Genetic and Epigenetic Impacts of L1 Retrotransposition in Mouse and Man

THESIS SYNOPSIS

Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

By

MANOJ KANNAN
1996PH29707P

Under the Supervision of
Dr. DAVID E. SYMER

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
PILANI (RAJASTHAN)

2015

1.INTRODUCTION

An interesting fact the Human Genome Sequencing Project revealed is that just around two percent of the genome consists of genes, while the rest of it consists of repetitive sequences and a group of diverse elements called transposons, which are mobile genetic elements (Lander et al., 2001). This was also found to be by and large true in case of other mammalians genomes as well (Waterston et al., 2002). Transposons have come a long way from being called “selfish DNA” or “parasite”, to being recognized as “an important player in the mammalian genomes”, playing roles in altering the genome landscape and evolution and in normal biological processes such as bringing about phenotype variation and cellular diversity, as well as abnormalities such as cancers. Approximately half of each sequenced mammalian genome is comprised of various classes of transposable elements (TEs), mobilized by different mechanisms and accumulated over evolutionary time (Akagi et al., 2013; Lander et al., 2001; Levin and Moran, 2011; Waterston et al., 2002).

Long interspersed elements (LINEs, L1s) constitute a major class of mammalian retrotransposons, comprising ~19% and 21% of the mouse and human genomes, respectively It has been hypothesized that approximately half of the mammalian genome has resulted from L1-mediated mobilization. Full-length L1s (of about six kilo base pairs in humans, and seven kilo base pairs in mouse) contain an internal sense-stranded promoter in the 5' untranslated region (UTR), two open reading frames (ORF1 and ORF2) and a 3' UTR with a polyadenylated tail. The 5' UTR of human L1 also contains an antisense promoter, with about 10% activity as the sense-strand promoter (Speek, 2001), while there has been evidence for the initiation of antisense transcription from the ORF1 region of mouse L1 (Zemojtel et al., 2007). While L1 ORF1 encodes a nucleic acid-binding chaperone protein, ORF2 encodes an endonuclease, reverse transcriptase and a zinc finger-like protein. Both ORFs are required for autonomous retrotransposition.

Thousands of full-length elements in three young L1 subfamilies (T_F, G_F and A) reside in the mouse genome. The mouse L1 subfamilies are defined by differences in their 5' UTR monomeric repeats. ORF2 contains the fewest nucleotide variants, whereas the 3' UTR has the most. Members of each subfamily have integrated into the mouse genome after the evolutionary split between rat and mouse. Many L1 T_F, G_F and A integrants are polymorphic, reflecting recent ongoing retrotransposition.

Ongoing movement of endogenous L1 retrotransposons has resulted in several forms of genomic structural variation including insertion polymorphisms, deletion of larger genomic

fragments, exon shuffling, incorporation into transcription units through insertional mutagenesis or exaptation, and probably chromosomal translocations and inversions (Akagi et al., 2008; Gilbert et al., 2005; Kazazian et al., 1988; Moran et al., 1999; Symer et al., 2002). Until recently, L1 mobilization was thought to occur in germline cells or in early embryogenesis. However, recent work has established that L1 retrotransposons, along with other classes of mobile genetic elements, also can move actively in somatic cells, including several human cancers (Baillie et al., 2011; Coufal et al., 2009; Evrony et al., 2012; Iskow et al., 2010; Muotri et al., 2005)

Some integrants can initiate or disrupt cellular transcripts by introducing new promoters, splice sites, polyadenylation signals, and A/T-rich sequences in or near genes while in other cases the mechansisms underlying transcriptional disruption remain unclear. In addition, various classes of retrotransposon integrants can disrupt transcripts at a distance (Druker et al., 2004; Kaer et al., 2011; Li et al., 2012). Several copies of the L1 antisense promoters strewn about the genome has also resulted in widespread generation of transcripts originating from within these elements, both in human (Nigumann et al., 2002) and mouse (Akagi et al., 2008; Zemojtel et al., 2007), thus expanding the transcriptome greatly.

To define the genetic consequences of de novo L1 retrotransposition, several donor constructs exist today for researchers to track the mobilization of new insertions in transfected cells. The fist cell culture model was developed in 1996 by John V. Moran from the Kazazian lab (Moran et al., 1996). The system utilized the system used by Garfinkel and his colleagues to study yeast Ty1 element (Curcio and Garfinkel, 1991). This plasmid-based system for tracking L1 retrotransposition and movement involved placing in the 3' UTR region of the element, a reporter gene (antibiotic resistance), driven by its own strong promoter but interrupted by an artificial intron (AI), in the antisense orientation with respect to the L1. This mechanism ensures the reporter gene can be expressed only when the L1 undergoes transcription, retrotransposition and integration of the cDNA copy of the entire cassette into the genome of the cell.

Since a majority of full-length intragenic human L1s are oriented antisense (AS) to the ORFs of flanking genes, resulting AS L1 retrotransposon-initiated fusion transcripts (RIFTs) frequently include downstream spliced exons expressed in the canonical sense orientation. An antisense promoter (ASP) in L1Hs was first reported in 2001 (Speek, 2001). Studies have also shown the effect of RIFTs in influencing the expression neighbouring genes in humans.

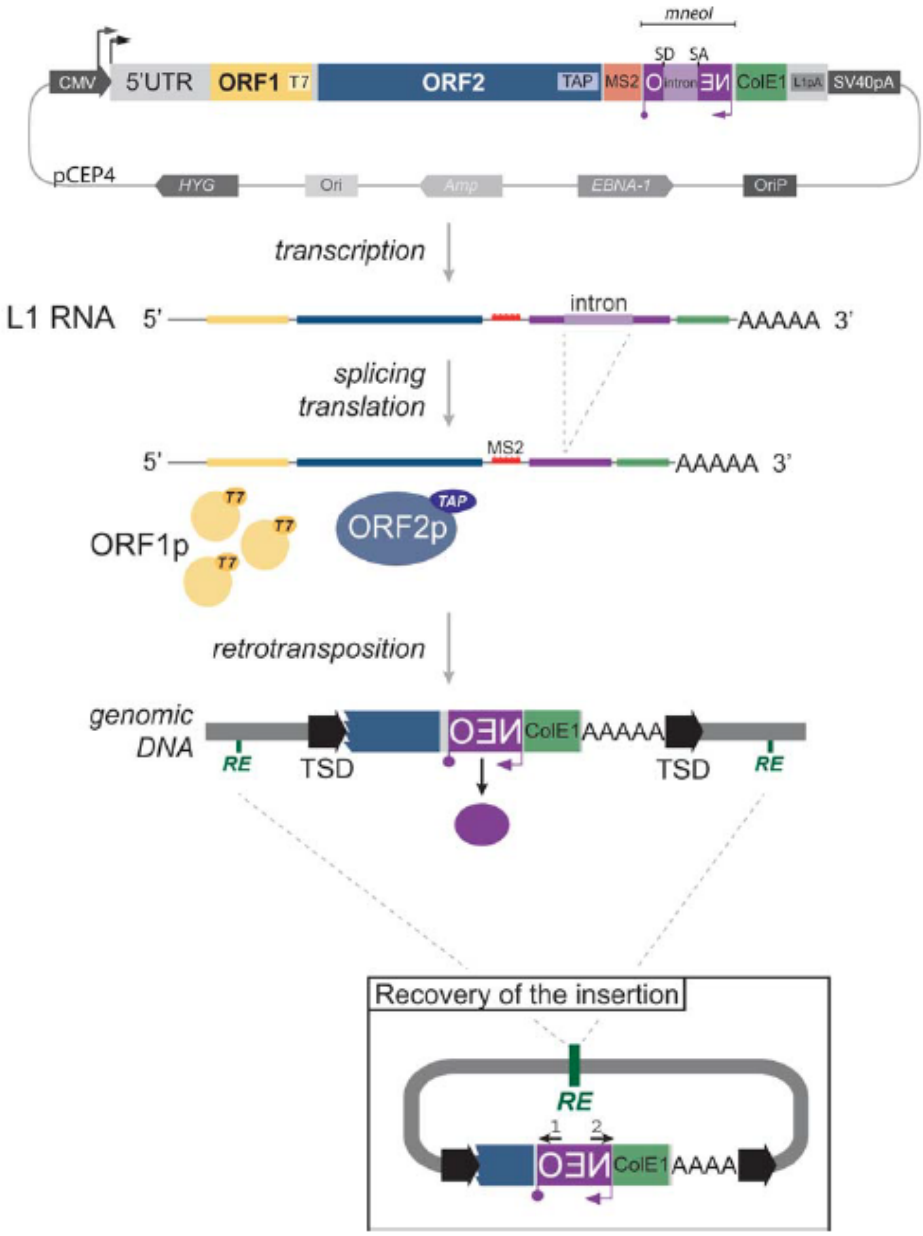


Figure 1 Engineered Long INterspersed Element (LINE-1) structure and cell-based strategies to study retrotransposition.

The widely used donor plasmid (expression vector) in LINE-1 retrotransposition assays consists of a retrotransposition-competent L1 subcloned into pCEP4 (flanked by a CMV promoter and an SV40 polyadenylation signal). The pCEP4 vector is an episomal plasmid encodes a protein (Epstein Barr virus nuclear antigen, EBNA-1) and cis-acting (OriP) sequences necessary for replication in mammalian cells; it also has a hygromycin resistance gene (HYG) that allows for the selection of mammalian cells containing the vector, as well as a bacterial origin of replication (Ori) and an antibiotic selection marker (Ampicillin, Amp) for plasmid propagation in bacteria.

The mneoI reporter cassette, located in the LINE-1 3′ UTR, contains the neomycin phosphotransferase gene (NEO, purple box, with its own promoter and polyadenylation signals, purple arrow and lollipop, respectively) in the opposite transcriptional orientation of L1 transcription. The reporter gene is interrupted by an intron (light purple box) with splice donor (SD) and splice acceptor (SA) sites in the same transcriptional orientation as the L1. This arrangement of the reporter cassette ensures that the reporter gene will only be expressed after a successful round

of retrotransposition. The addition of the ColE1 bacterial origin of replication (recovery of the insertion panel, green box) to a modified version of the mneoI reporter cassette allows the recovery from cultured cell genomic DNA of engineered LINE-1 retrotransposition events as autonomously replicating plasmids in Escherichia coli. The insertions also can be characterized by inverse polymerase chain reaction using divergent oligonucleotide primers (recovery of the insertion panel, black arrows: 1 and 2) that anneal to the reporter gene.

Other useful components of the system include: epitope tags (T7-tag in C-terminus of ORF1, yellow box, and TAP-tag in C-terminus of ORF2, blue box) for immunoprecipitation and detection of ORF1 protein by western blotting and immunofluorescence, and MS2 coat protein (orange box) in the 3′ UTR of LINE-1 for detecting the cellular localization of LINE-1 RNA by FISH.

[Figure reproduced and legend adapted from (Richardson et al., 2015), after obtaining copyright permission from ASM Press]

Much less is known about the control of L1 elements in mouse cells. Both sense and AS transcripts mapping to the 5' end of full-length mouse L1 elements are expressed in mouse ES cells. Mouse chimeric transcripts containing AS L1 sequences also have been identified (Akagi et al., 2008; Zemojtel et al., 2007). These results suggest that mouse L1 elements also may contain one or more antisense promoters (ASP). In fact, Zemojtel et. al's work had indicated that an ASP is likely to be present in ORF1 region of mouse L1, based on analysis of exonization events in mouse cDNA libraries. However, despite identification of AS L1 RIFTs in mouse testis and of sense and AS transcripts in mouse ES cells, a putative AS promoter had not been experimentally validated. Moreover, neither its activity in other tissues nor its possible biological roles have been described. **Chapter 3** of my thesis describes the work done to establish experimentally the existence of an antisense promoter in mouse L1, whose activity is driven by RNA polymerase II, and how it self-regulates the retrotransposition of L1 in a Dicer-independent manner. Evidence for abundant antisense transcription initiated by this promoter is also being shown. L1 mobilization is kept under control by various cellular mechanisms, including genetic and epigenetic mechanisms. The major and most prominent suppression of LINE-1 elements in several cell types is done through two primary mechanisms – epigenetic control, and small RNA-based control. Cytosine methylation is a key epigenetic regulatory mark that is localized predominantly within endogenous L1 retrotransposons and other transposable elements in mammalian genomes. Epigenetic transcriptional silencing based on histone modifications, of a marked LINE L1 in human embryonal carcinoma (EC) cells was reported (Garcia-Perez et al., 2010). **Chapter 4** describes how varying epigenetic marks (cytosine methylation and histone tail modifications) are deposited on *de novo* L1, depending on the cellular and developmental contexts, shown using both cell culture (cancer cell line, mouse ES cells) and mouse models.

2. SPECIFIC AIMS

We have seen why regulation of L1 activity is necessary, and how the cell accomplishes this. Our knowledge on both these areas is still not comprehensive. As part of the doctoral work presented here, I set out to study both genetic and epigenetic impacts/influence of LINE L1 transposition in the mammalian genome, with the following specific aims:

***Specific Aim #1*: To identify and characterize putative antisense promoter(s) located in mouse LINE L1, and to understand its impact on the retrotransposition of L1.**

This aim tests the hypothesis that there would exist within the ORF1 region of mouse L1 transposon, one or more sequences that would possess promoter activity and located antisense to the usual promoter of L1. After identifying the exact location of the antisense promoter (ASP), we will compare its strength with that of the sense promoter, and check the type of RNA polymerase binding to it. Then, we will check the effect of this ASP on retrotransposition of L1 and attempt to elucidate a mechanism for a possible decrease in retrotransposition owing to antisense RNA formation.

***Specific Aim #2*: To compare the epigenetic status of L1 reporter silencing in cancer cell lines with that in other cellular contexts such as embryonic stem cells and *in vivo* mouse tissues.**

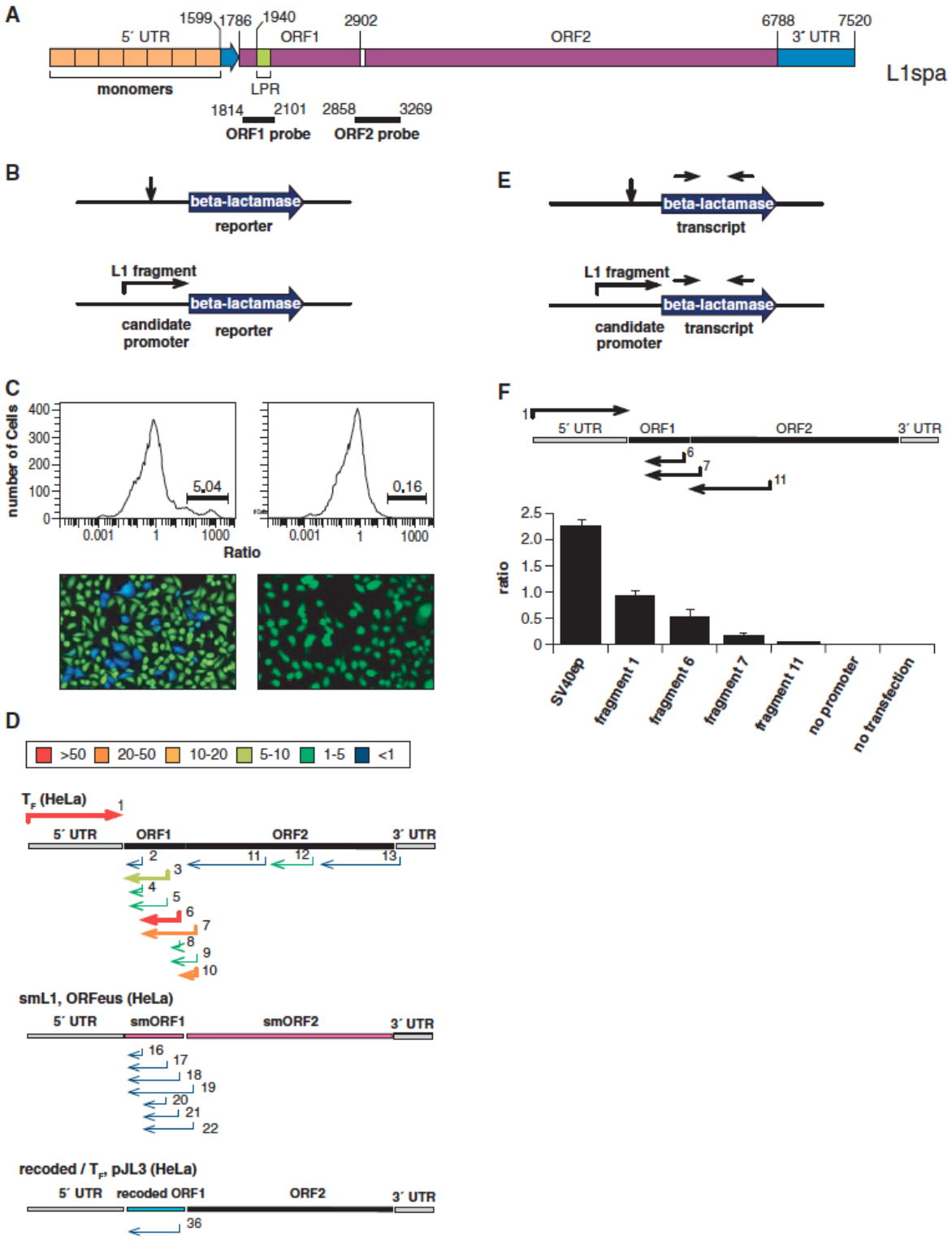
This aim tests the hypothesis that control of L1 retrotransposition in the cultured cancer cell lines (where it can undergo retrotransposition) would be epigenetic, as is the case with other cellular contexts where L1 is thought to be mobilized (e.g., ES cells, early developmental stages). We will assay the endogenous levels of L1 transcripts in various tissues of mouse, ES cells, etc., identify the most "permissive" ones, and study the effect of how various epigenetic factors including DNA methyltransferases, histone deacetylases and drugs affect the efficiency of a marked, highly efficient L1 and/or marker driven by endogenous L1 promoters.

While Chapter 3 describes the work done to address the first specific aim, Chapter 4 describes the work done to address the second specific aim.

3. AN ANTISENSE PROMOTER IN MOUSE L1 RETROTRANSPOSON

To characterize the putative AS promoter experimentally, we amplified 36 candidate promoter fragments spanning various regions of the L1 element and cloned each of them directionally upstream of a plasmid that contained TEM1 -lactamase reporter gene (Zlokarnik et al., 1998) lacking a promoter. The fragments, were derived from mouse L1 subfamilies TF, GF, A and FIII; fully synthetic synonymously recoded smL1 (more recently called ORFeus) (Han and Boeke, 2004); and a novel synonymously recoded ORF1 template that we generated with A/T content similar to native elements. As positive controls, a constitutively active SV40 promoter and arrays of sense strand L1 5' UTR monomers from TF and GF elements were engineered upstream of the TEM1 -lactamase reporter gene. As a negative control, no fragment was inserted upstream of the gene. To assay promoter activities of these fragments, we transfected resulting constructs individually into cultured mouse or human cell lines (CRL-2196 and HeLa). The highest level of AS promoter activity was found in L1 TF AS nucleotides 2823–2125, mapped as per L1spa coordinates (**Error! Reference source not found.**A). Various L1 subfamily members displayed distinct AS promoter activities, i.e. TF (~40% of positive control, i.e. L1 TF 5' UTR monomers in sense orientation) >> GF ≈ A (~10% of control) > F (~5% of control).

We used several independent experimental methods to identify AS L1 RIFTs. These included screens of phage cDNA libraries, RT-PCR followed by cloning and sequencing, bioinformatics surveys of transcript sequence databases, Northern blots and a novel RIFT assay using RT-PCR followed by exon microarray hybridization. These findings clearly established that many diverse RIFTs were expressed from AS promoters located in L1 ORF1 *in vivo*. Many of the AS L1 RIFTs identified by bioinformatics analysis were found in testis and embryonic cells at certain developmental stages, again suggesting a high level of tissue specificity. This search identified over 80 EST clones with AS alignment ~300 nt and over 90% identity with L1 at their 5' ends, of which 15 were full-length RIKEN cDNAs. In some cases, 3' paired ends of other EST clones were identified from the EMBL/EBI database using 5' clone IDs; 57 clones were sequenced from both ends.

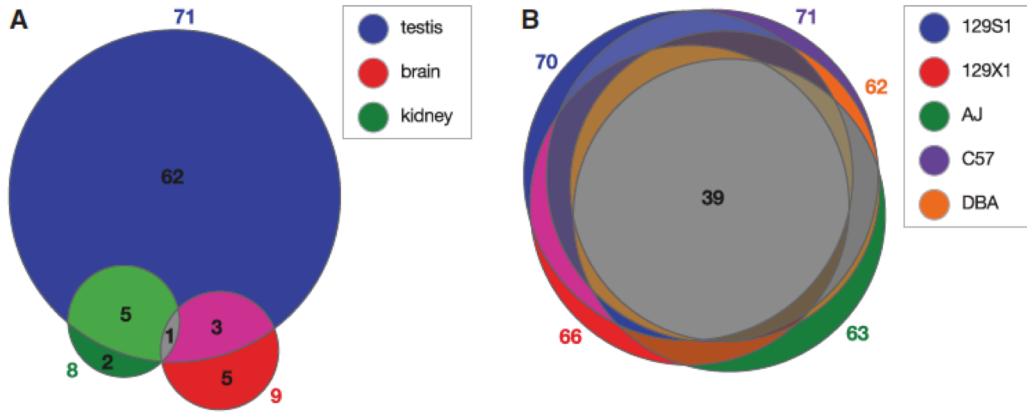


Mapping an active AS promoter within L1 ORF1.

(A) Schematic representation of an L1 T_F subfamily retrotransposon, L1spa, with coordinates indicated as used throughout this article. L1spa was identified in GenBank accession no. AF016099. Shown below the L1 schematic are the probes for phage cDNA library hybridization against ORF2 (2858–3269 nt) and ORF1 (1814–2101 nt). **(B)** Various DNA fragments were directionally engineered upstream of a promoter-less reporter gene, i.e. TEM1 β-lactamase. **(C)** Linearized DNAs containing various candidate promoter-reporter cassettes were transfected into

HeLa cells. Functional beta-lactamase protein expression was measured by staining cells with CCF2-AM, whose fluorescence emission shifts from green to blue on increased enzymatic cleavage. Cells expressing (left) or not expressing (right) β-lactamase were evaluated both by flow cytometry (top), which measured quantitative blue/green emission ratios (Knapp et al., 2003), and by fluorescence microscopy (bottom). **(D)** Fragments derived from various L1 positions and subclasses were numbered and directionally oriented as indicated. Their promoter strengths were assayed as described above. Key: colors and thicknesses indicate promoter activity scores for each fragment assayed. The highest scores (>50, red, thick line) indicate strongest promoter activities. **(E)** TEM1 transcript levels were measured using qRT-PCR (arrows: primer binding sites) to assess the candidate fragments' promoter activities. **(F)** The ratio of TEM1 β-lactamase to β-actin transcript concentrations was calculated (y-axis) after correction for amplification of contaminating plasmid or genomic DNA. As a positive control, SV40 early promoter was engineered upstream of the TEM1 reporter, and as negative controls, no promoter was included or no plasmid was transfected. The AS L1 promoter activity (fragment 6) is half that of the sense-stranded mouse L1 5' UTR promoter (fragment 1). Fragments are numbered as in (D).

To identify genes whose expression levels may be affected by AS L1 RIFTs, we probed Affymetrix mouse exon microarrays conventionally with total RNAs. We developed a novel assay using the arrays to screen specifically for AS L1 RIFTs that include downstream exons. In this "RIFT assay" technique, we prepared cDNAs from several tissues and mouse lineages by RT-PCR, using an AS L1-specific primer paired with an oligo-d(T) primer. Both assays, i.e. the RIFT assay and conventional expression profiling using exon microarrays, confirmed the expression of an AS L1 RIFT at the gene Arhgap15, initially found by screening a testis cDNA library. The RIFT assay also showed that distinct AS L1 RIFTs, although expressed in various tissues, were most abundantly expressed in testis. Several other RIFTs were identified in brain and kidney.



Comparison of AS L1 RIFTs expressed in various mouse tissues and strains.

Distinct AS L1 RIFTs were counted in Venn diagrams depicting shared (overlapping) and unique (distinct) RIFTs expressed in different mouse strains and tissues. Numbers indicate unique RIFTs in each group.

(A) AS L1 RIFTs expressed in B6 testis (n=71, blue), brain (n=9, red) and kidney (n=8, green)

(B) AS L1 RIFTs expressed in testis of five mouse strains: 129S1 (n=70, blue), 129X1 (n=66, red), A/J (n=63, green), B6 (n=71, purple) and DBA/2 J (n=62, orange).

We also found that AS L1 transcription limited L1 retrotransposition, as demonstrated both by altered L1 transcript levels (measured by RT-PCR) and mobilization on synonymous recoding of the AS L1 promoter in ORF1 in *cis* and upon overexpression of AS L1 RIFTs in *trans* (based on retrotransposition frequencies measured by cell culture assay). Hybrid L1s, containing either a recoded synonymous ORF1 segment from smL1 with decreased A/T content or a second recoded ORF1 segment with neutral changes in A/T content, exhibited higher rates of retrotransposition than that of native L1spa. Expression of AS L1 transcripts could result in formation of double-stranded RNA molecules that could affect chromatinization and silencing of the L1 template or trigger an interferon response. Such dsRNAs could form substrates for processing to small inhibitory RNAs through Dicer-dependent or -independent mechanisms.

We also found that Dicer played a modest <2-fold role in suppression of endogenous mouse L1 elements, similar to the results reported earlier for human L1s (Yang and Kazazian, 2006), and infer that AS L1 transcripts act mostly independent of Dicer in decreasing L1 expression and retrotransposition. We conclude that mouse L1s encode a built-in mechanism that regulates them and alters expression of neighboring genes.

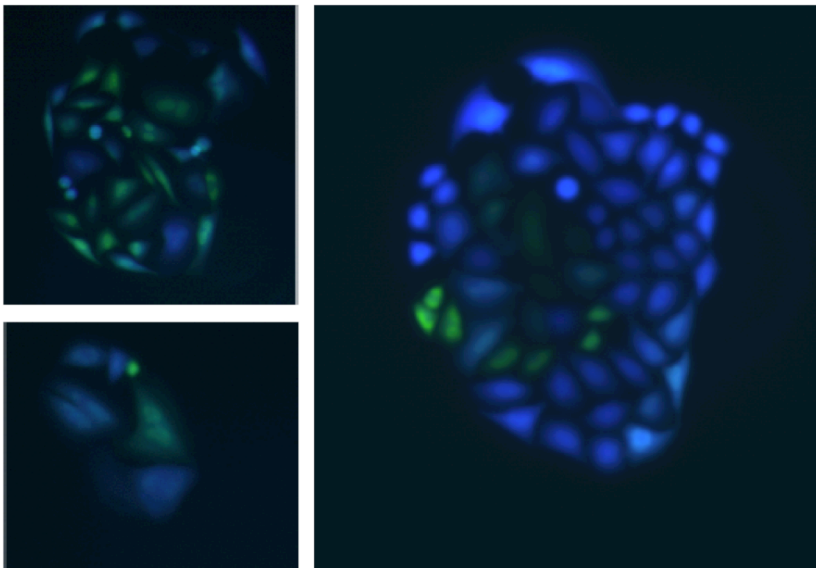
4. DYNAMIC SILENCING OF SOMATIC L1 INTEGRANTS REFLECTS THE DEVELOPMENTAL AND CELLULAR CONTEXTS OF THEIR MOBILIZATION

We launched a marked L1 in cancer cells, investigated the epigenetic status of the resultant integrants, and compared that with the integrant status in embryonic stem cells and adult tissues of mouse. We chose *TEM1* β-lactamase to generate an exquisitely sensitive reporter assay in living cells. Its expression levels can be quantified over a very large dynamic range, extending over four orders of magnitude (Zlokarnik et al., 1998).

Using this reporter system, here we show how different cellular contexts and mechanisms of transposition may impact epigenetic silencing of new insertions. We investigate L1 expression and silencing in cultured human cancer cells, mouse embryonic stem cells, and in tissues of “pseudofounder” transgenic mice (An et al., 2006) and their progeny. Here we describe strikingly different patterns of expression and epigenetic controls at newly mobilized L1 integrants in different contexts. The results revealed distinctive patterns of L1

reporter expression and associated epigenetic marks, depending on the genomic, cellular and developmental contexts of integration.

Upon L1 retrotransposition in cultured human cancer cells, we found that the expression of newly integrated L1 reporters was frequently silenced and the reporter expression was oscillating. This "oscillating state" was heritable, in the sense the daughter cells after mitotic cell divisions tended to display similar levels of reporter expression. This oscillation of gene expression ranged from almost completely silenced to high levels of expression.



Variegation of L1 reporter expression in subcloned populations of HeLa cells
Clones of cells harboring L1 reporter insertions show variegated pattern of expression of the reporter. Each of the three panel shows a colony that presumably arose from a single cell. Blue cells represent those in which retrotransposition has taken place and the reporter is active, while the green ones represent those in which the reporter is being silenced.

Regardless of their expression levels, the genomic L1 reporter integrants remained almost completely unmethylated, even after many cell divisions. L1 reporters were silenced rapidly by histone tail deacetylation, as demonstrated by strong, uniform reactivation of reporter expression upon treatment with diverse HDAC inhibitors. Taken together with the fact that the silencing could spread to the neighboring genes, the observed oscillation in reporter expression can be explained by the stochastic nature of L1 silencing. The oscillating pattern of green and blue cells we observe may well be due to L1-induced rapid changes in chromatin structure (Feng et al., 1999; Henikoff et al., 2004).

In contrast to the oscillating, HDAC-mediated silencing of newly retrotransposed L1 integrants in cultured cancer cells, new L1 insertions in mES cells were silenced by dense de novo CpG methylation. Pluripotent mouse embryonic stem (mES) cells can be viewed as a surrogate for the undifferentiated cells present in early embryos. In addition, we observed that new L1 integrants, present in identical orthologous genomic loci in all differentiated somatic tissues of adult pseudofounder transgenic mice, also were stably silenced by dense cytosine methylation. We surmise that this reflects maintenance of epigenetic controls

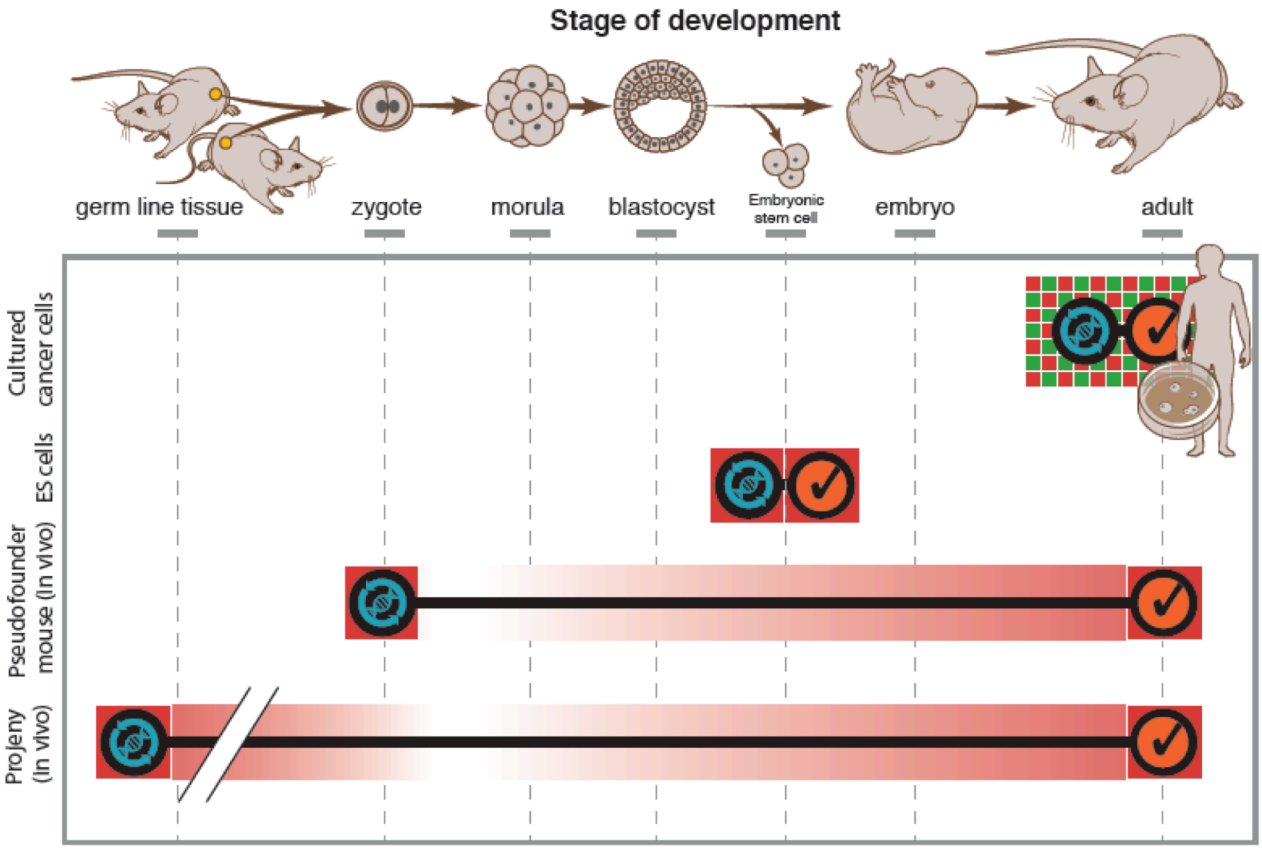
established at the time they were inserted – in early embryogenesis. Thus we conclude that the dense cytosine methylation in mES cells mimics that in early development.

Based on these results, we concluded that the distinctive types of epigenetic regulation established at new L1 insertions observed in mES cells or somatic tissues in vivo, in comparison with those in cultured somatic cells, could be related to the different cellular contexts or stages of differentiation in which L1 mobilization occurred initially. The resulting, distinct forms of silencing established at new transposon integrants in various contexts therefore may have important implications for the expression of L1 elements themselves, for the regulation of other genes neighboring the new insertions, and for chromosomal architecture. We conclude that *de novo* L1 retrotransposition can contribute to significant variability in epigenetic marks established in cellular genomes.

In control experiments, *piggyBac* (*PB*) DNA transposons were used to mobilize the same reporter genes, both in cultured cancer cells and in mES cells. Although our current reporters can be thought of as transgenic insertions in the various cells *in vitro* or *in vivo,* there are several fundamental differences: the target site preference of L1 retrotransposition vs. that of *PB* transposition vs. random integration during transgenesis. We observed two key difference in silencing of *de novo* L1 reporter integrants vs. *PB* integrants. First, we observed minimal or no oscillation of *PB* reporter expression; instead, their expression appeared to be mostly "all or none" in both HeLa and ES cells. Second, the percentage of cells in which *PB* reporter integrants were silenced was much lower than that with L1 integrants. We speculate that these differences between L1 and *PB* reporter expression and silencing may reflect different genomic target site preferences of their mobilization. The differences also could arise based on the different integration mechanisms of these TEs.

In summary, we showed here that the cellular, developmental and genomic contexts of new L1 insertions in somatic tissues are associated with epigenetic silencing marks established at the integrants. We hypothesize that these findings may have important practical implications for our evaluation and understanding of new TE insertions in various biological contexts. For example, they may facilitate the identification of the mobilization timing in an organism. That is, we would expect to find dense cytosine methylation at a new polymorphic L1 insertion that had been mobilized early in development of an individual or was passed through the germ line. This integrant might be detected at a high allele fraction (e.g. 50%, in heterozygosity). By contrast, a somatic L1 polymorphism would be expected to be mosaic and therefore present at a much lower allelic fraction in one tissue and not another, such as in a tumor and not in

matched normal tissue. It would be silenced only by histone deacetylation. Together these features would suggest that its mobilization occurred in differentiated somatic cells incapable of de novo methylation.



A model for L1 integrant silencing in various cellular and developmental contexts

L1 integrants are silenced differently, depending on whether the integration occurred before fertilization, during early development or in adult somatic tissues. In this schematic, the blue symbol represents the event when transposition is thought to have occurred, while the red check mark indicates when we studied the methylation status of those retrotransposed integrants. Both these events are separated by time (indicated by horizontal black lines), in which period the epigenetic marks are erased and reestablished (indicated by intensity of red shading). Red and green checkers indicates variegation of the reporter expression.

We also hypothesize that, in turn, these distinctive epigenetic marks established at new insertions may play important roles in the downstream impacts of new insertions. New insertions occurring early in development may much more significantly disrupt neighboring gene expression, because their allelic fraction would be higher and the silencing imparted at them would be expected to be stronger and more stable. We conclude that such greater disruptive impacts of such earlier integrants would be attributable to with bigger epigenetic effects at transcription, imparted by dense cytosine methylation at such insertions.

5.SUMMARY AND CONCLUSIONS

The raison d'être and the biological impacts of retrotransposons in mammalian genomes, chiefly L1 elements, still are not completely understood. Between 6 and 30% of human and mouse transcripts are initiated from transposable elements. However, the promoters driving such transcriptional activity are mostly unknown. As described in Chapter 3, we experimentally characterized an antisense (AS) promoter in mouse L1 retrotransposons for the first time, oriented antiparallel to the coding strand of L1 open reading frame-1. We found that AS transcription is mediated by RNA polymerase II. Rapid amplification of cDNA ends cloning mapped transcription start sites adjacent to the AS promoter. We identified more than 100 novel fusion transcripts, of which many were conserved across divergent mouse lineages, suggesting conservation of potential functions. To evaluate whether AS L1 transcription could regulate L1 retrotransposition, we replaced portions of native open reading frame-1 in donor elements by synonymously recoded sequences. The resulting L1 elements lacked AS promoter activity and retrotransposed more frequently than endogenous L1s. Overexpression of AS L1 transcripts also reduced L1 retrotransposition. This suppression of retrotransposition was largely independent of Dicer. Our experiments shed new light on how AS fusion transcripts are initiated from endogenous L1 elements across the mouse genome. Such AS transcription can contribute substantially both to natural transcriptional variation and to endogenous regulation of L1 retrotransposition.

Transposons make a huge target for epigenetic regulation. To survey host epigenetic responses to newly transposed insertions in diverse host contexts, as described in Chapter 4, we engineered a very sensitive, "real-time" L1-reporter construct to be used in cell culture assay. We found strikingly different patterns of expression and epigenetic controls established at newly mobilized L1 integrants in somatic cells and tissues including cultured human cancer cells, mouse embryonic stem cells, and in tissues of pseudofounder transgenic mice and their progeny. In cancer cell lines, the new L1 reporter integrants typically were silenced rapidly, but cytosine methylation was absent even after many cell divisions. L1 reporter expression was reversible, oscillated frequently, and was strongly and uniformly reactivated upon treatment with histone deacetylase inhibitors, suggesting that histone deacetylation silences such insertions. By contrast, *de novo* L1 integrants in pluripotent mouse embryonic stem (ES) cells underwent rapid, dense cytosine methylation. Similarly, dense cytosine methylation also was observed at new L1 integrants in several distinct somatic tissues of adult founder mice. We hypothesized that *de novo* methylation marks, established

at the time of transposition in early development, were maintained through development. As controls, reporters also were engineered into *piggyBac*, revealing relatively stable expression upon mobilization in both cultured cancer cells and ES cells. Pre-existing L1 elements in cultured human cancer cells were stably silenced by dense cytosine methylation, whereas their transcription modestly increased when DNA cytosine methylation was experimentally reduced.

My thesis work lends credence to the idea that L1 transposons continue to exert impacts on the host by contributing to the transcriptional repertoire of the cell (Han et al., 2004; Zemojtel et al., 2007), and by serving as targets for epigenetic marks such as cytosine methylation and histone tail modifications at the new spots where they integrate in the genome, corroborating data published earlier (Garcia-Perez et al., 2010; Grandi et al., 2015).

6. REFERENCES

Akagi, K., Li, J., Stephens, R.M., Volfovsky, N., and Symer, D.E. (2008). Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. Genome Res *18*, 869-880.

Akagi, K., Li, J., and Symer, D.E. (2013). How do mammalian transposons induce genetic variation? A conceptual framework. BioEssays *35*, 397-407.

An, W., Han, J.S., Wheelan, S.J., Davis, E.S., Coombes, C.E., Ye, P., Triplett, C., and Boeke, J.D. (2006). Active retrotransposition by a synthetic L1 element in mice. Proc Natl Acad Sci U S A *103*, 18662-18667.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M.*, et al.* (2011). Somatic retrotransposition alters the genetic landscape of the human brain. Nature *479*, 534-537.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., and Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. Nature *460*, 1127-1131.

Curcio, M.J., and Garfinkel, D.J. (1991). Single-step selection for Ty1 element retrotransposition. Proceedings of the National Academy of Sciences of the United States of America *88*, 936-940.

Druker, R., Bruxner, T.J., Lehrbach, N.J., and Whitelaw, E. (2004). Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. Nucleic Acids Res *32*, 5800-5808.

Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A.*, et al.* (2012). Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain. Cell *151*, 483-496.

Feng, Y.Q., Alami, R., and Bouhassira, E.E. (1999). Enhancer-dependent transcriptional oscillations in mouse erythroleukemia cells. Mol Cell Biol *19*, 4907-4917.

Garcia-Perez, J.L., Morell, M., Scheys, J.O., Kulpa, D.A., Morell, S., Carter, C.C., Hammer, G.D., Collins, K.L., O'Shea, K.S., Menendez, P.*, et al.* (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature *466*, 769-773.

Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol *25*, 7780-7795.

Grandi, F.C., Rosser, J.M., Newkirk, S.J., Yin, J., Jiang, X., Xing, Z., Whitmore, L., Bashir, S., Ivics, Z., Izsvak, Z.*, et al.* (2015). Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. Genome research *25*, 1135-1146.

Han, J.S., and Boeke, J.D. (2004). A highly active synthetic mammalian retrotransposon. Nature *429*, 314-318.

Han, J.S., Szak, S.T., and Boeke, J.D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature *429*, 268-274.

Henikoff, S., McKittrick, E., and Ahmad, K. (2004). Epigenetics, histone H3 variants, and the inheritance of chromatin states. Cold Spring Harb Symp Quant Biol *69*, 235-243.

Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. Cell *141*, 1253-1261.

Kaer, K., Branovets, J., Hallikma, A., Nigumann, P., and Speek, M. (2011). Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. PLoS One *6*, e26099.

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.

Knapp, T., Hare, E., Feng, L., Zlokarnik, G., and Negulescu, P. (2003). Detection of beta-lactamase reporter gene expression by flow cytometry. Cytometry A *51*, 68-78.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W*., et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nature reviews Genetics *12*, 615-627.

Li, J., Akagi, K., Hu, Y., Trivett, A.L., Hlynialuk, C.J., Swing, D.A., Volfovsky, N., Morgan, T.C., Golubeva, Y., Stephens, R.M*., et al.* (2012). Mouse endogenous retroviruses can trigger premature transcriptional termination at a distance. Genome research *22*, 870-884.

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

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.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Nigumann, P., Redik, K., Matlik, K., and Speek, M. (2002). Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics *79*, 628-634.

Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Perez, J.L., and Moran, J.V. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr *3*, MDNA3-0061-2014.

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol *21*, 1973-1985.

Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. (2002). Human L1 retrotransposition is associated with genetic instability in vivo. Cell *110*, 327-338.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P*., et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Yang, N., and Kazazian, H.H., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol *13*, 763-771.

Zemojtel, T., Penzkofer, T., Schultz, J., Dandekar, T., Badge, R., and Vingron, M. (2007). Exonization of active mouse L1s: a driver of transcriptome evolution? BMC Genomics *8*, 392.

Zlokarnik, G., Negulescu, P.A., Knapp, T.E., Mere, L., Burres, N., Feng, L., Whitney, M., Roemer, K., and Tsien, R.Y. (1998). Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. Science *279*, 84-88.