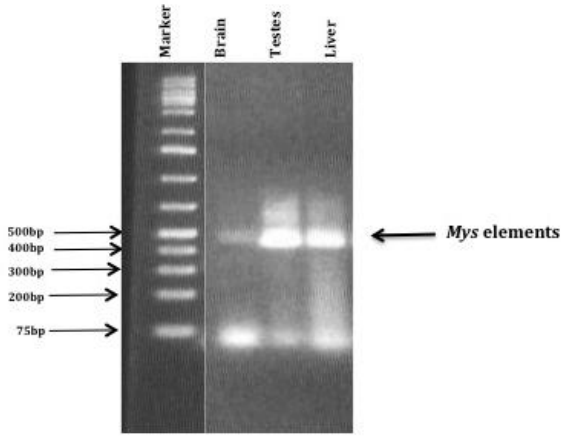


Figure 9. RT-PCR products of *mys* ORF1 from deer mouse tissues analyzed on a 1.5% agarose gel. Marker used was GeneRuler 1kb Plus DNA ladder from Thermo Scientific, (Waltham, MA)

We analyzed sequences from four clones isolated from brain RNA and six from the testes (figure 10) excluding one clone from the testes that was evidently not a *mys* element (data not shown), and contrasted these to the corresponding portion of the *P. leucopus* full-length *mys*-1 sequence. An alignment of the DNA sequences demonstrated 93% or greater sequence identity to *mys*-1, and minimally 91.5% similarity to each other, but mostly 96% or greater. The clone labeled Te4 from testes displayed the greatest nucleotide sequence identity to *mys*-1 (99.5%). The translated amino acid sequence from the Te4 clone yielded only a single variant in the analyzed ORF1 region, in relation to *mys*-1 (figure 11, involving an isoleucine substitution for a valine). Three testes clones exhibited a complete open reading frame for the region analyzed, four (2 testes, 2 brain) stopped just short but at the same position, one had a stop codon near the end, and one did not exhibit a notable open reading frame. An alignment of the amino acid sequences (figure 11) demonstrated a common frameshift near the carboxyl end, not specific for brain or testes and potentially indicating two distinct protein variants, characterized as group 1 (along with *mys*-1) and group 2. To determine the potential of each expressed

Table 2. Trace BLAST of *mys* cDNA sequnces and *P. leucopus mys*-1 to the *P. maniculatus* WGS database. Values represent coverage (C) and highest identity (I) to a minimum of eight hits. Group 1 and Group 2 frameshift variants are shown in figure 11.

Clone	ORF	Trace BLAST	ORF FS Variant
Pl mys-1	complete	100% C 98-99% I	Group 1
Br1	complete	100% C 98-99%I	Group 1
Br2	Nearly complete	100% C 98-99% I	Group 2
Br3	Stop at 392	100% C 98% I	Group 1
Br4	Stop at 396	100% C 98% I	Group 2
Te1	Stop at 396	99% C 97% I	Group 2
Te2	Complete, but unique internal frameshift	100% C 94% I	Group 1
Te4	Complete	100% C 98-99% I	Group 1
Te5	No ORF	100% C 97-98% I	Group 2
Te7	Complete	100% C 98-99% I	Group 2
Te8	Stop 396	100% C 97-98% I	Group 2

sequence serving as a *mys* master gene, we performed a Trace BLAST to the *P. maniculatus* WGS project. We were inclined to choose Te4 as a potential master gene by features such as an ORF spanning the entire clone and the greatest nucleotide and amino acid similarity to *P. leucopus mys*-1. When examining a pattern of high similarity to numerous sequenced fragments in the *P. maniculatus* genome project, clones with a complete or nearly complete ORF had the greatest level of similarity to a large number of sequences (table 2), typically 98-99%. This is the same value obtained by *mys*-1. Although Te2 showed a complete ORF, there was an internal frameshift caused by an insertion and a deletion, and had the lowest identity to individual sequences. Te7 contained a full-length ORF for the sequenced region, but was in the "group 2" frameshift lineage. Therefore, the possibility for two active *mys* subfamilies exist in *P. maniculatus*, which have distinct amino acid sequences differences toward the carboxyl terminus. Interestingly, and consistent with this hypothesis, the consensus of all Br and Te transcripts generated a sequence with only four variants of the 440

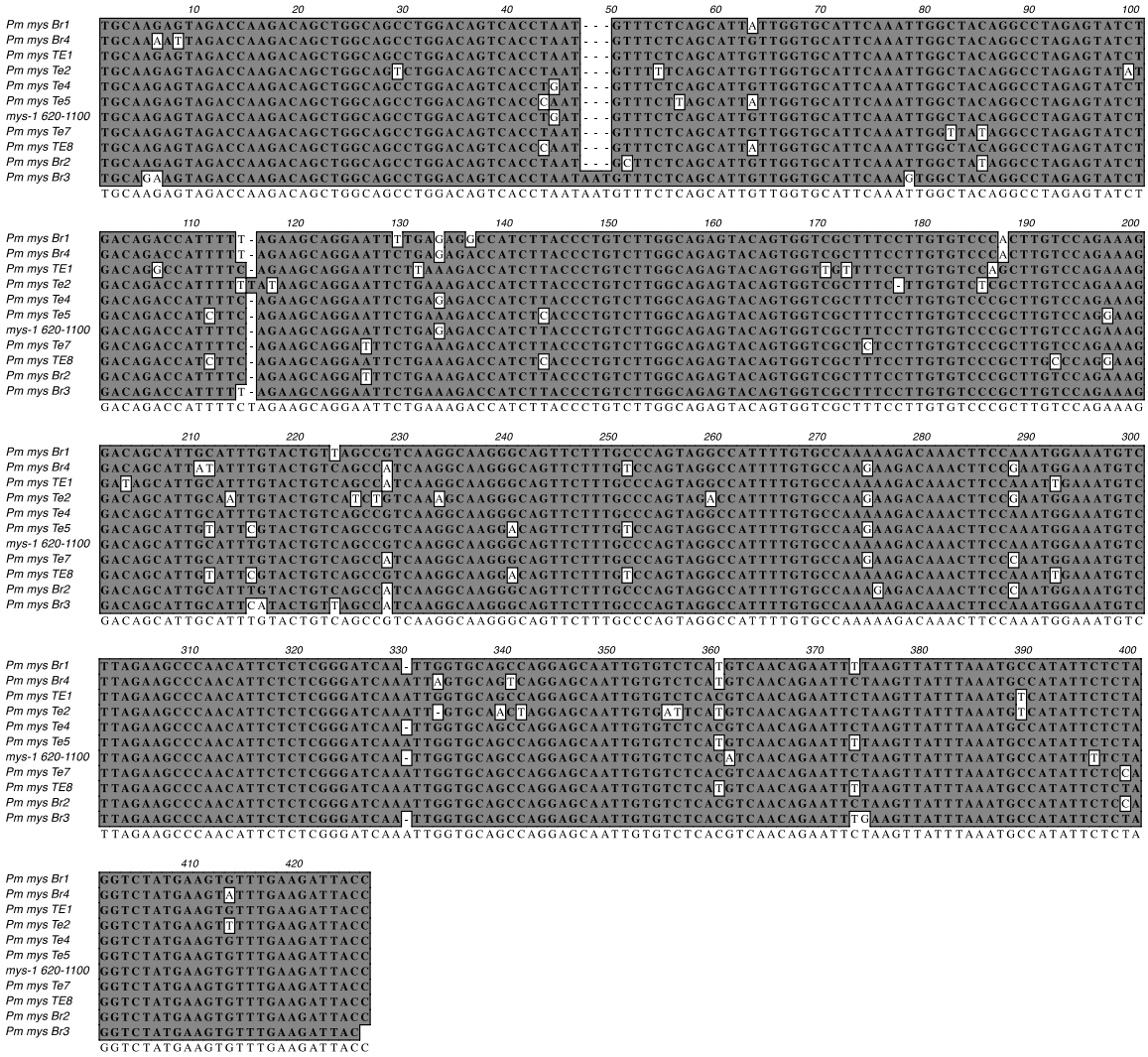


Figure 10. Alignment of nucleotide sequences of *mys*-RT PCR products from deer mouse brain and testes and the corresponding segment of the full-length *mys*-1 element in *P. leucopus*.

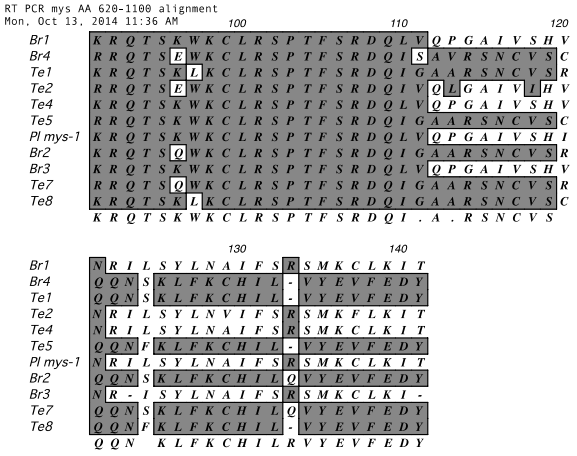


Figure 11. Alignment of translated *mys* RT-PCR products from deer mouse brain and testes, and the genomic *P. leucopus mys*-1 element.

corresponding nucleotides (99% match) from the *P. leucopus mys*-1 element (figure 12). In general, there does not appear to be specificity of *mys* expression in comparing the brain and testes of the deer mouse, or that a single locus is responsible for generating inherited *mys* insertions, unless a locus exists having low expression but high retrotransposition competency.

Discussion

We utilized RT-PCR from tissues of *Peromyscus maniculatus* to discern the expression of SINE

(ID) and LINE (L1 and *mys*) loci to assess variation in somatic versus germ-line tissue, and determine if there is a limited subset of loci being expressed that would potentially exhibit features of a master gene. These characteristics would include having a high conservation to individual integrations, possibly having a higher level of expression in relation to other loci, and for LINEs displaying conservation of an open reading frame.

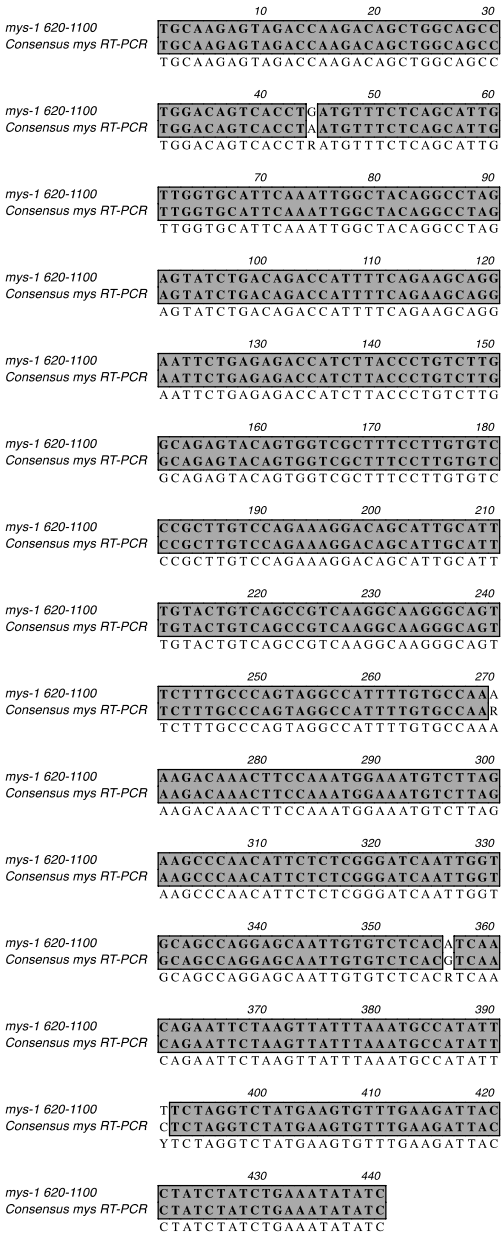


Figure 12. Alignment of consensus sequence of brain and testes *mys* RT-PCR products with an ORF1 segment of a full-length *mys* in *P. leucopus.*

We analyzed the expression of ID elements in the deer mouse, removing potential elements incorporated in mRNA by utilizing size fractionation of cDNA. The BC1 gene was the only observed ID-containing locus expressed in brain based on three analyzed clones, and apparently highly expressed in this tissue, consistent with the finding of BC1 being the primary ID-containing transcript of the rat brain [17]. The one sequence with a single variant could very likely have been the result of *Taq* error. The deer mouse has distinctive variants in BC1 in relation to the ID consensus sequence (IDCS) of rodents, which is the same consensus as ID elements isolated from a *P. maniculatus* genomic DNA library [18], simplifying the differentiation of BC1 to alternative ID-contributing loci. We found BC1 expression occurs in the testes, but based on relative intensities of RT-PCR, apparently not to the same extent. Most of the ID-containing loci expressed in testes are not derived from BC1. The higher proportion of exact matches to IDCS among sequences in a *Peromyscus maniculatus* genomic DNA library (www.hgsc.bcm.edu/other-mammals/peromyscus-genome-project) in relation to exact matches of the ID portion of BC1 supports BC1 as not being the predominant master gene of heritable ID elements. One testes-based transcript exactly matched to an equivalent number of genomic sequences, as did BC1, supporting equivalent potential as a master gene. Since none of the isolated transcripts from testes demonstrated an exact match to IDCS, then there is a possibility that we did not identify the predominant ID master gene, and the level of RNA might not be the limiting factor regarding the efficiency of SINE retrotransposition. This is consistent with the finding of many *Alu* loci representing various *Alu* subfamilies in humans that are transcriptionally active [36, 37, 38]. Additionally, a detailed investigation of *Alu* retrotransposition in cell culture that evaluated different features of *Alu* elements determined retrotransposition competency is likely not the result of a single feature, but rather a combination of sequence features [39].

Previously, stability of BC1 RNA was proposed to play a contributing role in retrotransposition [17] but our finding of three of eleven BC1 transcripts being associated with equivalent or fewer recent genomic integrations than a potential IDCS-equivalent transcript or a transcript exactly matching IDCS, then stability associated with RNA levels might not be the determining factor for retrotransposition competency. RNA stability would be expected to have been associated with a higher proportion of observed transcripts.

We analyzed expression of L1 non-LTR elements as well as the LTR *mys* element found in a limited taxonomic group of rodents [23, 40]. DNA methylation has been proposed to regulate transcription of LINEs [41], particularly in somatic cells as LINE activity along with global DNA demethylation has been associated with tumorigenesis [42]. Additionally L1 retro-transposition events were identified in the genome of the postmortem human brain [11] and were shown to be mostly associated with more recently active L1 subfamilies. Expression of a large number of L1 elements was identified in human lymphoblastoid cell lines, including over 400 full-length [43]. Since full-length L1 elements are also expressed in mouse spermatocytes [28], we analyzed expression of L1 loci in brain and testes, representing somatic and germ-line activity in the deer mouse. Using L1 ORF2-based sequences we observed heterogeneity of L1 sequences derived from RNA, indicating various L1 loci are expressed, with no clear delineation between expressed brain and testes L1 elements. However, seven of eight analyzed clones exhibited an ORF for the entire length of the sequence. In the hunt for a master gene, two clones, one from the brain and one from the testes translated into identical amino acid sequences exactly matching to the consensus sequence. One testes clone also exactly matched 112 segments from the *P. maniculatus* genome project, along with numerous 99% matches supporting this transcript being derived from a candidate master gene.

Contrasting amino acid sequences corresponding to recently active human, mouse and expressed deer mouse L1 elements, we identified 49 of 84 amino acids conserved among the three divergent species suggesting highly selected and functionally important amino acids. Variation at specific sites in L1 ORF2 has been shown to impact retrotransposition efficiency [44]. Additionally, although the generation of chimeric human and mouse L1's by switching ORF1 and ORF2 domains had been known to have little impact on both L1 and *Alu* retrotransposition, substituting a portion of the mouse cys domain into human ORF2P significantly reduces L1 and *Alu* mobilization [45]. Notably, our analyzed amino acid sequences correspond to the end of the RT domain and beginning of the cys-1 domain [45] and we find only three variants, among the last 22 amino acids, between mouse, human and the deer mouse consensus sequence. The 86.4% identity we found in this region among these divergent species is consistent with the importance of the cys domains and selection for efficient retrotransposition.

LTR endogenous retroviral sequences have been inactive in the human genome [30], *i.e.* not generating new copies, in contrast to the *mys* LTR ERV in which recent activity has been reported in the *P. maniculatus-P. leucopus* species complex [23]. However, studies have supported the re-activation of the HERV-K element in human melanoma cells [19, 42]. LTR-ERVs including MaLR, IAP, and Etn have been actively integrating in the genome of *M. musculus*, with the latter two elements accounting for 15% of all spontaneous mouse mutants. IAP transcripts have been identified in somatic tissue in an age-dependent manner [2] and most notably in testes as determined using reporter constructs in transgenic mice [30]. Here we analyzed expression of the *mys* element in germ-line and somatic tissue in the deer mouse. We found the element to be expressed in brain, liver and testes. Upon sequencing individual clones, our data support

the expression of several *mys* loci in brain and testes with a minority potentially serving as master genes. Since demethylation of ERVs has been associated with ERV activity [46] it is possible that this mechanism is involved in *mys* activation based on the highly similar cDNA sequences and genomic sequences suggesting recent activity.

We utilized the testes for analyzing expressed retrotransposon loci, as these would be potentially heritable. Previous findings of full-length L1 RNA and ORF1p (protein) expression during spermatogenesis [28] and the presumptive ID master gene (BC1) expression in testes [17] support heritable male germ-line integrations events. However, our findings suggest BC1 is not the predominant master gene of ID elements. Transcripts from a variety of loci were identified for all three retrotransposon types and either a very limited number of master genes resembling the consensus are actively retrotransposing or there are a larger number of source genes derived from a single ancestral gene that when combined will resemble the consensus sequence. Our findings are consistent with the activity of more than one master gene for each SINE and LINE, and these master genes are expressed in both germ-line and somatic cells. There may, however, be a predominantly active L1 source gene in *P. maniculatus* based on the large number of exact genomic matches to one of the testes-derived transcripts.

Acknowledgement

This work was funded in part by an Eastern Michigan University Faculty Research Fellowship to DHK.

Reference

1. Lander ES, Linton LM, Birren B, *et al*. 2001. Initial sequencing and analysis of the human genome. *Nature*. *409*: 860-921.
2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, *et al*. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*. *420(*6915): 520-62.
3. Hanks DC and Kazazian HH. 2012. Active human retrotransposons: Variation and disease. *Curr Opin Genet Dev*. *22*(3): 191-203.
4. Maxwell PH. 2014. Consequences of ongoing retrotransposition in mammalian genomes. *Adv Genet. 2014*(4): 129-142.
5. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM. 2008. LINE1 ORF1 protein enhances *Alu* SINE retrotransposition. *Gene. 419*: 1-6
6. Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*. *332*: 164-66.
7. Kass DH, Raynor ME, Williams TM. 2000. Evolutionary history of B1 retroposons in the genus *Mus*. *J Mol Evo. 51*: 256-64.
8. Brosius J and Tiedge H. 2001. Neuronal BC1 RNA: Intracellular transport and activity-dependent modulation. *Results Probl Cell Differ*. *34*: 129-138
9. de Souza FS, Franchini LF, Rubinstein M. 2013. Exaptation of transposable elements into novel *cis*-regulatory elements: Is the evidence always strong? *Mol Biol Evol*. *30*: 1239-1251.
10. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell. 87*: 917-27.
11. Baillie JK, Barnett MW, Upton KR, *et al.* 2011. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature. 479*: 53
12. Slotkin RK and Martienssen R. 2007. Transposable elements and epigenetic regulation of the genome. Nat Rev Genet. 8(4): 272-85.
13. Whitelaw E and Martin DI. 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nature Genet. 27*: 361-365.
14. Tiedge H, Jr RT, Weinstock PH, Aranico O, Brosius J. 1991. Dendritic location of neural BC1 RNA. *Proc Natl Acad Sci U S A*. *88*: 2093-2097.
15. Deininger PL, Batzer MA, Hutchison III CA, Edgell MH. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet. 8*: *307*-311.
16. Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P. 1994. Rodent BC1 RNA gene as a master genefor ID element amplification. *Proc Natl Acad Sci USA*. 91(9): 3607-3611.
17. Kim J, Kass DH, Deininger PL. 1995. Transcription and processing of the rodent ID repeat family in germline and somatic cells. *Nucl Acids Res*. 23 (12): 2245-2251.
18. Kass DH, Kim J, Deininger PL. 1996. Sporadic amplification of ID elements in rodents. *J Mol Evol. 42*: *7*-14.
19. Katoh I and Kurata S. 2013. Association of endogenous retroviruses and long terminal repeats with human disorders. *Front Oncol. 3*: 234.
20. Dewannieux M and Heidmann T. 2005. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J Mol Bio. 349*: 241-247.
21. Ishihara H, Tanaka I, Wan H, Nojima K, Yoshida K. 2004. Retrotransposition of limited deletion type of intracisternal A-particle elements in the myeloid leukemia cells of C3H/He Mice. *J Radiat Res*. *45*: 25-32.
22. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de legemaat LN, Mager DL. 2006. Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLOS Genetics*. *2:* 1.
23. Lee RN, Jaskula JC, van de Bussche RA, Baker RJ, Wichman HA.1996. Retrotransposon *Mys* was active during evolution

of the *Peromyscus leucopus-maniculatus* complex. *J Mol Evol. 42*: 44-51.
24. Sawby R and Wichman HA. 1997. Analysis of orthologous retrovirus-like elements in the white-footed mouse, *Peromyscus leucopus*. *J Mol Evol. 44*:74-80.
25. Dewannieux M and Heidmann T. 2013. Endogenous retroviruses: acquisition, amplification and taming of genome invaders. *Curr Opin Virol. 3*: 646-656.
26. Levin HL and Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet*. *12*: 615-627.
27. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH Jr. 2009. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev*. *23*: 1303-1312.
28. Branciforte D and Martin SL. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol.14*: 2584-2592.
29. Schmid C and Maraia R. 1992. Transcriptional regulation and transpositional selection of active SINE sequences. *Current Opinion Genet Dev. 2*: 874-82.
30. Dupressoir A, Heidmann A, Ura C.1996. Germ line-specific expression of intracisternal A-Particle retrotransposons in transgenic mice. *Mol Cell Biol. 16*: 4495-503.
31. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH Jr. 2002. A mouse model of human L1 retrotransposition. *Nat Genet. 32*: 655-660.
32. Casavant NC, Sherman AN, Wichman HA. 1997. Two persistent LINE1 lineages in *Peromyscus* have unequal rates of evolution. *Genetics*. *142*: 1289-1298.
33. Altschul SF, Gisg W, Miller W, *et* al. 1990. Basic local alignment search tool. *J Mol Bio*. 215: 403-410.
34. DeBerardinis RJ, Goodier JL, Ostertag EM, Kazazian HH Jr.1998. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet*. *20*: 288-290.
35. Hohjoh H, Minakami R, Sakaki Y. 1990. Selective cloning and sequence analysis of the human (LINE-1) sequences which transposed in the relatively recent past. *Nucleic Acids Res*. 18: 4099-4104.
36. Sinnett D, Richer C, Deragon JM, Labuda D. 1992. Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. *J. Mol. Biol*. 226:689–706.
37. Liu W and Schmid CW. 1993. Proposed roles for DNA methylation in *Alu* transcriptional repression and mutational inactivation. *Nucleic Acids Res*. *21*: 1351-1359.
38. Shaikh TH, Roy AM, Kim J, Batzer MA, Deininger PL.1997. cDNAs derived from primary and small cytoplasmic *Alu* (sc*Alu*). *J Mol Biol. 271*: 222-234.
39. Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. 2009. Diverse *cis* factors controlling *Alu* retrotransposition: What causes *Alu* elements to die? *Genome Res*. 19:545-555
40. Wichman HA, Potter SS, Pine DS. 1985. *Mys*, a family of mammalian transposable elements isolated by phylogenetic screening. *Nature*. *317*: 77-81.
41. Yu F, Zingler N, Schumann G, Strätling WH. 2001. Methyl-CpG-binding protein 2 represses LINE1 expression and retrotransposition but not *Alu* transcription. *Nucleic Acids Res. 29*: 4493-5401.
42. Stengel S, Fiebig U, Kurth R, Denner J. 2010. Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. *Genes Chromosomes Cancer*. *49*: 401-411.
43. Rangwala SH, Zhang L, Jr HH. 2009. Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol. 10*: 100.
44. Lutz SM, Vincent BJ, Kazazian HH, Batzer MA, Moran JV. 2003. Allelic heterogeneity in LINE1 retrotransposition activity. *Am J Hum Genet*. *73*: 1431-1437.
45. Wagstaff B, Kroutter E, Derbes R, Belancio VP, Roy-Engel AM. 2013. Molecular Reconstruction of extinct LINE-1 elements and their interaction with nonautonomous elements. *Mol Biol Evol*. *30:* 88-99.
46. Gotzinger N, Sauter M, Roemer K, Mueller-Lantzsch N. 1996. Regulation of human endogenous retrovirus-K Gag expression in teratocarcinoma cell lines and human tumors. *J. Gen. Virol*. 77**:** 2983-2990.