Genetic and Epigenetic Impacts of L1 Retrotransposition in Mouse and Man

THESIS

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

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

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

2015

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI (RAJASTHAN)

CERTIFICATE

This is to certify that the thesis entitled "**Genetic and Epigenetic Impacts of L1 Retrotransposition in Mouse and Man**" and submitted by **Mr. Manoj Kannan,** ID No. **1996PH29707P**, for award of Ph.D. degree of the Institute embodies original work done by him under my supervision.

Signature of the Supervisor:

Name: **Dr. DAVID ERIC SYMER, M.D., Ph.D.**

Designation: **Director, Genomics Shared Resource, Human Cancer Genetics Program**
Assistant Professor, Dept. of Molecular Virology, Immunology and Medical Genetics
The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, USA.

Date: 14 December, 2015

In loving dedication to my mentor and well-wisher

Gaurasundara Dasa

Who inspired me to complete my Ph.D. as a service to God

ACKNOWLEDGEMENTS

After more than a decade of embarking, I stand at the verge of completing my journey of obtaining a doctoral degree. Providential help was manifested through numerous people whom I have known in each of the three institutions I worked at: BITS Pilani (India), National Cancer Institute at Frederick (part of National Institutes of Health, USA) and Ohio State University, Columbus (USA). I am grateful for their gestures of kindness and support.

First, I am indebted and thankful to my Ph.D. supervisor, Dr. David Symer, for selecting me to work with him, for providing the best possible doctoral training he could, and for showing indomitable patience, determination and support to help me graduate; his former lab members for providing an environment conducive to learn and do science; and Amudhan Venkateswaran, my friend and senior at BITS, who introduced me to Dr. Symer.

Those at BITS Pilani have been patiently and eagerly waiting since 2008 for my completion of the degree, while extending their cooperation in every possible way. In particular, I thank the current and former heads of the institution – Professors S. Venkateswaran, L.K. Maheshwari, Raghurama G., Ranendra N. Saha, A.K. Sarkar, G. Sundar and B.N. Jain; Deans Drs. Ravi Prakash, S.K. Verma, Ajit Pratap Singh, Hemant Jadhav; and heads of the Department of Biological Sciences at BITS – Drs. Rajesh Mehrotra, Jitendra Panwar, Shibasish Chowdhury, Ashis K. Das, and Prof. S.K. Ray who encouraged me to foray into biology.

I also thank my immediate family for their motivation, and my community of close friends at Pilani - Haripada, Harihara, Lilakishore, Sundar and Shyamahari prabhus, who graciously treated me as part of their own families.

I firmly believe it is Lord Krishna, the Supreme Personality of Godhead, who enabled me to complete this endeavour, despite my obstinacy and recalcitrance. My gratitude will be borne out by how best I utilize this gift in His service.

Manoj Kannan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	artificial intron
AS	antisense
ASP	antisense promoter
CMV	Cytomegalovirus
FISH	Fluorescent in situ hybridization
GFP	Green fluorescent protein
HDAC	Histone deacetylase Inhibitor
Hygro	hygromycin
LINE	Long Interspersed Nuclear Element
Neo	neomycin phosphotransferase
nt	nucleotide
ORF	open reading frame
PB	piggyBac (a transposon)
qRT-PCR	quantitative reverse transcriptase polymerase chain reaction
RIFT	Retrotransposon-Initiated Fusion Transcript
SINE	Short Interspersed Nuclear Element
smL1	synthetic mouse LINE L1 element
SV40	Simian virus 40
TE	transposable element
TsA	Trichostatin-A
TSS	transcription start site
UTR	untranslated region

1.INTRODUCTION

1.1 MAMMALIAN L1 RETROTRANSPOSONS

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). In fact, transposable elements are present in almost every eukaryotic organism whose genome has been sequenced so far (with rare exception such as that of *Ashbya gossypii*, a filamentous fungus) (Huang et al., 2012). Transposons have come a long way from being called "selfish DNA" or "parasite" (Orgel and Crick, 1980; Yoder et al., 1997), to being recognized as "an important player in the mammalian genomes" (Rebollo et al., 2012), playing roles in altering the genome landscape and evolution (Hedges and Batzer, 2005) and in normal biological processes such as bringing about phenotype variation (Akagi et al., 2013) and cellular diversity (Thomas et al., 2012), as well as abnormalities such as cancers (Rodic and Burns, 2013). 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).

Why is the study of transposons significant? For one, their substantive presence in mammalian genomes is a reason compelling enough to study their roles, which, until a decade ago, were largely unknown. Secondly, it may be easily argued that unrestrained movement of these genomic DNA elements may cause havoc, among other things, by interrupting genes or causing chromosomal breakage. There is evidence for both of these: transposition has been associated with human disease as well as genomic instability in cells.

Long interspersed elements (LINEs, L1s) constitute a major class of mammalian retrotransposons, comprising ~19% and 21% of the mouse and human genomes, respectively (Lander et al., 2001; Waterston et al., 2002). It has been hypothesized that approximately half of the mammalian genome has resulted from L1-mediated mobilization (Ostertag and Kazazian, 2001a). 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 (Callahan et al., 2012; Martin et al., 2005), ORF2 encodes an endonuclease (Feng et al., 1996), reverse transcriptase and a zinc finger-like protein (Cost et al., 2002). Both ORFs are required for autonomous retrotransposition (Moran et al., 1996). Very recently, a third open reading frame, named ORF0, located in the 5'-UTR and oriented in the antisense direction, was reported to be present in primate-specific L1 elements (Denli et al., 2015). ORF0 has its own promoter and encodes a protein product that localizes to promyelocytic leukemia (PML)-adjacent bodies. Owing to the presence of splice donor sites, ORF0 is also a potential source of chimeric fusion transcripts with downstream genomic sequences, thereby contributing to L1-mediated transcript diversity.

Thousands of full-length elements in three young L1 subfamilies (T_F, G_F and A) reside in the mouse genome (Jachowicz and Torres-Padilla, 2015). 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 (Sookdeo et al., 2013). 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 (Akagi et al., 2008; Akagi et al., 2010).

The genome sequencing projects placed the number of L1 fragments present in humans to be 516,000, while that in mouse as 600,000. However, work from the Kazazian lab concluded that only about 80–100 L1s in humans (Brouha et al., 2003; DeBerardinis et al., 1998; Goodier et al., 2001) and about 3,200 in mouse (all three families inclusive) remain capable of hopping, i.e., retrotransposition-competent (RC-L1) (Brouha et al., 2003; DeBerardinis et al., 1998; Goodier et al., 2001). The rest of the L1 elements were found to be either truncated from the 5' ends (and hence lacking promoters), or harboring mutations (inversions), thereby rendering them inactive. Among the several thousand L1 copies found interspersed in the genome, only those that remain retrotransposition-competent were thought to be potential mutagens that affect genomic integrity through insertional mutagenesis and generation of double-stranded breaks (DSBs) in the DNA. However, it was recently shown that those L1s truncated in ORF2 (and therefore, retrotransposition-incompetent) but retaining a functional endonuclease domain not only cause cellular toxicity when overexpressed in mammalian cells, but also have the potential to contribute to genomic instability by mobilizing *Alu*s and introducing DSBs (Kines et al., 2014), thereby underscoring the need for cellular controls on L1 mobility.

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 (Figure 1). 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. Thus the new genomic cDNA integrants all had been mobilized and inserted by a retrotransposition-dependent mechanism. Even if the reporter gene is transcribed from its own promoter, such transcripts cannot be successfully expressed due to the presence of the intron that cannot be spliced (being in the antisense orientation with respect to the reporter gene). Although most of the insertions were truncated from their 5' ends, many insertions included intact copies of the spliced reporter gene, and its expression was assayed as a surrogate for retrotransposition.

The reporter genes that have been used for this retrotransposition assay include the antibiotic resistance gene, *neo* (encoding neomycin phosphotransferase) that provides G418 resistance (Moran et al., 1996), enhanced green fluorescent protein (EGFP) (Ostertag et al., 2000), blasticidin deaminase (Goodier et al., 2007),firefly luciferase (Xie et al., 2011), and most recently, by us, the highly sensitive TEM1 β-lactamase (Li et al., 2014). Despite being considerably different from the *in vivo* process of retrotransposition, (Kazazian and Goodier, 2002), the cell culture system for L1 retrotransposition has been extensively used by researchers to study the behavior and biology of L1 retrotransposon (Rangwala and Kazazian, 2009).

Ongoing movement of endogenous L1 retrotransposons has resulted in several forms of genomic structural variation including insertional 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).

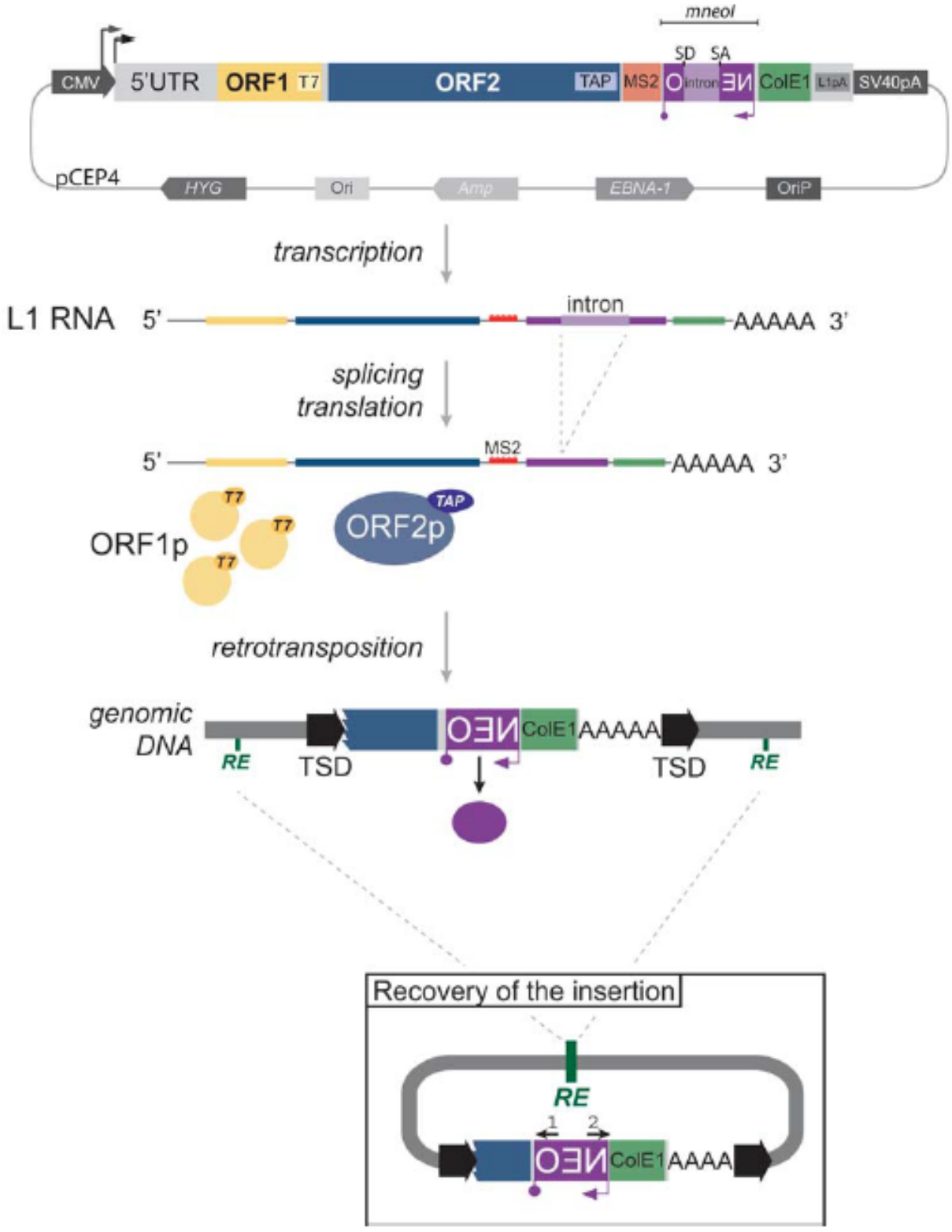


Figure 1 Engineered Long INterspersed Element (LINE-1) structure and cell-based strategies to study retrotransposition.

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]

Until recently, L1 mobilization was thought to occur in germline cells or in early embryogenesis (van den Hurk et al., 2007). 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).

The potential impacts of such mobilization on transcriptional regulation have not been fully characterized (Symer and Bender, 2001). Most of our existing knowledge about gene disruption due to transposon integrants can be attributed to characterization of diseases caused by them in mouse and man (Callinan and Batzer, 2006; Chen et al., 2006; Chen et al., 2005; Kazazian et al., 1988). There are around 100 known disease insertions known till date

(Hancks and Kazazian, 2012). The first reported association of mobile DNA with human disease came from Haig Kazazian's laboratory at John Hopkins University (Kazazian et al., 1988). They found two L1 insertions in the fourteenth exon of factor VIII gene of two individual haemophilia A patients. Later on, from the parents of one of the two patients, they isolated and characterized the active L1 element that caused this insertion (called L1.2) and also reported the reverse transcriptase activity of ORF2 (Dombroski et al., 1991; Mathias et al., 1991).

The first reported case of association of a mobile genetic element with cancer was of an L1 element in colon cancer (Miki et al., 1992). The 3' end of the L1 and a poly-A tract was found inserted in the last exon of the adenomatous polyposis coli (APC) gene, and this event had arisen from a somatic retrotransposition event, as evidenced by the presence of 8-bp target site duplication, a characteristic feature of L1 integration events. Another instance of a somatic L1 insertion was reported in a choromosomal translocation in desmoplastic small round cell tumour (Liu et al., 1997). With the recent advent of technologies that are able to effectively capture the rare somatic insertions that occur in cancers, L1-mediated retrotransposition has been shown to play etiological role in lung tumour (Iskow et al., 2010), colorectal cancer (Solyom et al., 2012), hepatocellular carcinoma (Shukla et al., 2013), and a variety of other cancers (Lee et al., 2012). While these instances establish a clear role for L1s in cancers, whether L1s are the "drivers" or "passengers" in the cancer development process is yet to clearly determined (Rodic and Burns, 2013). Cancer is also a disease of epigenetic deregulation; the connection between L1, epigenetics and cancer is described later in the text.

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 (Belancio et al., 2008; Chen et al., 2006; Han et al., 2004; Perepelitsa-Belancio and Deininger, 2003; Speek, 2001). 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.

1.2 GENOMIC CONTROL OF L1 RETROTRANSPOSITION

In order to maintain genome integrity, L1 elements in the genome are kept under tight and efficient control so that they don't undergo retrotransposition promiscuously. As mentioned earlier, L1 is capable of being mobilized only in the germ cells, during a window of time during early embryonic development, and in neural precursor cells during normal development. Such restricted mobility of L1 and the abundance of pre-existing L1s have precluded direct mechanistic studies *in vivo* on endogenous L1s, but experiments employing marked L1s in cultured cells (HeLa, mouse and human ESCs, embryonal carcinoma cells, and gonadal cell lines such as NTera1), as well as transgenic mice models have generated substantial information on L1 movement and control.

Of the many host factors one may expect to find, only some that are responsible for L1 movement (or repression) have been discovered. One group of positive regulators that enable L1 retrotransposition are transcription factors that bind to the internal promoter located in the 5' UTR, viz., (i) YY1 binding to +21 to +13 on the antisense strand (Becker et al., 1993); (ii) SRY (SOX family) binding to +472 to +477 and +572 to +577 (Tchenio et al., 2000); (iii) RUNX3 binding to +83 to +101 (Yang et al., 2003). Another positive regulator of efficient transposition and RNP formation is the poly (A) binding protein C1 (PABPC1) (Dai et al., 2012).

All other factors that control L1 mobility in the cell are negative regulators, and include (i) proteins belonging to the APOBEC3 (apoprotein B-editing catalytic polypeptide 3) family that inhibit L1 and Alu retrotransposition (Schumann, 2007); (ii) KZNF protein family and its associated repressive complex proteins that inhibit L1s in ES cells (Castro-Diaz et al., 2014); (iii) the Aicardi–Goutières syndrome related proteins Trex1 and SAMHD1 that inhibit L1 retrotransposition in cultured cells (Stetson et al., 2008; Zhao et al., 2013); (iv) the RNA helicase MOV10 that inhibits retrotransposition of human L1 in cell culture (Li et al., 2013); (v) heterogenous nuclear ribonucleoprotein L (hnRNPL) that also inhibits retrotransposition in cultured cells (Peddigari et al., 2013). But 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. Let us look at these briefly.

Quoting Conrad Waddington who coined the term 'epigenetics' in 1942, it is "a branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being." A more modern definition is: *study of the set of phenomena that bring about heritable change in gene expression and function, without actually changing the sequence of nucleotides*. The field of epigenetics, although can be thought of as a part of the

broad discipline of genetics, has evolved as an independent field of research investigation in the last decade or so. While those phenomena associated with the DNA sequence, how they are controlled and the gene expression profiles fall in the field of genetics, those that influence the former, but are not present within the DNA sequence constitute epigenetics. Outstanding examples of such epigenetic phenomena include the occurrence of patterns of methylated cytosine bases in DNA (DNA methylation) and covalent modifications to the histone proteins around which the double helix is wrapped (histone modifications). Other events include RNA interference, X-chromosome inactivation and genomic imprinting (Allis, 2015; Armstrong, 2014).

Cytosine methylation is a key epigenetic regulatory mark that is localized predominantly within existing L1 and other transposable elements (TEs), and is strongly associated with their silencing (Yoder et al., 1997). In fact, a 'genome defense hypothesis' posits that the DNA methylation evolved in mammals primarily for silencing the TEs. These repressive marks are heritable and in general are stably maintained. In normal somatic cells, L1 retrotransposons are heavily methylated at CpG dinucleotides, but hypomethylation of these TEs has been observed in most cancers, potentially resulting in their increased transcription and mobilization (Alves et al., 1996; Florl et al., 1999; Iskow et al., 2010; Menendez et al., 2004; Suter et al., 2004). During early development, there occur two waves of demethylation of the genome that serve as windows for retrotransposition to occur (Bodak et al., 2014). While the histone tail modification controls established at *de novo* TE integrants in mammalian genomes have not been studied until recently (Garcia-Perez et al., 2010), epigenetic marks including cytosine methylation and various histone modifications have been characterized at newly inserted, exogenous retroviruses and transgenes in different experimental systems. Transcriptional gene silencing of newly integrated retroviruses and transgenes may depend on position effects and the differentiation status of host cells. Epigenetic processes controlling L1 activity in the cell are revisited in Section 4.1.

The second type of control of transposon movement is small RNA-based. Existence of three different types of small RNAs – micro RNA (miRNA), endogenous small interfering RNA (endo-siRNA), and Piwi-interacting RNA (piRNA) – are found in mammals, of which two classes have been shown to be involved in L1 regulation. In male gametes, PIWI-piRNA plays a role in repressing L1s (Aravin et al., 2007), while in mouse oocytes both piRNA and siRNA work in conjunction to silencing L1s (Watanabe et al., 2008).

1.3 ORGANIZATION OF THE THESIS

Chapter 3 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.

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. While integrant reporters in cancer cell lines are silenced by histone tail modifications (and de-repressed by inhibiting histone deacetylases) but show almost no methyl marks in the cytosines, the scenario is different in ES cells and mouse tissues: the integrants are heavilty methylated. Use of a sensitive reporter gene – TEM1 β-lactamase, tagged to L1 also revealed the dynamic "switching" or "oscillation" between silent and active states of the reporters in cultured cancer cells when integrated using L1 as the vehicle. No such marked variation is observed when the reporter enters the cell via another DNA delivery vehicle (*piggyBac*), thus revealing how the cell treats L1 integrants differently from others.

This study adds to the existing knowledge about how the LINE L1 retrotransposon influences mammalian transcriptome, and how the mammalian epigenetic machinery controls and differentially treats new insertions of L1.

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

The entire work described in this chapter has been published (Li et al., 2014).

3.1 BACKGROUND OF THE STUDY

The potential biological impacts of endogenous transposable elements (TEs) in human and mouse appear to depend on their genomic context, their sequence structure and other factors (Akagi et al., 2013). Endogenous TEs have been shown to affect neighboring gene expression in various ways (***Figure 2***). For example, they have been reported to initiate a surprising number, between 6 and 30%, of human and mouse transcripts (Faulkner et al., 2009). Since a majority of full-length intragenic human L1s are oriented antisense (AS) to the ORFs of flanking genes (Symer et al., 2002; Szak et al., 2002), resulting AS L1 retrotransposon-initiated fusion transcripts (RIFTs) frequently include downstream spliced exons expressed in the canonical sense orientation. Other human AS L1 RIFTs are noncoding (Conley et al., 2008; Mourier and Willerslev, 2009; Rangwala et al., 2009). Mouse endogenous retroviruses have been shown to disrupt overlapping gene expression (Druker et al., 2004; Li et al., 2012). Human L1s may affect expression of overlapping genes, including the Met proto-oncogene and others (Kaer et al., 2011; Weber et al., 2010).

An antisense promoter (ASP) in L1Hs was first reported in 2001 (Speek, 2001). In this study the presence of L1 RIFTs, initiated from an ASP located within the 5' UTR of L1Hs, at around nt 500 position of the L1 element, was shown. Use of a differential screening strategy of an NTera1 cDNA library, by employing ORF1 and 5' UTR probes, led to the preliminary identification of antisense transcription originating from within the 5' UTR of the L1s. Four of the cDNAs mapped to known genes and the corresponding chimeric transcripts were identified in cell lines by RNAse protection assay. Several of these cDNAs were shown to be correctly spliced. Also, by using a series of deletions and using a luciferase reporter system, Mart Speek narrowed down the critical location of ASP to nucleotides 399 to 467 which also has binding sites for transcription factors like Sp1 and SOX, suggesting their roles in supporting the transcription from the L1Hs ASP. Subsequently his group also showed by bioinformatics analysis, the existence in the human genome of several chimeric expressed sequence tag (cEST) sequences containing the L1 ASP region, demonstrating the potential influence of the antisense transcription of L1s located in introns of genes (Nigumann et al., 2002).

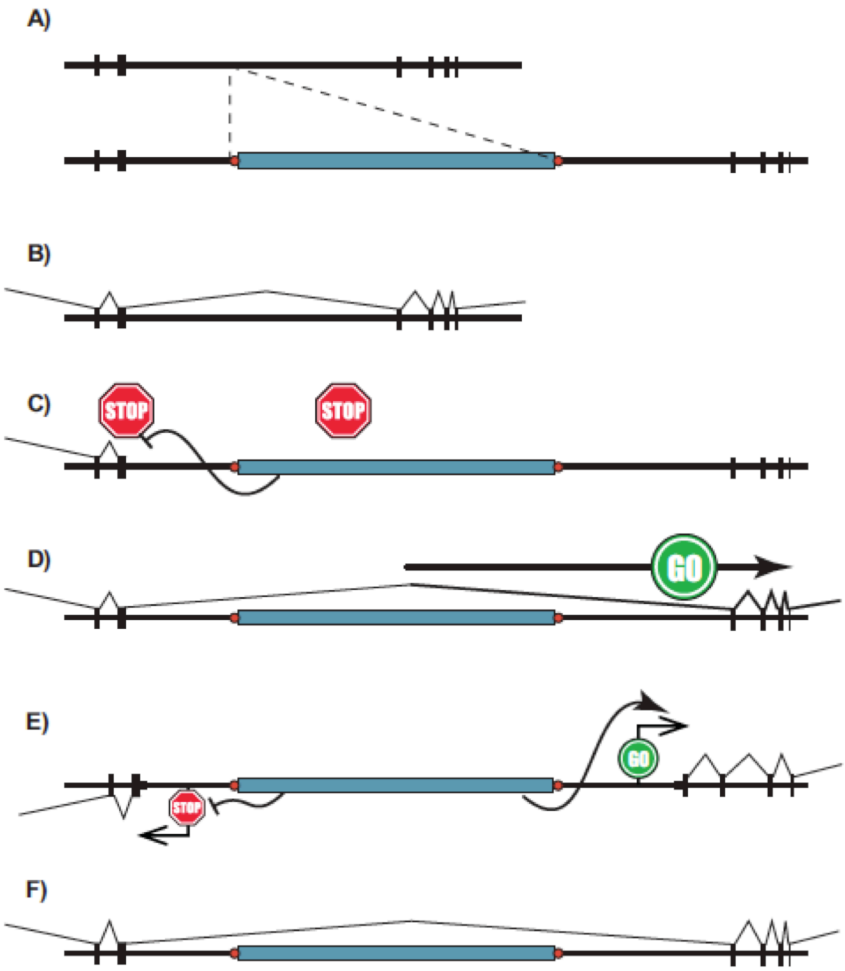


Figure 2 Potential impacts of a transposon insertion on the neighbouring genes

Shown here are models of possible genomic and transcriptional variation due to mammalian transposon integrants. A: Schematic of a gene without a TE insertion (top) and with a polymorphic TE insertion (bottom, indicated by blue rectangle). Black vertical rectangles, gene exons; horizontal black lines, genomic sequences including introns; red circles, target site duplications. B–E: Effects of TE insertions on gene expression. B: Typical gene expression with RNA splicing between exons, removing introns in the absence of a transposon insertion. C: Premature transcriptional termination (Stop) triggered by an intronic TE acting at long genomic distances [24, 82]. D: TE-mediated transcriptional activation or upregulation (GO), due to internal promoter or enhancer activity. E: Intergenic TEs may incorporate enhancer or silencer activities that variably affect gene expression upstream and/or downstream. These effects can help TEs influence transcriptional regulatory networks. F: We hypothesize that many TE integrants may have no detectable effect on gene expression, depending on TE age and structure, tissue-specific differences and genomic context.

[Figure reproduced and legend adapted from (Akagi et al., 2013), after obtaining copyright permission from John Wiley and Sons]

The role of the ASP in tissue-specific expression of genes has also been shown previously (Matlik et al., 2006). Other studies have also shown the effect of RIFTs in influencing the expression neighbouring genes in humans (Cruickshanks and Tufarelli, 2009; Kim and Hahn, 2010). More recently, an aberrantly activated ASP of an L1 element was shown to give rise to an antisense RNA, LCT13, the expression of which was linked to the epigenetic silencing of a metastasis suppressor gene tissue factor pathway inhibitor 2 (TFPI-2) in breast and colon cancer (Cruickshanks et al., 2013).

Because L1s contain both sense and antisense promoters in their sequence (the sense promoter located in the 5' UTR in both mouse and human, while the antisense located in the 5' UTR in human and the ORF1 in mouse, as we show here), bidirectional transcription in L1 presents the possibility of involvement of the RNA interference, yet another epigenetic process, as a potentially regulatory mechanism for L1 mobilization, as is the case with the regulation of other transposable elements. Single-stranded transcripts also can be processed to small RNAs, regardless of whether they are initiated within or outside of L1 elements. Resulting L1-specific small RNAs could mediate transcriptional and/or posttranscriptional gene silencing (Aravin and Bourc'his, 2008; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008; O'Donnell and Boeke, 2007; Yang and Kazazian, 2006; Yu et al., 2008).

The proof of principle that RNAi machinery could act on human L1 and control its transposition in cell culture was given, using $L1_{RP}$ and cancer cell lines (HeLa and HCT116) (Soifer et al., 2005). In the study, it was shown that long dsRNA of various L1 regions (ORF1, ORF2 and 5' UTR) can be cleaved by the Dicer to yield functional siRNA, and the L1 siRNA can act on L1 hybrid transcripts using the RNAi pathway. They also showed that the 5' UTR region of $L1_{RP}$ could be targeted by RNAi, thereby affecting the retrotransposition of the RC-L1 in cell culture. The first report of detection of small RNAs against human LINE elements came out in 2006, from the Kazazian lab, using 293 and HeLa cell lines (Yang and Kazazian, 2006). In this study, small RNAs of about 20 nucleotides were detectable, and a knockout of Dicer1 led to a marginal increase in L1 transcription and retrotransposition, thereby implicating the role of RNAi in silencing in vivo. However, further attempts to substantiate this claim have not been very successful, and more investigations are required (Kazazian, 2011). Subsequently however, RNAi has been shown to play a role in controlling human L1 mobility in other cell types and cancers. More recently, endogenous siRNA was shown to repress L1 by increasing the L1 promoter methylation in human breast cancer cells (Chen et al., 2012).

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

embryonic stem (ES) cells (Chow et al., 2010). Mouse chimeric transcripts containing AS L1 sequences also have been identified (Akagi et al., 2008; Zemojtel et al., 2007). Together, these results suggest that mouse L1 elements also may contain one or more antisense promoters (ASP). In fact, Zemojtel et. al. had indicated that an antisense promoter is likely to be present in ORF1 region of mouse L1, based on analysis of exonization events (i.e., representation of L1 sequences in transcripts derived from genes in whose introns they lie) 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. Also, while it is known that RNA interference plays a role in repressing LTR retrotransposons like IAP and MuERV-L in preimplantation mouse embryos (Svoboda et al., 2004), the role of RNAi in controlling LINE-1 mobility and expression in mouse remained to be explored.

In this chapter, I describe how we identified an active mouse AS L1 promoter within ORF1, immediately proximal to the transcription start sites (TSS) of AS L1 RIFTs. We found that the resulting AS mouse L1 RIFTs, including spliced, unspliced and many noncoding RNAs, were initiated by L1s interspersed genome-wide. Our results indicate that AS L1 RIFTs contribute to the diverse transcriptome (including long noncoding RNAs) expressed in various tissues (Conley et al., 2008; Mourier and Willerslev, 2009; Van de Lagemaat et al., 2003a). We also provide evidence for how the antisense promoter also helps to limit mouse L1 retrotransposition through a Dicer-independent mechanism (Yu et al., 2008). I performed experiments involving promoter activity assays, retrotransposition assay, and phage library screening in the laboratory of Dr. David Symer, in collaboration with Dr. Jingfeng Li, who performed the other experiments. The bioinformatics analyses were done by Dr. Keiko Akagi.

3.2 MATERIALS AND METHODS

3.2.1 Maintenance of mouse colonies, cells lines, and isolation of genomic DNA and RNA

Mice were maintained and euthanized according to approved Institutional Animal Care and Use Committee protocols (National Cancer Institute, Frederick, MD, USA; Ohio State University, Columbus, OH, USA). Mouse strains and purified genomic DNA were purchased from the Jackson Laboratory (Bar Harbor, ME, USA). A mouse spermatocyte cell line (CRL2196) was purchased from the American Type Culture Collection. HeLa cells were provided by Dr John V. Moran (University of Michigan). HCT116 *Dicer* ex5 knockout cells were provided by Dr Bert Vogelstein (Johns Hopkins). Genomic DNA and pooled total RNAs were isolated from CRL2196 cells and from various tissues, ages and lineages of mice as indicated, using standard methods and Trizol (Invitrogen), respectively.

3.2.2 Cell culture and L1 retrotransposition assay

HeLa cells were cultured in DMEM with 10% heat-inactivated fetal bovine serum and 2% penicillin/streptomycin (Gibco). Cells at ~75% confluence in six-well plates or T25 flasks were transfected with plasmid DNA mixed with FuGENE HD (Roche) at a ratio of 1 μg to 3 μl FuGENE. To quantify transfection efficiency, GFP expression was assessed by fluorescence microscopy in cells transfected with pEGFP-N1 (Clontech). Both stable and transient transfection assays were performed. In the stable assay (Moran et al., 1996; Symer et al., 2002), cells were treated with 0.2 mg/ml Hygromycin for various periods, starting 3 days after transfection. Hygromycin-resistant (HygroR) cells then were grown without selection for several days, prior to selection for NeoR L1 integrants in 0.4 mg/ml G418 (Invitrogen) for two weeks. In the transient assay (Wei et al., 2000), NeoR L1 integrants were selected directly. After discrete colonies formed in either assay, cells were washed in 1X phosphate buffered saline (PBS), fixed in 2% formaldehyde/0.2% glutaraldehyde in 1X PBS, washed and stained using 0.4% Giemsa (Sigma) at RT overnight and then counted.

To assess effects of AS L1 transcripts overexpressed in trans, we co-transfected HeLa cells with smL1 donor plasmid together with AS smL1 fragment-expressing constructs. One μg of pCEP4/smL1/Neo donor plasmid DNA was mixed with 1 μg of various smL1 AS fragment-expressing constructs or empty vector pCEP4, respectively, in FuGENE 6. Two micrograms of pCEP4 vector was used as another negative control. The transient assay for retrotransposition was performed to test impacts of the AS smL1 fragments on retrotransposition, by plating cells at various dilutions and counting resulting Neo^R colonies.

A similar experiment was performed to assess inhibition of L1spa retrotransposition (from donor plasmid pTN201) on expression of AS L1 transcripts.

3.2.3 Quantitation of β-lactamase expression

Genomic DNA fragments representing four mouse L1 subfamilies (Goodier et al., 2001) (T_F GenBank accession number AF016099; G_F , AC068252; A, AY053456 and FIII, AC002406) and a synthetic L1 element smL1 (Han and Boeke, 2004) were amplified by PCR using Platinum Taq HiFi (Invitrogen) and forward and reverse primers incorporating BglII and NcoI restriction sites, respectively. Amplicons included fragments of L1 T_F (represented by L1spa in pTN201), sense promoter (primers DES1212 and DES1213), AS promoter (DES1218 x DES1220, DES1218 x DES1221) and AS ORF2 (DES1459 x DES1460). Promoter candidates were cloned directionally upstream of TEM1, a β-lactamase reporter gene, in plasmid pBLAK-b, which lacks a promoter (Invitrogen) (Figure 9 and Figure 10). They were confirmed by sequencing. One microgram of Bgl II-digested (linearized) plasmid DNA was transiently transfected into CRL2196 or HeLa cells using FuGENE 6 (Roche). As positive and negative controls, plasmids with and without the SV40 promoter upstream of TEM1 were used (pBLAK-c and pBLAK-b, respectively).

To quantify β-lactamase protein expression, cells were stained with CCF2/AM substrate (Invitrogen) (Knapp et al., 2003; Zlokarnik et al., 1998) by replacing culture medium with 1ml loading solution [2 ml of a 1-mM CCF2/AM solution, 16 ml of solution B, 10 ml of 250mM Probenicid (Sigma) and 972 ml Hanks' Balanced Salt solution, (HBSS)] per 9.6 cm^2 well. Cells were incubated in the dark at room temperature (RT) for 1 h with gentle shaking, washed with HBSS and visualized using an Axiovert 200M inverted fluorescence microscope (Zeiss) with blue/aqua and β-lactamase ratio filter sets (Chroma Technology Corp.) and ORCA-ER high resolution digital camera (Hamamatsu Photonics) using Openlab software (version 4.0.2, Improvision). Flow cytometric analysis was performed using a BD LSR II flow cytometer with a 405nm violet laser, 440/ 40nm (blue) and 530/30nm (green) filters, and FACSDiva software (BD Biosciences). Ratios of blue to green intensities were collected as a linear parameter. Each flow cytometry session included positive and negative controls to normalize output.

TEM1 β-lactamase expression also was quantified by quantitative reverse transcriptase-mediated PCR (qRT-PCR). Promoter candidates were linearized by BglII digestion and transfected into HeLa cells using FuGENE 6. Total RNAs were isolated ~48 h after transfection using RNeasy kit (Qiagen). Standard curves were based on serial dilutions of

control plasmids. First strand cDNAs were synthesized using oligo-d(T) (DES2633) primer and the SuperScript double-stranded cDNA synthesis kit (Invitrogen). As further controls, RNAs were treated with and without reverse transcriptase. qRT-PCR was performed on an iCycler (Bio-Rad) or Step One Plus (Applied Biosystems) instrument, using SYBR Green Supermix (Bio-Rad). TEM1 β-lactamase transcript concentrations were calculated by interpolation, after subtracting for input plasmid DNA contamination. Beta-actin transcript levels were calculated for each sample. Each sample was measured in triplicate. Results are presented for each sample as the normalized ratio of TEM1 β-lactamase to β-actin transcript levels.

3.2.4 Chromatin immunoprecipitation of RNA polymerases

Anti-mouse RNA polymerase III subunit RPC39 mouse monoclonal antibody was purchased from Santa Cruz (catalog no. SC-21753), and anti-mouse RNA polymerase II mouse monoclonal antibody (Cat. 39097), from Active Motif. For chromatin immunoprecipitation, the Magna ChIP G Tissue kit (Millipore) was used, and the manufacturer's instructions were followed for performing the procedure.

3.2.5 Identification of TSS of TE-initiated fusion transcripts

We performed 5' RACE cloning using a second generation 5'/3' RACE kit (Roche) with an AS L1 ORF1-specific primer, DES1947 for first strand cDNA synthesis.

3.2.6 Phage library screens for mouse transcripts containing L1 sequences

Double-stranded DNA probes for mouse L1 ORF2 and ORF1 transcripts (Figure 9) were amplified by PCR from L1spa (Genbank Accession No. AF016099), a representative full-length T_F template (Naas et al., 1998). The primer pairs used to amplify fragments from L1spa ORF2 and ORF1 were DES1165 x DES1166 and DES 1167 x DES1168, respectively. The resulting PCR products were gel purified and radiolabeled by random nonamer priming. Bacteriophage cDNA libraries from mouse testis (Clontech) and thymus (Stratagene) were hybridized with an ORF2 probe, followed by an ORF1 probe. Commercial bacteriophage cDNA libraries from mouse testis (Clontech) and thymus (Stratagene) were plated at ~50,000 plaques per dish, transferred to Hybond-N filters (Amersham) and hybridized with the ORF2 probe. Filters were washed at 65°C in 0.1X SSC and 0.1% SDS and exposed for autoradiography. This procedure was repeated with the ORF1 probe to identify $ORF1^{+}ORF2^{-}$ clones, which were purified upon additional rounds of hybridization. Phage plaques were converted to plasmids and sequenced using BigDye version 3.1 (Applied Biosystems) on a 96-capillary sequencer (Transgenomic Spectrumedix), using primers DES886 and DES837 (5' and 3' ends, testis cDNA)

and standard M13R and M13F oligonucleotides (5' and 3' ends, thymus).

3.2.7 Computational identification of AS L1 RIFTs

A BLASTN search of mouse EST databases from testis and other tissues was conducted using AS L1spa (TF subfamily) ORF1 as query, i.e. AS nucleotides 2225–1838 (cf. coordinates, Figure 9).

3.2.8 Identification of RIFTs using exon microarrays

To develop a novel assay to identify L1 RIFTs, we modified the manufacturer's protocol of the Affymetrix GeneChip mouse exon microarray. First strand cDNA synthesis was performed on total RNA isolated from various tissues and lineages, using SuperScript II reverse transcriptase (Invitrogen), and a primer including both T7 promoter and oligo-d(T) sequences. Polyadenylated cDNAs containing AS L1 sequences were amplified using a primer for a particular L1 ORF1 AS template sequence paired with the T7 promoter primer. Resulting doublestranded RIFT cDNAs containing T7 sequences were used as templates for in vitro transcription, following standard procedures (Affymetrix). Resulting AS RNA was purified; a second round of first strand cDNA synthesis was performed with reverse transcriptase, dUTP and random primers; cRNA was hydrolyzed using RNaseH, and resulting sense strand DNA was purified. Products were fragmented with uracil DNA glycosylase and apurinic/apyrimidinic (AP)-endonuclease I. Terminal labeling was performed with terminal deoxynucleotidyl transferase (TdT), and resulting labeled fragments were hybridized to the exon microarray.

Resulting raw data from an Affymetrix microarray chip reader were analyzed for transcript expression, using Partek Genomics Suite software. CEL files (MoEx-1_0-st-v1) were imported using RMA background correction and quantile normalization. Probe intensities were transformed to log base 2. We defined signals with intensity> mean+one standard deviation (i.e. ~log2 intensity > 7) as high expression probes and counted the number of consecutive high expression probes per annotated gene. On alignment with the reference mouse genome, candidate fusion transcripts were scored positive if a neighboring AS L1 could be identified within 30 kb of an overlapping RefSeq gene and/or within 100 kb of the upregulated probe(s). We also required five consecutive high expression probe intensities (corresponding to adjacent exons in a given gene) in exon microarray data; length of predicted initiating genomic L1 integrant had to be more than 5 kb; and its subtype had to be L1 T_F, A, G_F or F as per RepeatMasker (www.repeatmasker.org).

3.2.9 Measuring promoter activity of L1 ASP fragments: Derivation of a new scoring system for transcription initiation activity

We sought to quantify promoter activities of candidate fragments using flow cytometric assay readouts of reporter expression (fluorescence levels). Because the reporter provides higher sensitivity and cells when analyzed by a cytometer fall in a wide range of fluorescence values, we designed a system to attribute a numeric value to the strength of the promoter, in terms of the proportion of cells that showed blue fluorescence as compared to the control blue cells.

All raw data files (.scf) obtained in a single flow cytometry run were imported into a separate workspace in FlowJo software (version 8.4.3 for Macintosh). Each run included a control for green cells (negative control) and blue cells (positive control). A subset was created by gating cells on forward and side scatter parameters. Keeping the green control cells as the reference, a second gate ("non-green cells" gate) was created that excluded all the green cells of any given sample. The data was graphically displayed as ratio of blue and green fluorescence values of each cell ("scaled ratio", X-axis) and the cell number (Y-axis); the axes limits were adjusted to fit the display within range. The statistic mean of the ratio of the cells in the final gate was calculated for each sample. Using the table editor, the blue/green ratio of each sample was normalized to that of the control blue cells (that had the highest score) to report the final score as follows: the mean ratio of each sample was multiplied with the total number of cells to get the total score. This score was divided by the score obtained for blue cells, and the quotient was multiplied by 100. So the final score of every sample in a given sample set (workspace) is the percentage of blueness with respect to the control blue cells' final score, which was taken to be100%.

3.2.10 Construction of modified L1-reporter plasmids

In order to know if the antisense promoter affected the L1's retrotransposition in cell culture assay, we replaced the putative antisense promoter region of L1spa present in the plasmid pTN201 (Naas et al., 1998) with the corresponding region of the recoded synthetic mouse L1 (smL1) (obatined from the plasmid pCEPsmL1). This plasmid, designated pMK28, was used in the retrotransposition assays, along with other L1 constructs (Figure 19). The series of subcloning steps done for constructing this plasmid is given below in detail (including those of all intermediate plasmids made; sections 3.2.10.1 to 3.2.10.6), while the construction details are given in brief for the other plasmids used (sections 3.2.10.7 and 3.2.10.8).

3.2.10.1 Construction of pMK20:

As a first step in replacing the ORF1 of the T_F element (L1spa) with the ORF1 of smL1, an 8.185 kb NotI-XhoI fragment of the 19.775 kb pTN201 vector was subcloned into the smaller pBluescript KS(+) for ease of working. The resultant plasmid is pMK20.

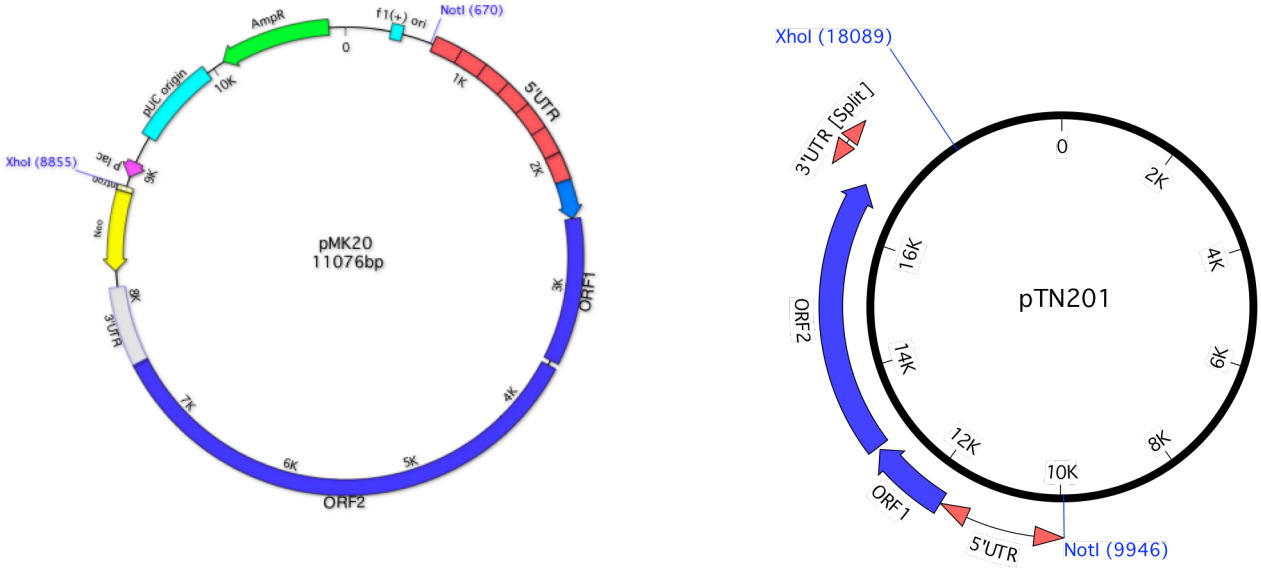


Figure 3 Annotated vector maps of pMK20 (plasmid intermediate for sublconing) and pTN201 (donor for L1spa, a mouse T_F L1)

3.2.10.2 Construction of pMK21:

Another subcloning was done to further reduce the size. A PstI-HindIII fragment of pMK20 was subcloned into pBluescriptKS(+) to give pMK21 (5.812 kb). This 2.865 kb fragment includes a small portion of ORF1, inter-ORF region (where the PacI site is to be introduced by mutagenesis) and upto about two-thirds of ORF2. This plasmid was used as a template to perform site-directed mutagenesis.

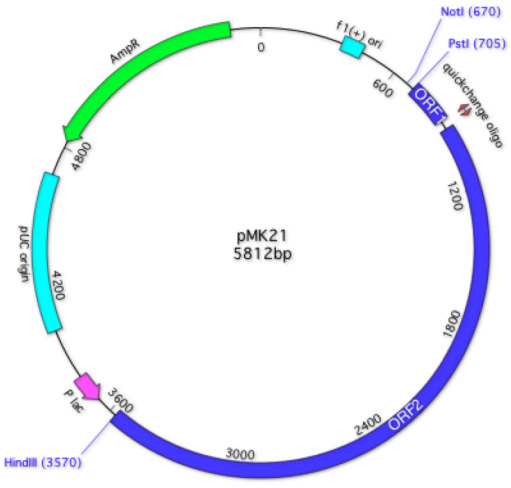


Figure 4 Annotated vector map of pMK21, a plasmid intermediate used for subcloning

3.2.10.3 Construction of pMK22:

A PacI site was introduced in the inter-ORF region in pMK21 using Stratagene's QuickChange Mutagenesis kit resulting in the plasmid pMK22 (5.812 kb). This site was to facilitate modular transfer of cassettes and fragments in the future.

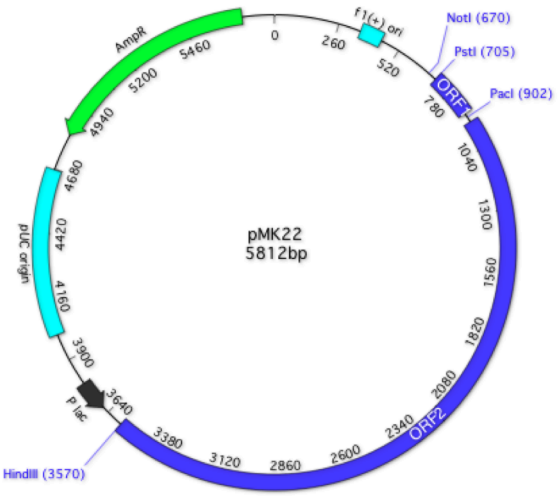


Figure 5 Annotated vector map of pMK22, a plasmid used to transfer cassettes modularly during subcloning

3.2.10.4 Construction of pMK22pCEPsmL1:

pMK22 was restricted with NotI and PacI to drop out a 0.2kb piece to be replaced with the 2.9kb NotI-PacI fragment from pCEPsmL1 (5'UTR and ORF1), resulting in pMK22smORF1(8.555 kb). So, this plasmid contains: 5' UTR - smORF1- inter-ORF region (with PacI site) - truncated ORF2 of L1 T_F, flanked on the 5' end by NotI site and the 3' end by HindIII site.

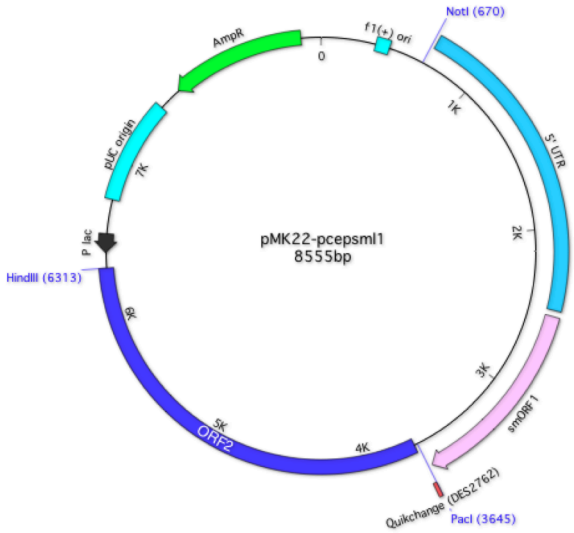


Figure 6 Annotated vector map of pMK22-pCEPsmL1, an intermediary plasmid that contains synthetic ORF1 (from ORFeus) but ORF2 (partial) from L1spa

3.2.10.5 Construction of pMK27:

The 5.6kb NotI-HindIII fragment of pMK22smORF1 was fitted back into the NotI-HindIII backbone(5.4 kb) of pMK20 to give pMK27 (10.999 kb). This plasmid was Sanger sequenced (Big Dye 3.1, Applied Biosystems; Transgenomic Spectrumedix), revealing a missense mutation in ORF2, i.e., Ala756Ser, along with two noncoding mutations present in pTN201. The missense mutation was repaired by replacement of the ~1.4kb EcoRI fragment with the same size fragment from pTN201, yielding the final version of pMK27.

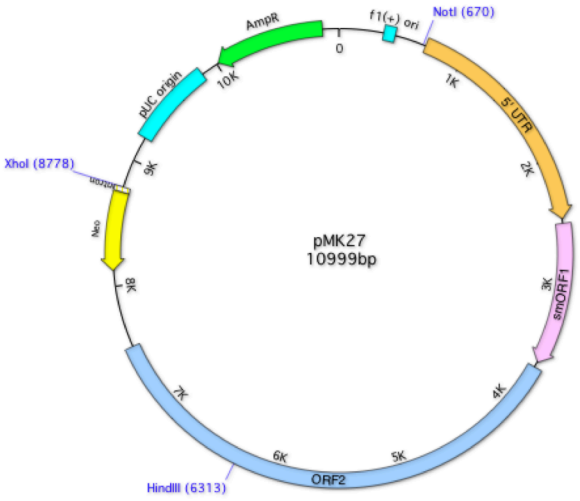


Figure 7 Annotated vector map of pMK27, an intermediate subcloning plasmid to be used to generate pMK28

3.2.10.6 Construction of pMK28:

The 8.1kb NotI-XhoI fragment of pMK27 was ligated with the ~11.7kb NotI-XhoI pTN201 fragment to give pMK28 (19.914 kb). This plasmid is same as pTN201 except for the ORF1 sequence. pMK28 has smORF1, whereas pTN201 has wild type ORF1 (L1 T_F). pMK28 was used in retrotransposition studies and showed ~5-fold increase in L1 activity in HeLa cells, compared to pTN201.

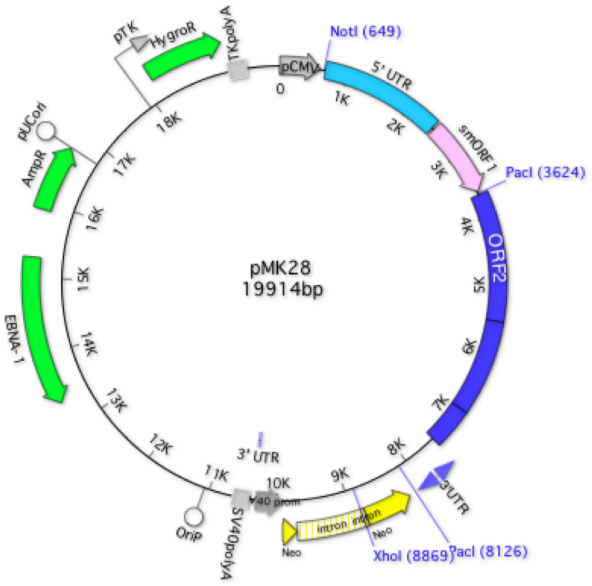


Figure 8 Annotated vector map of pMK28, an episomal donor plasmid that contains a marked hybrid L1 which, in turn, has its ORF1 from smL1, while ORF2 from L1 T_F

3.2.10.7 Construction of pJL2 and pJL3:

To preserve A/T content and synonymous amino acids of native L1s, while maximally changing codon usage, we also designed a novel nucleotide recoded L1 ORF1 fragment, corresponding to 2123–2932 nt from L1spa (Figure 11). This fragment (Blue Heron Bio), which also included 50-nt flanking arms on both ends for recombineering, was cloned into pUC MinusMCS, resulting in plasmid pJL2. The recoded L1 ORF1 fragment from pJL2 was amplified by PCR using DES3353xDES3354 and Platinum Taq DNA polymerase High Fidelity (Invitrogen), gel purified (Qiagen), mixed with PstI-linearized pMK20 and co-transformed into electrocompetent DY380 bacteria, bearing the lambda red recombination system for recombineering (Lee et al., 2001). After heat shock at 42°C for 15 min, to induce the lambda system, bacteria were cultured on LB+Carb agar plates at 32°C overnight. Candidate clones containing recombinant pMK20::pJL2 were screened by PCR, by PstI digestion, and verified by sequencing. Candidate and control plasmids were digested with NotI and XhoI at 37°C overnight. An 8.1-kb fragment containing the synthetic ORF1 was ligated to an 11.7-kb NotI-XhoI fragment from the pTN201 backbone. The final construct, pJL3 or L1spa::recoded-L1-ORF1, was verified by Sanger sequencing.

3.2.10.8 Construction of AS L1 transcript overexpression plasmids:

Various AS L1 transcript overexpression plasmids were engineered from L1 T_F template fragments generated by PCR using HiFi platinum Taq (Invitrogen), using primer pairs DES2880 x DES2881 (L1spa nucleotides 2150–1286); DES2880 x DES2882 (nucleotides 2150–1636) and DES2879 x DES2881 (nucleotides 2823–1286). Resulting PCR products were digested with NotI and BamHI, electrophoresed on agarose gels, purified and ligated to linearized pCEP4 backbone. Fragments from AS synthetic L1 elements were generated similarly using these primer sets: DES3818 x DES3820 (amplicon mapped to corresponding coordinates in L1spa, nt 2150–1121), DES3819 x DES3820 (nt 2150–1801), DES3818 x DES3821 (nt 2823– 1121) and DES3819 x DES3821 (nt 2823–1801). Products were digested with NheI and BamHI and ligated to similarly linearized pCEP4. These plasmids were used for transfection.

3.2.11 Investigating the role of Dicer in regulating L1 retrotransposition

To assay L1 retrotransposition in Dicer ex5 –/– HCT116 human colorectal cells, we had use an L1 cassette tagged with a reporter other than Neo^R, because that cell line already has Neo^R (Cummins et al., 2006). Hence we decided to use L1 donors marked by TEM1 β-lactamase-artificial intron (AI) from a construct pDES46 (which contained human L1.3; see Section 4.2.1 for details). The TEM1 β-lactamase - artificial intron (AI) reporter cassette and a portion of pCEP4 backbone were excised from pDES46 by digestion with NotI and BstZ17I. The resulting ~13-kb backbone fragment was gel-purified. Native, hybrid or fully synthetic mouse L1 constructs in pTN201, pJL3, pMK28 and pCEP4/smL1 were digested with BamHI, and the ends were filled in by Klenow, before digestion with NotI. Each of the resulting ~6.5-kb L1 fragments was ligated with the NotI - BstZ17I fragment (pCEP4 backbone) obtained earlier. Positive candidates were confirmed by Sanger sequencing. Resulting L1 donor plasmids, marked by the TEM1-AI reporter, included pTN201/TEM1, pJL3/TEM1, pMK28/TEM1 and pCEP4/smL1/TEM1.

To compare retrotransposition with native or hybrid L1 donors marked with TEM1-AI reporter in HCT116 wild-type versus Dicer –/– cell lines, transfectants were selected for ten days on 400 μg/ml hygromycin. Expression levels of spliced TEM1 transcripts were assayed by qRT-PCR using primers DES3062 and DES3063.

3.3 RESULTS

3.3.1 Mapping an active AS promoter in mouse L1 ORF1

Previously, others as well as our group, had identified mouse AS L1 RIFTs (Akagi et al., 2008; Zemojtel et al., 2007). Based on the approximate 5' ends of these RIFTs and widespread expression, we hypothesized that an active initiating promoter could reside in an AS orientation within ORF1 of mouse L1. To characterize this putative 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 β-lactamase reporter gene (Zlokarnik et al., 1998) lacking a promoter (Figure 9 and Figure 10). The fragments, were derived from mouse L1 subfamilies T_F, G_F, 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 T_F and G_F elements were engineered upstream of the TEM1 β-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). As described in Section 3.2.9, promoter strength scores were assigned to each fragment, based on β-lactamase reporter enzymatic activity expression visualized by microcopy and quantitated by cytometry. TEM1 transcript levels were measured by qRT-PCR (Figure 9 and Figure 10).

The highest level of AS promoter activity was found in L1 T_F AS nucleotides 2823–2125, mapped as per L1spa coordinates (Figure 9A). Various L1 subfamily members displayed distinct AS promoter activities, i.e. T_F (~40% of positive control, i.e. L1 T_F 5' UTR monomers in sense orientation) >> G_F ≈ A (~10% of control) > F (~5% of control). For these functional promoter assays, we chose particular elements to represent the subfamilies, i.e. L1spa for L1 T_F subfamily, L1 G_F62 for the G_F subfamily, and L1Md_A2 for the A subfamily (Figure 11). Within ORF1, these individual surrogates were 99.8, 99.7 and 99.9% identical to the consensus subfamily sequences, respectively. Differences between the subfamily consensus sequences and the individual surrogates were predicted at 1944A>G and 2261G>C (i.e., L1 T_F >L1spa, coordinates of sense strand, L1 T_F reference element nucleotide listed first); 1963C>T, 2687T>A, 2716T>C and 2857A>C (L1_G_F>L1 G_F62); and 2857G>A (L1_A>L1Md_A2).

A qRT-PCR assay for reporter transcript expression confirmed that L1 T_F AS promoter activity was robust, i.e. again, approximately half that of the L1 5' UTR sense promoter (Figure

9). Low but detectable promoter activities were observed in older L1 subfamilies including F, FII and/or FIII (Figure 10). By contrast, virtually no promoter activity was detected in various fragments derived from the sense (coding) orientation of ORF1, AS ORF2, L1 3' UTR, smL1 or a novel recoded ORF1 sequence which we designed to contain A/T content comparable with natural L1 sequences (Figure 9 and Figure 10).

We examined a potential basis for the broad range of AS promoter activities among different mouse L1 subfamilies. Although they are defined mainly by differences between 5' UTR sequences, their sequences within ORF1 also are distinct (Figure 11). Comparison of representative L1 subfamily amino acid sequences encoded by ORF1 indicated that the particular portions comprising the AS promoter were more conserved, but still distinct, between subfamilies, compared with the flanking, proximal and distal portions of ORF1 (Figure 11). By contrast, the L1 subfamily sequences within ORF2, which do not contain this AS promoter, were nearly identical (not shown). These results suggested that ORF2 and the AS promoter segment within ORF1 may have undergone strong purifying selection. A recent analysis of the evolution of mouse and human L1s confirmed that the mouse ORF1 coiled-coil domain has undergone much less adaptive change than that of human elements (Sookdeo et al., 2013).

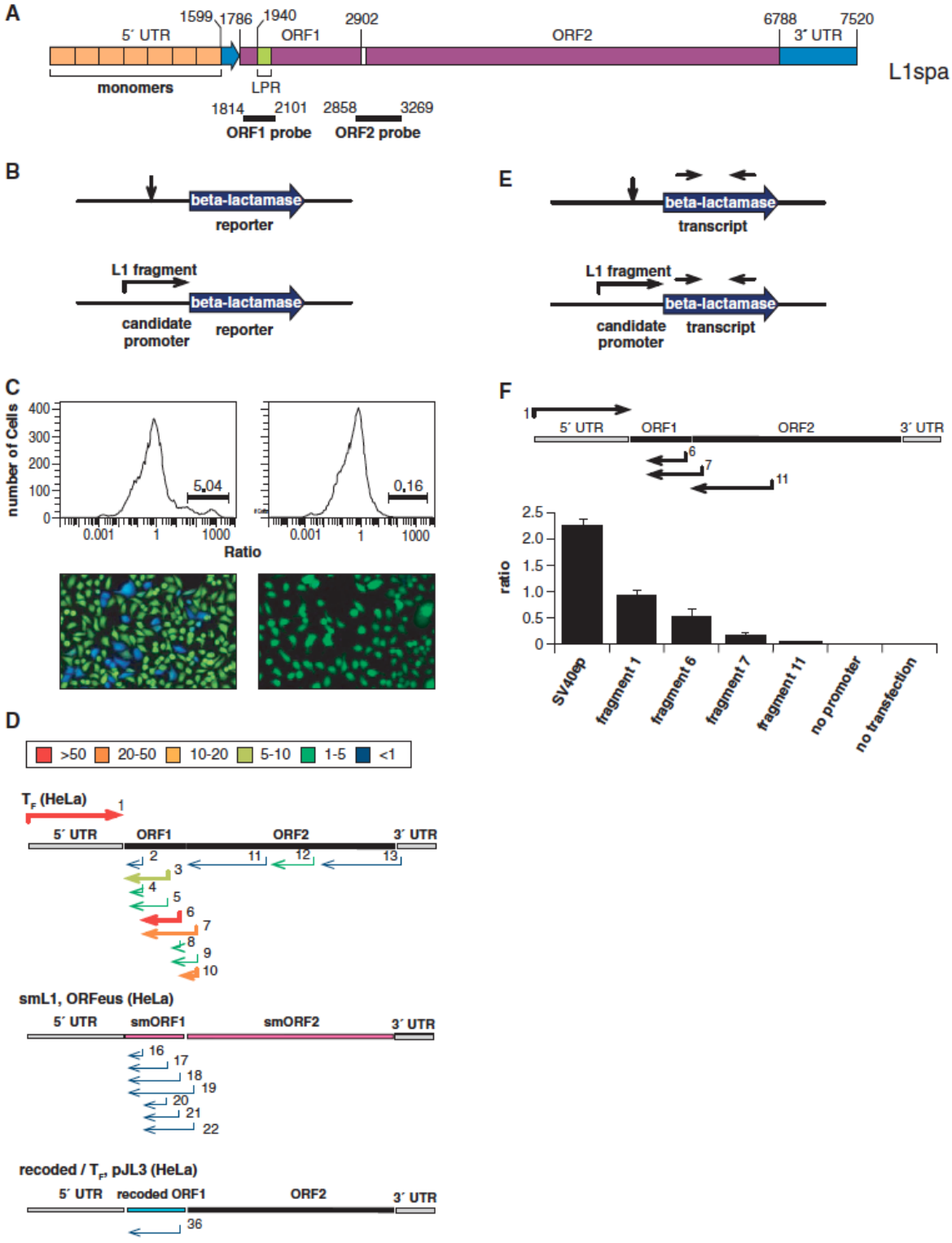


Figure 9 Mapping an active AS promoter within L1 ORF1.

(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). See Section 3.2.6 for description.

(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).

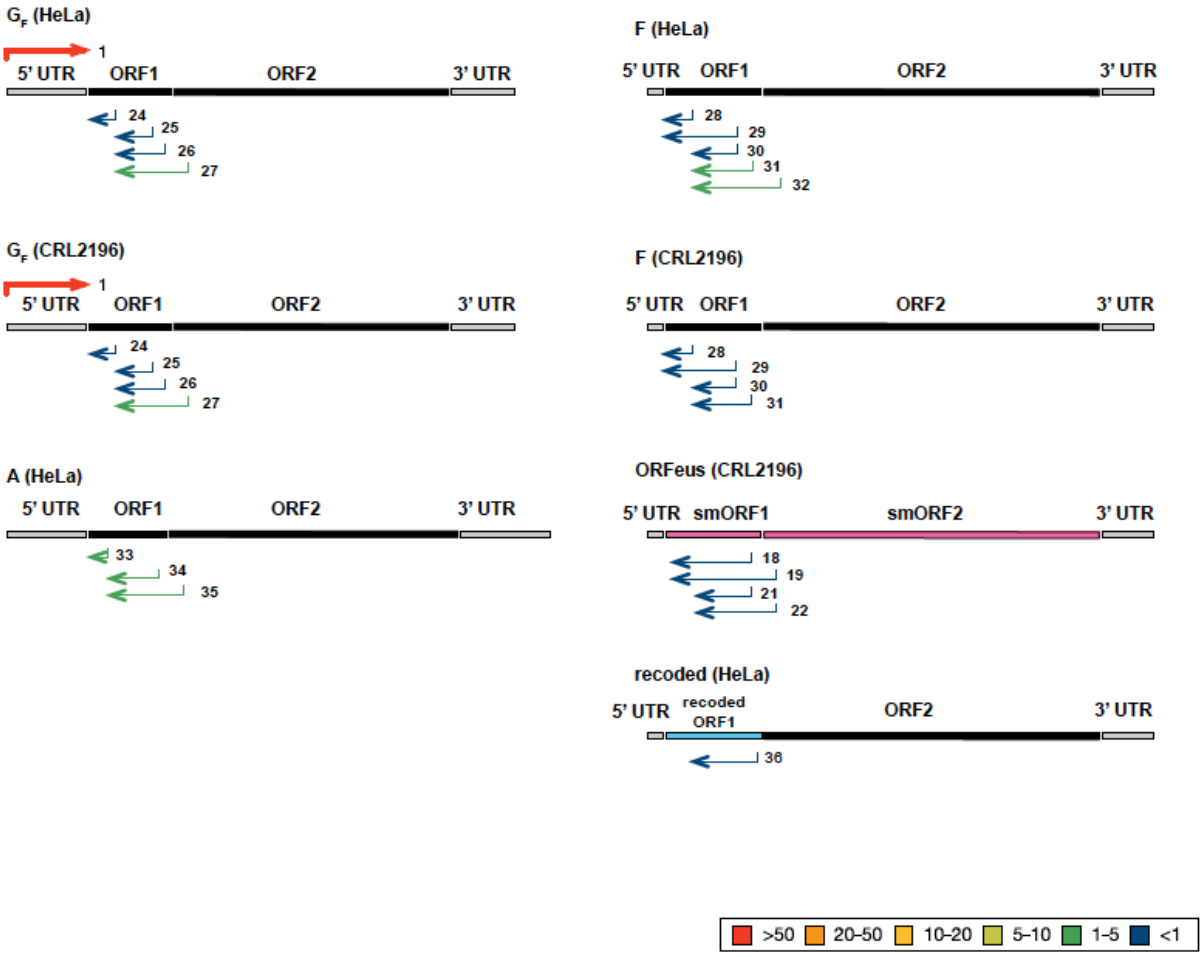


Figure 10 Additional reporter assays identify AS promoter activity within ORF1 of certain other L1 elements.

Additional mouse L1 fragments were engineered upstream of the β-lactamase TEM1 reporter in the orientation indicated by arrows, and assayed for promoter activity. Linearized DNAs containing the promoter-reporter cassette were transfected into cultured HeLa (human cervical carcinoma) and CRL-2196 (mouse spermatocyte) cells. Reporter expression was detected by staining cells with CCF2-AM. Promoter activity scores (colors in key, top) were assigned to each strand-specific candidate tested. See Section 3.2.9 for a description of the scoring system. Fragments are mapped to L1spa coordinates (arrows). The sources of fragments and the cell lines used for transfection also are indicated.

Table 2 Details of oligonucleotide primers used in the ASP study

Frag.	Subfamily	Oligo IDs	Position in L1spa	Length	Ori.	Forward oligo (5'- 3')	Reverse oligo (5'- 3')
1	T_F	1212/1213	14-1786	1773bp	s	agatctATTAGTCTGAACAGGTGAGAGG	ccatggCTGGTAATCTCTGGAGTTAG
2	T_F	1298/1797	2150-1801	350bp	as	ccatggACGGAGGAATCTTACTAACAGG	agatctGAGTTTTCTTTATTGTGTCTACTTCC
3	T_F	1298/1219	2648-1801	848bp	as	ccatggACGGAGGAATCTTACTAACAGG	agatctCTGGTGTAATTCTGATAGGCTTG
4	T_F	1299/1797	2150-1883	268bp	as	ccatggCCCAACACACCTGAGAACCT	agatctGAGTTTTCTTTATTGTGTCTACTTCC
5	T_F	1299/1219	2648-1883	766bp	as	ccatggCCCAACACACCTGAGAACCT	agatctCTGGTGTAATTCTGATAGGCTTG
6	T_F	1218/1220	2823-2125	699bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctTTCGTGGAGAGATAATGCGTG
7	T_F	1218/1221	3136-2125	1012bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctCCTTTCATTCTGAGGTAGTGTC
8	T_F	2093/1220	2823-2626	198bp	as	ccatggCAAGCCTATCAGAATTACACCAG	agatctTTCGTGGAGAGATAATGCGTG
9	T_F	2093/1221	3136-2626	511bp	as	ccatggCAAGCCTATCAGAATTACACCAG	agatctCCTTTCATTCTGAGGTAGTGTC
10	T_F	1221/1798	3136-2803	334bp	as	ccatggCACGCATTATCTCTCCACGAA	agatctCCTTTCATTCTGAGGTAGTGTC
11	T_F	1459/1460	4420-2931	1490bp	as	ccatggCACAAGAACAGAATGCCACC	agatctTCTTTGAAGGTCTGATAG
12	T_F	1461/1462	5296-4471	826bp	as	ccatggGTATTCTACCCAACTCATTTTATG	agatctTTCACTTCCTTCGTTAG
13	T_F	1463/1464	6905-5378	1528bp	as	ccatggATGGATTGGCAGGACCAAC	agatctGTGTTTTGTTCCCACTTCTAAG
14	T_F	2096/2097	1801-2150	350bp	s	agatctACGGAGGAATCTTACTAACAGG	ccatggGAGTTTTCTTTATTGTGTCTACTTCC
15	T_F	2096/2098	1801-3136	1336bp	s	agatctACGGAGGAATCTTACTAACAGG	ccatggCCTTTCATTCTGAGGTAGTGTC
16	Synthetic	2010/2012	2150-1801	350bp	as	ccatggTGACCAACCGCAACCAGGAC	agatctAGGTCCAGGATGGTCTTGTTC
17	Synthetic	2010/2013	2648-1801	848bp	as	ccatggTGACCAACCGCAACCAGGAC	agatctGGATGCGGTCCTTGTTCAGG
18	Synthetic	2010/2099	2823-1801	1023bp	as	ccatggTGACCAACCGCAACCAGGAC	agatctTTGGTCTTGTCGTGGAACACC
19	Synthetic	2010/2100	3136-1801	1336bp	as	ccatggTGACCAACCGCAACCAGGAC	agatctAACAGAAGGTGGGGTCCTGC
20	Synthetic	2011/2013	2648-2125	524bp	as	ccatggAACAAGACCATCCTGGACCTG	agatctGGATGCGGTCCTTGTTCAGG
21	Synthetic	2011/2099	2823-2125	699bp	as	ccatggAACAAGACCATCCTGGACCTG	agatctTTGGTCTTGTCGTGGAACACC
22	Synthetic	2011/2100	3136-2125	1012bp	as	ccatggAACAAGACCATCCTGGACCTG	agatctAACAGAAGGTGGGGTCCTGC
23	G_F	1214/1215	14-1786	1773bp	s	agatctCCATCTTCAGCTCCAGACAG	ccatggCTGGCAATCTCTGGAGTTAG
24	G_F	2031/1797	2150-1801	350bp	as	ccatggCGTAAGAATCCTACTAACAGAAG	agatctGAGTTTTCTTTATTGTGTCTACTTCC
25	G_F	1218/1219	2648-2125	524bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctCTGGTGTAATTCTGATAGGCTTG
26	G_F	1218/1220	2823-2125	699bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctTTCGTGGAGAGATAATGCGTG
27	G_F	1218/1221	3136-2125	1012bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctCCTTTCATTCTGAGGTAGTGTC
28	F	2036/1797	2150-1801	350bp	as	ccatggGAGATTACAAGATGGTGAAAGG	agatctGAGTTTTCTTTATTGTGTCTACTTCC
29	F	2036/1219	2648-1801	848bp	as	agatctGAGTTTTCTTTATTGTGTCTACTTCC	agatctCTGGTGTAATTCTGATAGGCTTG
30	F	1218/1219	2648-2125	524bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctCTGGTGTAATTCTGATAGGCTTG
31	F	1218/1220	2823-2125	699bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctTTCGTGGAGAGATAATGCGTG
32	F	1218/1221	3136-2125	1012bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	agatctCCTTTCATTCTGAGGTAGTGTC
33	A	2032/1797	2150-1801	350bp	as	ccatggCGTAAGAATCCTACTAACAGAAATC	agatctGAGTTTTCTTTATTGTGTCTACTTCC
34	A	1218/2138	2823-2125	699bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	aacgcgtTTCGTGGAAAGATATTGTGTG
35	A	1218/2137	3136-2125	1012bp	as	ccatggGGAAGTAGACACAATAAAGAAAACTC	aacgcgtCCTTTCACTCTGAGGTAGTGTC
36	recoded	4320/4321	2823-2125	699bp	as	ccatggAGAGGTTGATACGATTAAAAAGACGC	agatctTTTGTTGAAAGGTAGTGTGTA

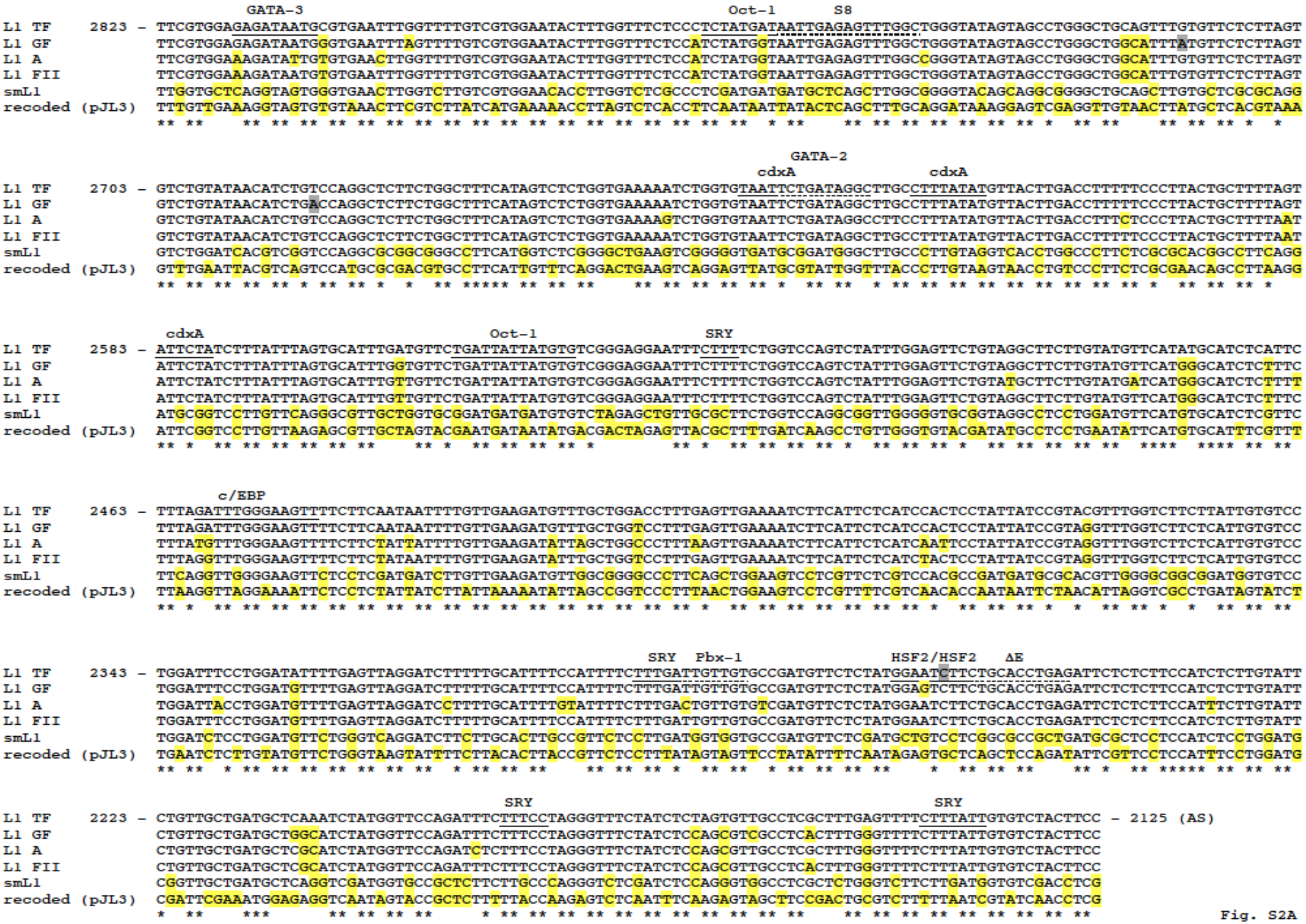
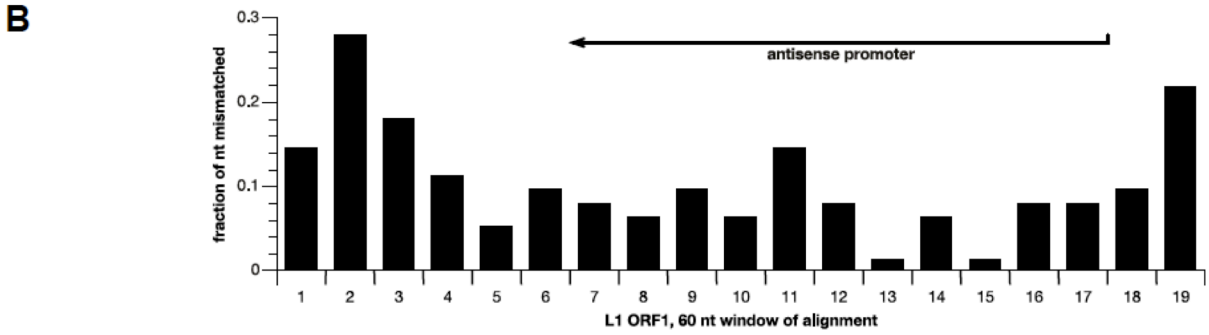

GATA-3
Oct-1
S8
L1 TF 2823 - TTCGTGGAGAGATAATGCGTGAATTTGGTTTTGTCGTGGAATACTTTGGTTTCTCCCTCTATGATAATTGAGAGTTTGGCTGGGTATAGTAGCCTGGGCTGCAGTTTGTGTTCTCTTAGT
L1 GF TTCGTGGAGAGATAATGGGTGAATTTAGTTTTGTCGTGGAATACTTTGGTTTCTCCATCTATGGTAATTGAGAGTTTGGCTGGGTATAGTAGCCTGGGCTGGCATTTATGTTCTCTTAGT
L1 A TTCGTGGAAAGATATTGTGTGAACTTGGTTTTGTCGTGGAATACTTTGGTTTCTCCATCTATGGTAATTGAGAGTTTGGCCGGGTATAGTAGCCTGGGCTGGCATTTGTGTTCTCTTAGT
L1 FII TTCGTGGAAAGATAATGTGTGAATTTGGTTTTGTCGTGGAATACTTTGGTTTCTCCATCTATGGTAATTGAGAGTTTGGCTGGGTATAGTAGCCTGGGCTGGCATTTGTGTTCTCTTAGT
smL1 TTGGTGCTCAGGTAGTGGGTGAACTTGGTCTTGTCGTGGAACACCTTGGTCTCGCCCTCGATGATGATGCTCAGCTTGGCGGGGTACAGCAGGCGGGGCTGCAGCTTGTGCTCGCGCAGG
recoded (pJL3) TTTGTTGAAAGGTAGTGTGTAAACTTCGTCTTATCATGAAAAACCTTAGTCTCACCTTCAATAATTATACTCAGCTTTGCAGGATAAAGGAGTCGAGGTTGTAACTTATGCTCACGTAAA
GATA-2
cdxA
cdxA
L1 TF 2703 - GTCTGTATAACATCTGTCCAGGCTCTTCTGGCTTTCATAGTCTCTGGTGAAAAATCTGGTGTAATTCTGATAGGCTTGCCTTTATATGTTACTTGACCTTTTTCCCTTACTGCTTTTAGT
L1 GF GTCTGTATAACATCTGACCAGGCTCTTCTGGCTTTCATAGTCTCTGGTGAAAAATCTGGTGTAATTCTGATAGGCTTGCCTTTATATGTTACTTGACCTTTTTCCCTTACTGCTTTTAGT
L1 A GTCTGTATAACATCTGTCCAGGCTCTTCTGGCTTTCATAGTCTCTGGTGAAAAGTCTGGTGTAATTCTGATAGGCCTTCCTTTATATGTTACTTGACCTTTCTCCCTTACTGCTTTTAAT
L1 FII GTCTGTATAACATCTGTCCAGGCTCTTCTGGCTTTCATAGTCTCTGGTGAAAAATCTGGTGTAATTCTGATAGGCTTGCCTTTATATGTTACTTGACCTTTTTCCCTTACTGCTTTTAAT
smL1 GTCTGGATCACGTCGGTCCAGGCGCGGCGGGCCTTCATGGTCTCGGGGCTGAAGTCGGGGGTGATGCGGATGGGCTTGCCCTTGTAGGTCACCTGGCCCTTCTCGCGCACGGCCTTCAGG
recoded (pJL3) GTTTGAATTACGTCAGTCCATGCGCGACGTGCCTTCATTGTTTCAGGACTGAAGTCAGGAGTTATGCGTATTGGTTTACCCTTGTAAGTAACCTGTCCCTTCTCGCGAACAGCCTTAAGG
cdxA
Oct-1
SRY
L1 TF 2583 - ATTCTATCTTTATTTAGTGCATTTGATGTTCTGATTATTATGTGTCGGGAGGAATTTCTTTTCTGGTCCAGTCTATTTGGAGTTCTGTAGGCTTCTTGTATGTTCATATGCATCTCATTC
L1 GF ATTCTATCTTTATTTAGTGCATTTGGTGTTCTGATTATTATGTGTCGGGAGGAATTTCTTTTCTGGTCCAGTCTATTTGGAGTTCTGTAGGCTTCTTGTATGTTCATGGGCATCTCTTTC
L1 A ATTCTATCTTTATTTAGTGCATTTGTTGTTCTGATTATTATGTGTCGGGAGGAATTTCTTTTCTGGTCCAGTCTATTTGGAGTTCTGTATGCTTCTTGTATGATCATGGGCATCTCTTTT
L1 FII ATTCTATCTTTATTTAGTGCATTTGTTGTTCTGATTATTATGTGTCGGGAGGAATTTCTTTTCTGGTCCAGTCTATTTGGAGTTCTGTAGGCTTCTTGTATGTTCATGGGCATCTCTTTC
smL1 ATGCGGTCCTTGTTCAGGGCGTTGCTGGTGCGGATGATGATGTGTCTAGAGCTGTTGCGCTTCTGGTCCAGGCGGTTGGGGGTGCGGTAGGCCTCCTGGATGTTCATGTGCATCTCGTTC
recoded (pJL3) ATTCGGTCCTTGTTAAGAGCGTTGCTAGTACGAATGATAATATGACGACTAGAGTTACGCTTTTGATCAAGCCTGTTGGGTGTACGATATGCCTCCTGAATATTCATGTGCATTTCGTTT
c/EBP
L1 TF 2463 - TTTAGATTTGGGAAGTTTTCTTCAATAATTTTGTTGAAGATGTTTGCTGGACCTTTGAGTTGAAAATCTTCATTCTCATCCACTCCTATTATCCGTACGTTTGGTCTTCTTATTGTGTCC
L1 GF TTTAGATTTGGGAAGTTTTCTTCAATAATTTTGTTGAAGATGTTTGCTGGTCCTTTGAGTTGAAAATCTTCATTCTCATCCACTCCTATTATCCGTAGGTTTGGTCTTCTCATTGTGTCC
L1 A TTTATGTTTGGGAAGTTTTCTTCTATTATTTTGTTGAAGATATTAGCTGGCCCTTTAAGTTGAAAATCTTCATTCTCATCAATTCCTATTATCCGTAGGTTTGGTCTTCTCATTGTGTCC
L1 FII TTTAGGTTTGGGAAGTTTTCTTCTATAATTTTGTTGAAGATATTTGCTGGTCCTTTGAGTTGAAAATCTTCATTCTCATCTACTCCTATTATCCGTAGGTTTGGTCTTCTCATTGTGTCC
smL1 TTCAGGTTGGGGAAGTTCTCCTCGATGATCTTGTTGAAGATGTTGGCGGGGCCCTTCAGCTGGAAGTCCTCGTTCTCGTCCACGCCGATGATGCGCACGTTGGGGCGGCGGATGGTGTCC
recoded (pJL3) TTAAGGTTAGGAAAATTCTCCTCTATTATCTTATTAAAAATATTAGCCGGTCCCTTTAACTGGAAGTCCTCGTTTTCGTCAACACCAATAATTCTAACATTAGGTCGCCTGATAGTATCT
SRY
Pbx-1
HSF2/HSF2
ΔE
L1 TF 2343 - TGGATTTCCTGGATATTTTGAGTTAGGATCTTTTTGCATTTTCCATTTTCTTTGATTGTTGTGCCGATGTTCTCTATGGAATCTTCTGCACCTGAGATTCTCTCTTCCATCTCTTGTATT
L1 GF TGGATTTCCTGGATGTTTTGAGTTAGGATCTTTTTGCATTTTCCATTTTCTTTGATTGTTGTGCCGATGTTCTCTATGGAGTCTTCTGCACCTGAGATTCTCTCTTCCATCTCTTGTATT
L1 A TGGATTACCTGGATGTTTTGAGTTAGGATCCTTTTGCATTTTGTATTTTCTTTGACTGTTGTGTCGATGTTCTCTATGGAATCTTCTGCACCTGAGATTCTCTCTTCCATTTCTTGTATT
L1 FII TGGATTTCCTGGATGTTTTGAGTTAGGATCTTTTTGCATTTTCCATTTTCTTTGATTGTTGTGCCGATGTTCTCTATGGAATCTTCTGCACCTGAGATTCTCTCTTCCATCTCTTGTATT
smL1 TGGATCTCCTGGATGTTCTGGGTCAGGATCTTCTTGCACTTGCCGTTCTCCTTGATGGTGGTGCCGATGTTCTCGATGCTGTCCTCGGCGCCGCTGATGCGCTCCTCCATCTCCTGGATG
recoded (pJL3) TGAATCTCTTGTATGTTCTGGGTAAGTATTTTCTTACACTTACCGTTCTCCTTTATAGTAGTTCCTATATTTTCAATAGAGTGCTCAGCTCCAGATATTCGTTCCTCCATTTCCTGGATG
SRY
SRY
L1 TF 2223 - CTGTTGCTGATGCTCAAATCTATGGTTCCAGATTTCTTTCCTAGGGTTTCTATCTCTAGTGTTGCCTCGCTTTGAGTTTTCTTTATTGTGTCTACTTCC - 2125 (AS)
L1 GF CTGTTGCTGATGCTGGCATCTATGGTTCCAGATTTCTTTCCTAGGGTTTCTATCTCCAGCGTCGCCTCACTTTGGGTTTTCTTTATTGTGTCTACTTCC
L1 A CTGTTGCTGATGCTCGCATCTATGGTTCCAGATCTCTTTCCTAGGGTTTCTATCTCCAGCGTTGCCTCGCTTTGGGTTTTCTTTATTGTGTCTACTTCC
L1 FII CTGTTGCTGATGCTCGCATCTATGGTTCCAGATTTCTTTCCTAGGGTTTCTATCTCCAGCGTTGCCTCACTTTGGGTTTTCTTTATTGTGTCTACTTCC
smL1 CGGTTGCTGATGCTCAGGTCGATGGTGCCGCTCTTCTTGCCCAGGGTCTCGATCTCCAGGGTGGCCTCGCTCTGGGTCTTCTTGATGGTGTCGACCTCG
recoded (pJL3) CGATTCGAAATGGAGAGGTCAATAGTACCGCTCTTTTTACCAAGAGTCTCAATTTCAAGAGTAGCTTCCGACTGCGTCTTTTTAATCGTATCAACCTCG
Fig. S2A

Figure 11 Alignments of L1 subfamily and recoded ORF1 sequences spanning the AS promoter.

*(A) AS L1 subfamily consensus sequences, corresponding to the active AS promoter region centered on ORF1 (*Figure 9*) were generated for four subfamilies, T_F, G_F, A and FII by querying the mouse reference genome (UCSC mm10) using seed sequences L1GF62, L1spa and L1Md_A2. In each case, >80 genomic elements were found that aligned at >97% identity to the query sequence. Consensus sequences were defined by majority rule. They were aligned to corresponding regions of synthetic mouse L1 (smL1, also called ORFeus) (Han and Boeke, 2004), and to our novel, recoded L1 in pJL3. Coordinates and the reference sequence were based on L1spa (L1 T_F). Predicted transcription factor binding sites were identified using TFSEARCH: Searching Transcription Factor Binding Sites (v. 1.3) (http://www.cbrc.jp/research/db/TFSEARCH.html). As shown here, they are underlined and labeled above the sequences. Yellow highlights: single nucleotide differences between subfamilies; gray highlights: differences between consensus elements and individual surrogates; asterisks below aligned nucleotides: conserved among all 6 aligned elements.*



(B) Plot of the fractional mismatch in 60 nt windows spanning the ORF1 region, presented in the sense orientation (based on L1spa reference sequence).

Arrow: approximate location of AS promoter.

3.3.2 RNA polymerase II transcribes AS L1 fusion transcripts

To confirm localization of AS promoter activity to mouse L1 ORF1 sequences and to define the RNA polymerase responsible for transcriptional initiation from it, we immunoprecipitated both RNA polymerases (pol) II and III, either of which plausibly could bind to and initiate fusion transcription from various endogenous TE sequences. As shown in Figure 12A, RNA pol II localized specifically to the ORF1 fragment that contains AS promoter activity, i.e. nucleotides 2125–2823. Notably, ChIP-PCR also demonstrated that RNA pol II bound to ORF1 nucleotides 1528–2061, mapping to L1 template sequences, downstream of the AS promoter, that are expressed as AS L1 fusion transcripts. As a control, ChIP-PCR analysis of SINE B2 sequences confirmed that both RNA pol II and RNA pol III bound to those sequences (Lunyak et al., 2007).

To confirm the role for RNA pol II in transcribing AS L1 RIFTs, we treated a mouse spermatocyte cell line, CRL2196, with alpha-amanitin (Figure 12B), a potent and specific RNA pol II inhibitor. We assayed AS L1 RIFT expression by qRT-PCR, demonstrating substantial inhibition by this drug both in general and at individual loci. Together with ChIP-PCR, our results indicated that AS L1 RIFTs were transcribed by RNA pol II.

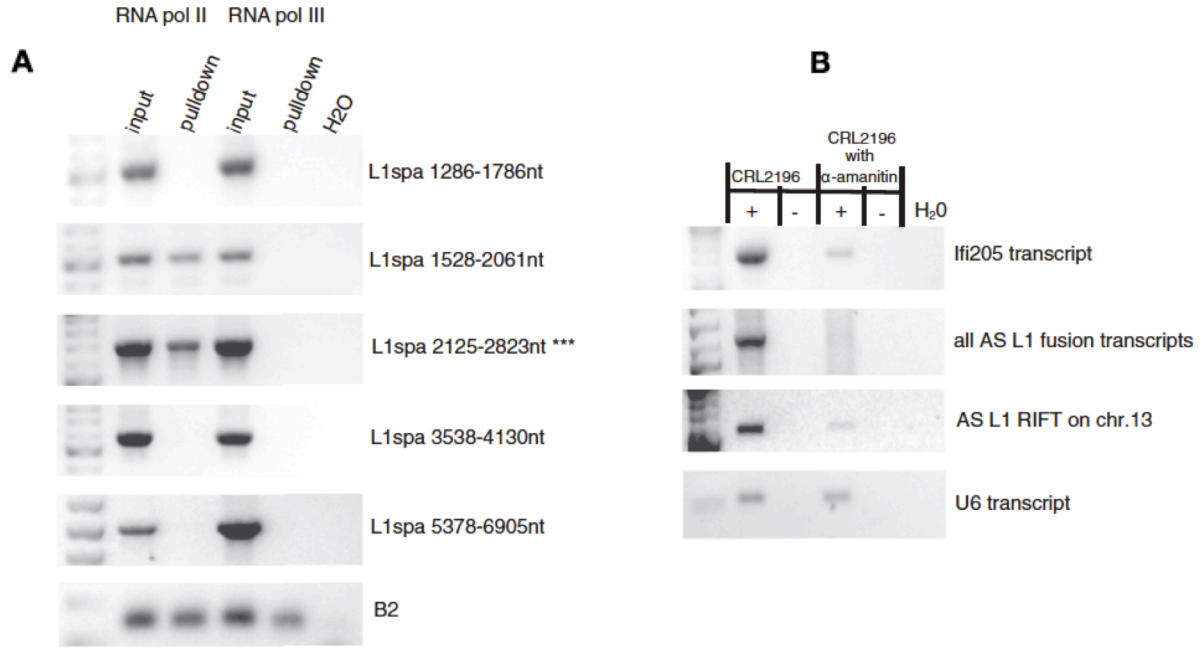


Figure 12 RNA polymerase II transcribes AS L1 fusion transcripts.

(A) Chromatin immunoprecipitation (ChIP) with anti-RNA polymerase II (left) and anti-RNA polymerase III (right) antibodies, followed by PCR amplification of target L1 or SINE B2 genomic sequences as indicated (right), showed specific enrichment (pulldown) of pol II at the AS L1

promoter in mouse testis (asterisks, L1 ORF1 sequences). Coordinates from L1spa reference are shown (right, cf. Figure 9*). RNA pol II also immunoprecipitated proximal L1 sequences, i.e. templates for transcribed AS fusion transcripts. As a control, both pol II and pol III pulled down SINE B2 elements genome-wide (bottom) as expected.*

(B) Mouse spermatocytes were treated with a-amanitin (RNA polII inhibitor) as indicated (top). Total RNAs were isolated, and reverse transcriptase was added as indicated (+ or – ; top) before PCR amplification of various cDNAs as indicated (right). As a negative control, U6 transcripts (RNA pol III, not inhibited by a-amanitin) were amplified (bottom).

3.3.3 Identification of diverse AS L1 RIFTs

To find mouse transcripts that included sequences from genomic L1 templates, we screened full-length transcripts represented in bacteriophage libraries. Although 6–30% of all mouse and human transcripts were recently reported to be initiated from TEs including L1s (Faulkner et al., 2009), we observed that only ~0.1 and 0.03% of all transcripts in phage cDNA libraries representing testis and thymus, respectively, hybridized with an L1 T_F subfamily probe for ORF2 (hereafter called ORF2$^+$ transcripts). Sequential hybridization with an L1 T_F ORF1 probe (Figure 9A) revealed an additional 0.06% of testis transcripts and 0.02% of thymus transcripts, identifying those that contained 5' L1 ORF1 but not ORF2 sequences. Of 940 testis cDNA clones hybridizing with either probe, 363 (~39%) were ORF1$^+$. Similarly, of 253 thymus cDNA clones hybridizing with either probe, 99 (~39%) were ORF1$^+$.

We hypothesized that such ORF1$^+$ ORF2$^-$ transcripts would include AS L1 RIFTs. This possibility was prompted by our previous identification of fusion transcripts in adult mouse tissues, mapping to L1 elements (Akagi et al., 2008). Of 27 ORF1$^+$ ORF2$^-$ transcripts identified from testis, 21 (78%) contained AS L1 ORF1 sequences spliced with other exons in the sense orientation, forming AS L1 RIFTs. Additionally, 2 of 13 (15%) thymus cDNAs also were spliced AS L1 RIFTs. Other ORF1$^+$ ORF2$^-$ cDNAs either were unspliced AS RIFTs, reading antiparallel to ORF1 through the 5' UTR into flanking genomic sequences (4 in testis, 15% of total; 2 in thymus, 15%), or were prematurely polyadenylated, sense-strand transcripts (2 in testis, 7%; 9 in thymus, 69%) (Perepelitsa-Belancio and Deininger, 2003). Some RIFTs were initiated in other mouse strains by polymorphic L1s absent from the C57BL/6 J (B6) reference genome (Akagi et al., 2008; Akagi et al., 2010). These screens also showed that some AS RIFTs were readily detectable without PCR amplification.

We identified diverse spliced ORF1$^+$ ORF2$^-$ transcripts initiated across the genome in a

variety of chromosomal and tissue contexts, as illustrated by schematics of their genomic templates including the initiating L1 elements (Figure 13) (Peaston et al., 2004). To determine whether AS L1 RIFTs were expressed more broadly, we screened additional mouse strains and cell lines by qRT-PCR. We experimentally identified 41 additional AS L1 RIFTs expressed in cultured mouse spermatocyte cells or adult testes. Twelve (29%) aligned to genomic regions lacking a previously annotated gene, and two (5%) were initiated from polymorphic L1s absent from B6 mice (Akagi et al., 2010). In addition, we searched public expressed sequence tag (EST) libraries by BLAST alignments, revealing 15 additional full-length mouse testis ESTs as spliced AS L1 RIFTs. Fifty-seven EST clones contained AS L1 sequences in their 5' ends. Of these, 22 were spliced, but no splicing was observed within L1 sequences *per se*.

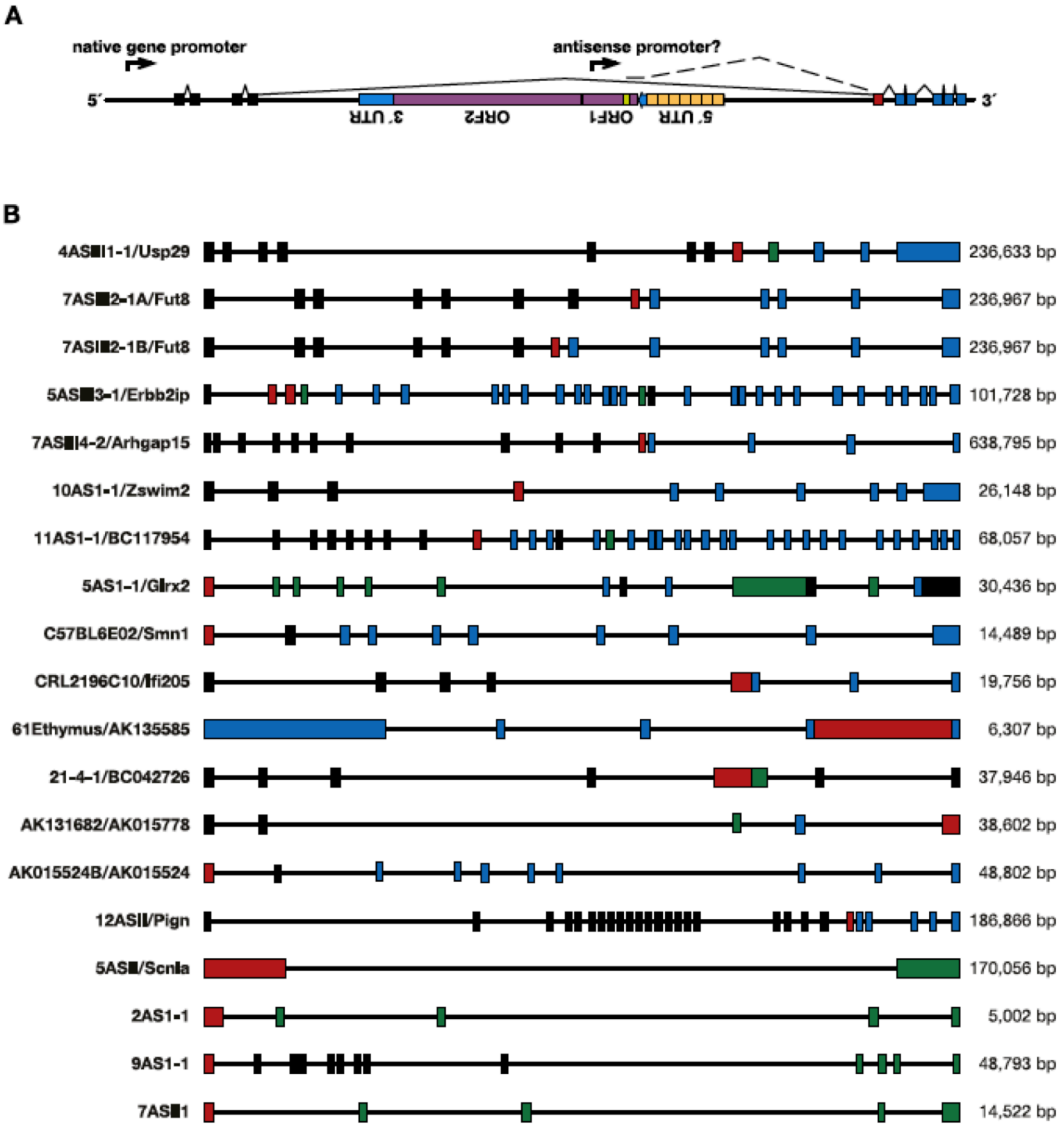


Figure 13 Genomic templates of AS L1 RIFTs.

(A) Schematic of an AS L1 integrant, located within an intron of an arbitrary gene.

(B) Schematics of diverse genomic templates and exons displaying the templates for specific AS L1 RIFTs, following a previously published format (Akagi et al., 2010).
Fusion AS L1-gene cDNAs were aligned to the B6 reference genome (UCSC browser, release mm8). Chromosomal sequences are represented by single horizontal lines, and spliced exons included in the cognate gene and/or RIFT are indicated by rectangles: (black) conventional transcript exons omitted from RIFTs; (red) first alternative first L1 exon; (blue) conventional transcript exons included in RIFTs; and (green) RIFT exons omitted from conventional transcripts. Also indicated are the cDNA clone and gene names (left) and the genomic length (bp) spanned by the L1 RIFT (right).

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.

To compare RIFT expression levels in different tissues, we re-assayed 17 RIFTs identified initially in adult testis or from a spermatocyte cell line. As expected, almost all of these RIFTs were confirmed in testis (Figure 14). Relatively few were expressed in other tissues assayed, but we did recover clones 1ASII1, additionally expressed in 11-day embryos; L1-5AS1-1, additionally expressed in brain; and CRL2196C10, widely expressed in most tissues assayed. We also assayed for overlapping spliced transcripts from cognate genes. Although AS L1 RIFTs that were spliced to downstream exons of *Erbb2ip*, *Usp29* and *Arhgap15* each were expressed in testis, the corresponding conventional transcripts of these genes (i.e., those lacking sequences from L1s) were not detectable there.

To identify genes whose expression levels may be affected by AS L1 RIFTs, we probed Affymetrix mouse exon microarrays conventionally with total RNAs. As commercial exon microarrays typically exclude probes for repetitive elements such as L1 retrotransposons, 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. At least 130 unique spliced AS L1 RIFTs were identified in adult testis, of which many were also identified in phage cDNA libraries. Thus, many transcripts were corroborated by independent methods.

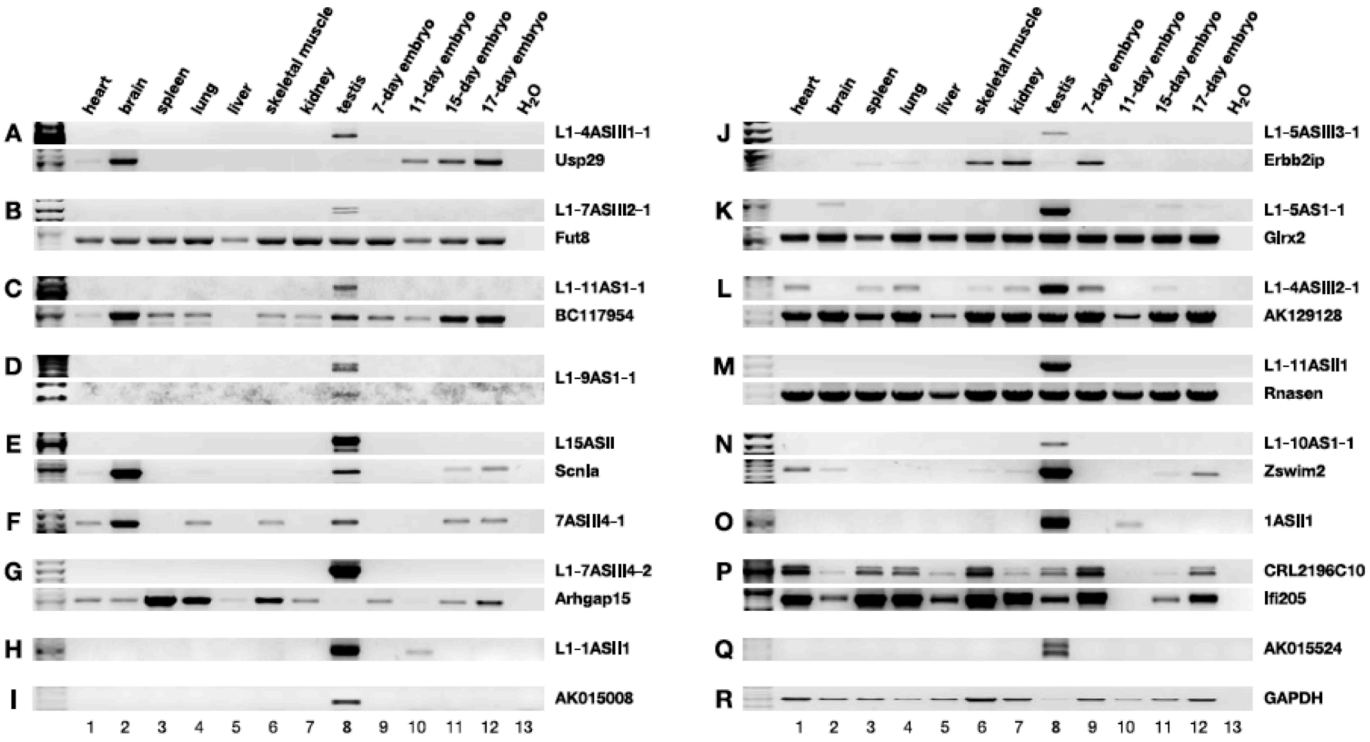


Figure 14 Tissue-specific expression of L1 RIFTs.

RT-PCR was performed for 17 pairs of L1 fusion transcripts (upper panel of each pair) and their overlapping cognate genes (where applicable; lower panels). A mouse multiple tissue cDNA panel, generated from a mouse strain known to include indicated L1 variants, was used in this screen. As a positive control, GAPDH transcripts were amplified by RT-PCR.

Both assays, i.e. the RIFT assay and conventional expression profiling using exon microarrays, confirmed the expression of an AS L1 RIFT at Arhgap15, initially found by screening a testis cDNA library. The initiating L1 integrant is polymorphic and is oriented antiparallel to the transcription unit of Arhgap15. The AS L1 RIFT, expressed in the same orientation as the overlapping gene's reading frame (Figure 15), was readily detectable in B6 but not others, consistent with the presence or absence of the initiating L1 element. Both assays showed that this AS RIFT contributed to overall *Arhgap15* RNA levels, in particular those measured at its 3' end.

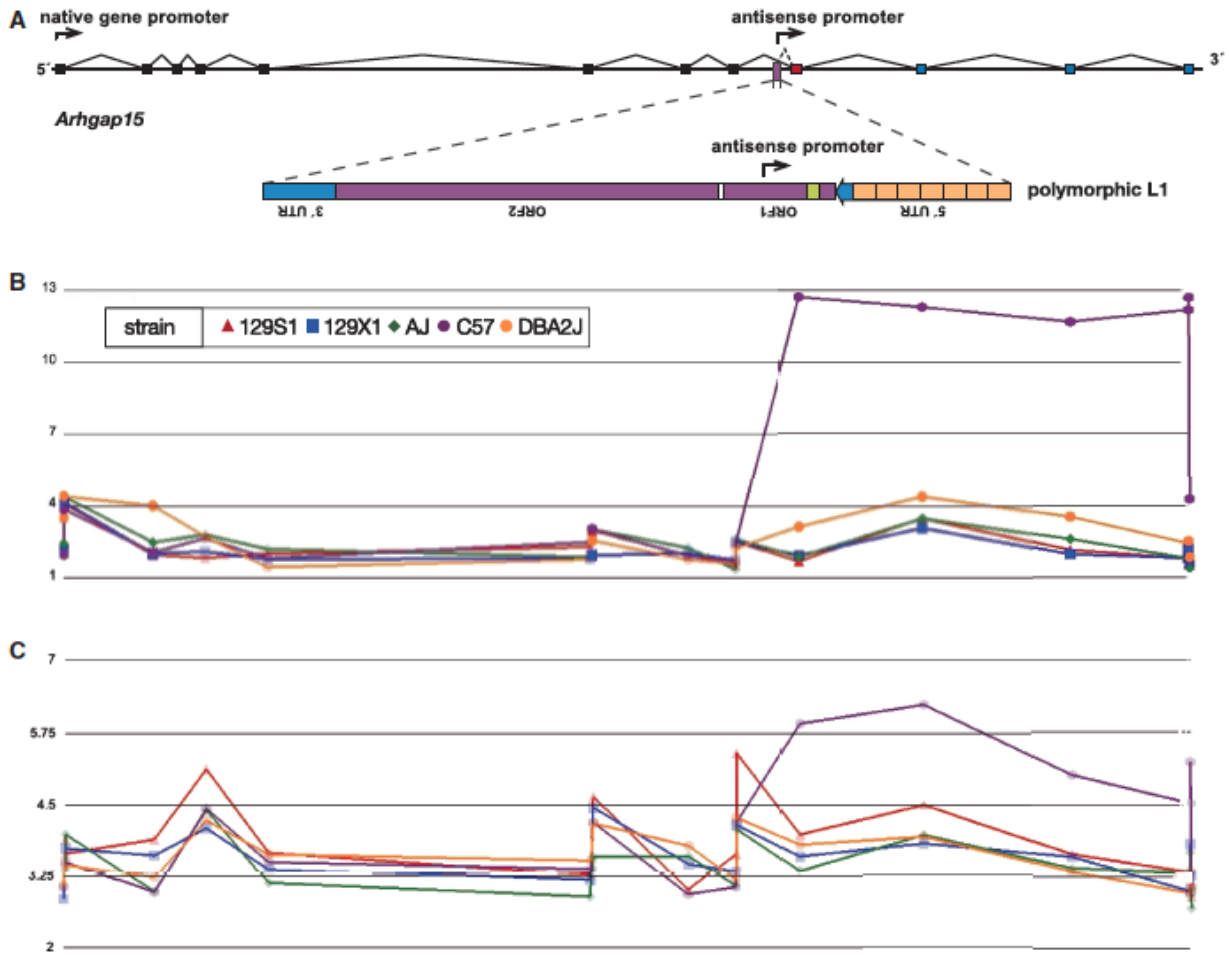


Figure 15 Contribution of an AS L1 RIFT to overall Arhgap15 gene expression in various mouse strains.

(A) Schematic representation of Arhgap15 exons, including a polymorphic AS L1 integrant in the B6 reference genome but not in other strains.

(B) AS L1 RIFT expression at Arhgap15 was detected in B6 mice, using the novel RIFT assay where we performed RT-PCR using AS L1 and oligo-d(T) primers, followed by hybridization of resulting cDNA products to an Affymetrix mouse exon microarray. We required five consecutive exon probes to be strongly positive to call RIFTs. Shown are genomic positions of probes within exons (x-axis) and hybridization signal intensities on a log scale (y-axis). Legend, inset: five mouse strains, different symbol colors and shapes.

(C) Conventional assay for Arhgap15 expression in total RNAs (see legend, B). The AS L1 RIFT in B6 mice affects total RNA expression levels at the 30 exons downstream of the polymorphic, initiating L1 integrant (see corresponding positions, A)

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 (Figure 16). Notably, a few RIFTs were expressed in more than one tissue. Thus most, but not all, RIFTs were expressed in a tissue-specific fashion. In addition, comparison of RIFTs expressed in five diverse strains highlighted that approximately half were conserved in all five strains (Figure 16), implying that potential biological functions of some RIFTs may be shared. Other RIFTs were expressed only in particular strains, consistent with the presence of the polymorphic L1 AS promoter in about half of these cases and with differential RIFT expression in the others.

Using targeted RT-PCR, we observed ~40% of extant L1 T_F subfamily members studied here initiated a nearby AS L1 RIFT. About 13% of L1 G_F elements, about 4% of A elements, and zero of one F element initiated RIFTs. Overall, about 19% of 68 genomic elements initiated RIFTs.

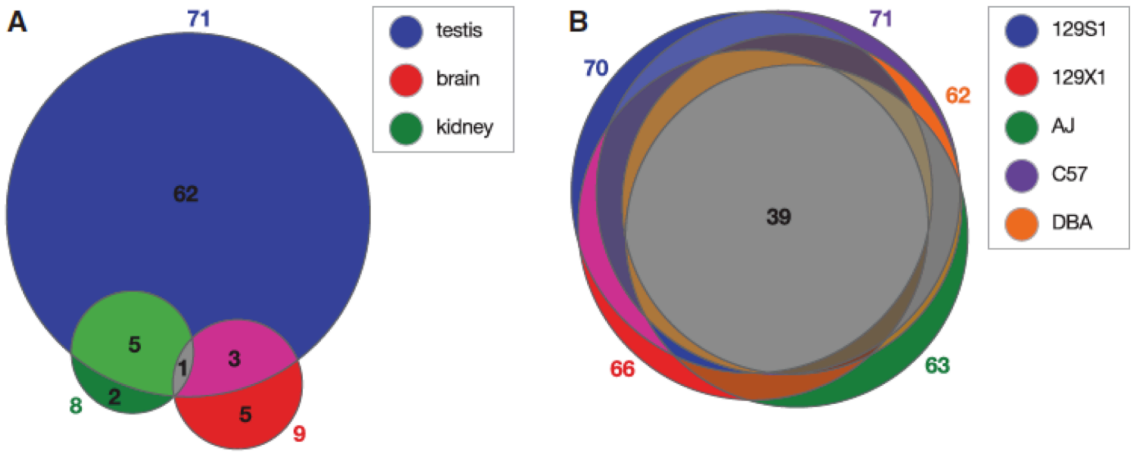


Figure 16 Comparison of AS L1 RIFTs expressed in various mouse tissues and strains.

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).

3.3.4 AS L1 RIFT TSS are proximal to the AS L1 promoter

To identify the 5' transcription start site (TSS) of AS L1 RIFTs, we performed 5' rapid amplification of cDNA ends (5' RACE) analysis on fusion transcripts expressed in testis, kidney and brain. A primer specific for the L1 T_F ORF1 template was paired with a standard RACE primer for amplification from total RNAs. A range of PCR product sizes was observed, revealing multiple nearby TSS (Figure 17). A large fraction of the 5' ends of transcripts recovered from all three tissues mapped to ORF1 nucleotides 2201–2244. In kidney and brain, additional TSS mapped to a wider range of ORF1 sequences, i.e. nucleotides 2210– 2306 and nucleotides 2210–2478, respectively. These results correlated well with the 5' ends of RIFT cDNAs identified in phage libraries. In addition, the 5' ends of 24 RIKEN cDNA clones, most of which were reported previously (Zemojtel et al., 2007), mapped to this same region. Thus, the 5' TSS of the fusion transcripts, determined experimentally by 5' RACE analysis and from cDNA clones, were closely adjacent to the experimentally mapped AS L1 promoter (Figure 9).

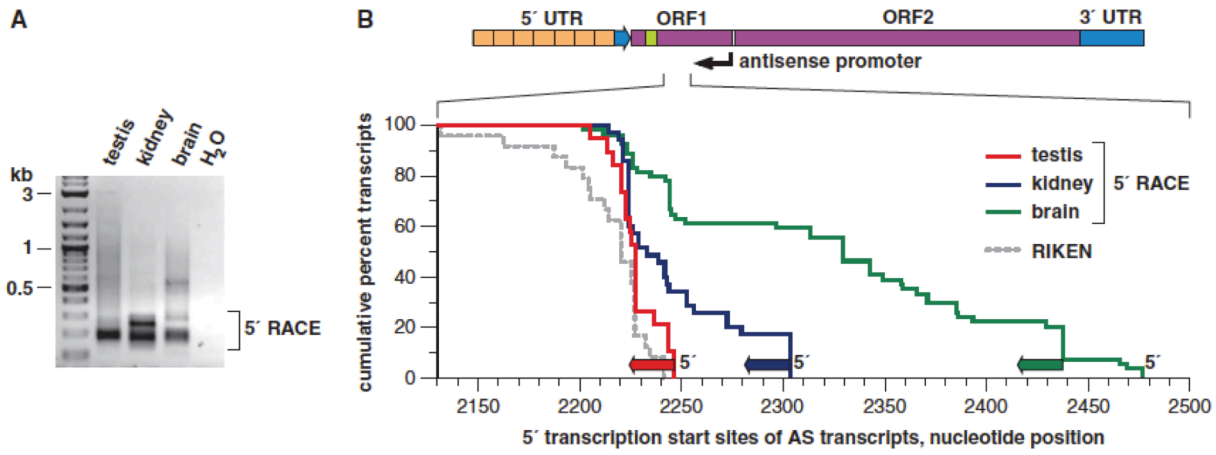


Figure 17 AS transcription start sites found by 5' RACE in multiple tissues.

(A) A 5' RACE was performed by PCR for 5' ends of AS L1 RIFTs, using total RNAs from testis, kidney and brain. Products were separated by agarose gel electrophoresis. Individual cloned 5' ends were sequenced from these pools.

(B) The cumulative positions of TSS for AS L1-gene RIFTs are plotted by summing 5' ends of individual transcripts, mapped against coordinates from L1spa. We analyzed 19 5' RACE clones from testis (red), 35 from kidney (blue) and 54 from brain (green). Also superimposed here are the cumulative positions of 5' ends from 24 RIKEN clones that align well with L1spa, although these formally are not ends determined by 5' RACE cloning.

We observed a candidate transcript-initiating TATAA sequence at position 2698 of AS L1 T_F ORF1 (Figure 11), but it is likely too distant from the RIFTs' 5' ends, identified by RACE (Figure 17), to account for them. Nevertheless, many mouse and human promoters lacking TATAA sequences have been identified previously, including variants of an 'initiator element (Inr)' sequence (Smale and Baltimore, 1989). We noted several variants of this sequence within the mapped AS promoter, some of which were immediately adjacent to observed TSS in the RIFTs.

3.3.5 Impacts of AS transcription on L1 transcription and retrotransposition

The synthetic mouse L1 element smL1 retrotransposes about 200-fold more than endogenous mouse L1s (Han and Boeke, 2004). Increased RNA polymerase II processivity, and increased expression of L1 ORF1 and ORF2 were proposed to be causes of this increase. Compared with mouse smL1, a synthetic human L1 (ORFeus-Hs) retrotransposed only about 3-fold more than the most active native human L1 elements (An et al., 2011). The exact basis for the differential increase in retrotransposition by synthetic mouse more than synthetic human L1s, over the corresponding native elements, is unknown. We noted that smL1 lacked the AS promoter activity harbored in ORF1 by native mouse L1s, thereby possibly contributing to marked increases in its expression and retrotransposition. To test this possibility, we replaced native ORF1 in L1spa with the synonymous fragment from smL1, forming a novel, hybrid full-length L1 donor, pMK28. To assess the role for A/T content in affecting L1 transcript levels, we also synthesized a second partially recoded hybrid L1 donor, i.e. as in pJL3. Like smL1, the recoded L1 in pJL3 also lacked AS promoter activity (Figure 9), but it had higher A/T content, similar to that of native mouse L1 elements. We also measured transcript levels expressed from these native or hybrid L1 donor elements using qRT-PCR. The lowest L1 transcript levels were observed for native L1 TF (L1spa), whereas the highest levels were seen for full-length smL1 (Figure 18). Intermediate levels were seen for the novel hybrid element containing recoded ORF1, harboring no AS promoter activity and neutral changes in A/T content, engineered upstream of native L1spa (T_F) ORF2. Somewhat higher expression was seen for the second hybrid L1 element, i.e. smL1/ L1spa in pMK28, which has lower A/T content in ORF1. The results suggested a potential contribution by native AS L1 promoters in reducing L1 transcription.

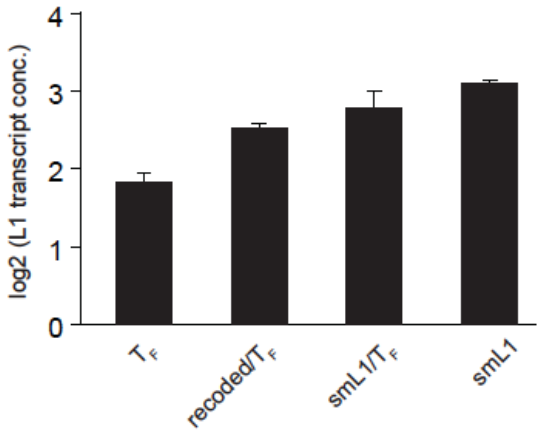


Figure 18 AS L1 promoter activity may contribute to reduced native L1 transcription.
To evaluate the effects of native AS ORF1 transcription upon transfected L1 retrotransposon transcript levels expressed from donor plasmids, we transiently transfected native, hybrid and fully synthetic elements. Transcript levels of the various L1 elements were measured by qRT-PCR assays of ORF2 levels, normalized to HygroR transcript levels (y-axis, presented as log2 of transcription concentration, based on delta Ct = –(Ct (ORF2) – Ct (Hygro))). Categories, x-axis: TF, native L1spa from pTN201; recoded/TF, neutral changes in A/T content of recoded ORF1 in pJL3; smL1/TF, low A/T content of ORF1 swapped from smL1 into L1spa, resulting in a hybrid L1 donor, pMK28; smL1, fully synthetic L1 ORF1 and ORF2 in pCEP4/smL1. Values, mean of duplicates; error bars, range.

We also compared mobilization of the various engineered L1s (Figure 19 and Figure 20). The hybrid L1 with reduced ORF1 A/T content in pMK28 retrotransposed at least 100-fold more than native L1spa. The partially recoded hybrid L1 in pJL3, with neutral changes in ORF1 A/T content, mobilized up to ~39-fold more than native L1spa. We conclude that synonymous disruption of the AS L1 promoter in ORF1, regardless of its A/T content, can increase retrotransposition substantially. These results are also consistent with evidence showing that longer L1 templates bearing reduced A/T content can result in increased transcript levels and retrotransposition (Han and Boeke, 2004). Thus, the AS L1 promoter helps to limit retrotransposition in *cis*.

To determine if overexpressed AS L1 transcripts could inhibit retrotransposition in *trans*, first we engineered AS smL1 fragments to overexpress them in the desired orientation. Four AS fragments from smL1, corresponding to AS L1spa coordinates 2119–1120, 2800– 1120, 2119–1812 and 2800–1812, each were cloned downstream of the CMV promoter and were co-

transfected with marked smL1 in a transient retrotransposition assay (Wei et al., 2000) (Figure 19 and Figure 21).

As a positive control, where smL1 could mobilize in the absence of overexpressed AS L1 transcripts in trans, empty pCEP4 was co-transfected with smL1. A negative control consisted of cells transfected with no smL1 donor and pCEP4 alone. The overexpression of AS smL1 transcripts in trans suppressed smL1 retrotransposition, by ~50–75% (Figure 19). This significant level of repression was comparable with that of human L1 siRNAs (Yang and Kazazian, 2006). In another experiment, several distinct native AS L1 T_F transcripts (generated from L1spa template at coordinates 2823–1286, 2150–1286 and 2150–1636; cf. Figure 9 and Figure 17) were overexpressed (Figure 21). These AS L1 transcripts overlapped in part with endogenous AS L1 RIFTs. Their expression in trans suppressed L1 retrotransposition at comparable levels, i.e. two- to five-fold (Figure 22).

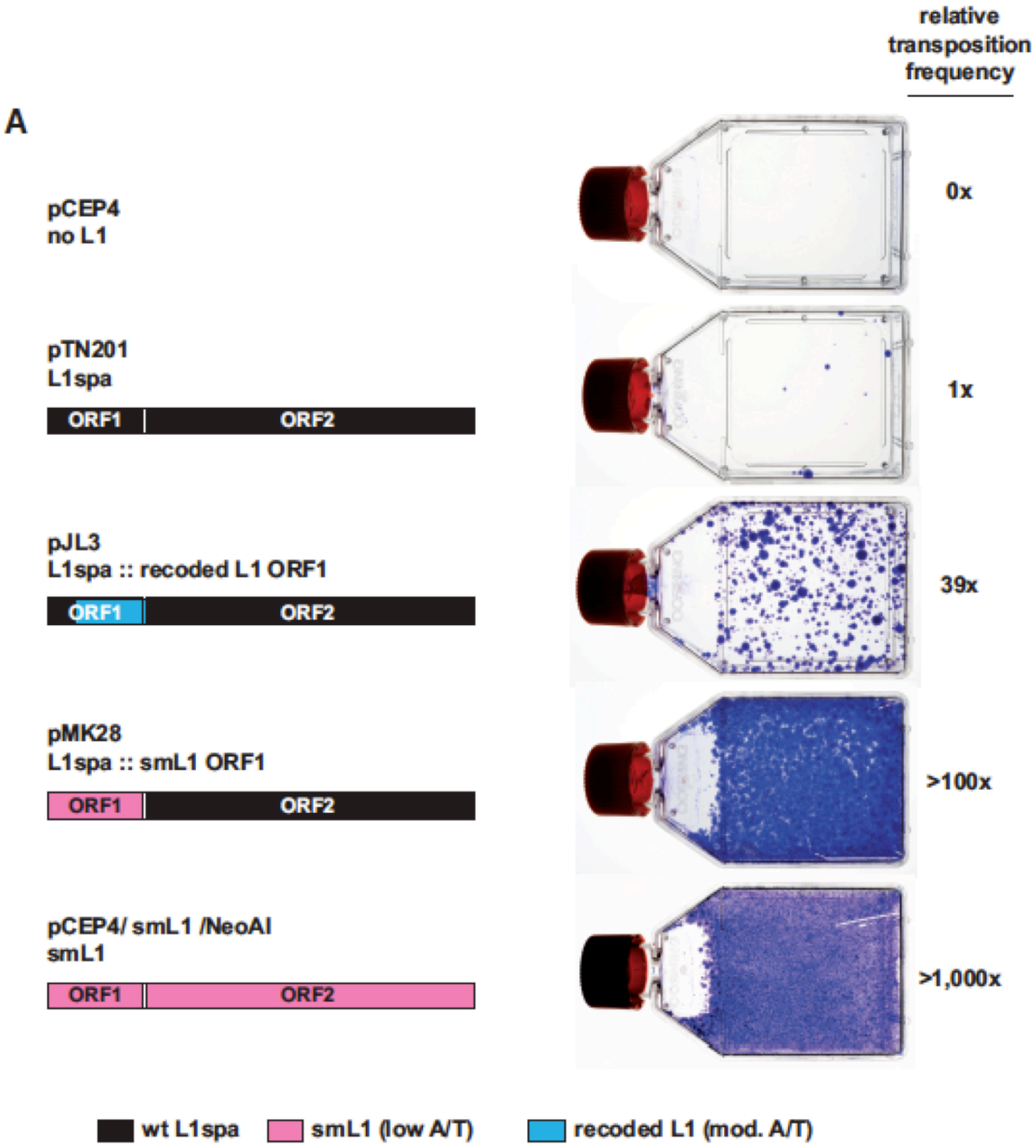


Figure 19 AS L1 transcription helps limit retrotransposition in cis

Native L1 ORF1 sequences in L1spa (black) were replaced either with a synonymously recoded fragment from smL1 with its markedly reduced A/T content (Han and Boeke, 2004) (pink), or a new recoded fragment that preserves A/T content more similar to that found in endogenous L1s (blue). Resulting marked L1 donors, i.e. pMK28 and pJL3, were assayed for retrotransposition by transfecting human HeLa cells. As controls, native L1spa (in pTN201), smL1 (in the same pCEP4 donor plasmid backbone and marked with NeoR/AI) and an empty donor plasmid (pCEP4) were transfected in parallel. Following selection on hygromycin, about one million Hygro^R cells were plated per flask, and new L1 integrants were selected for Neo^R, followed by staining of colonies. Retrotransposition frequencies are indicated relative to L1spa in pTN201 (right).

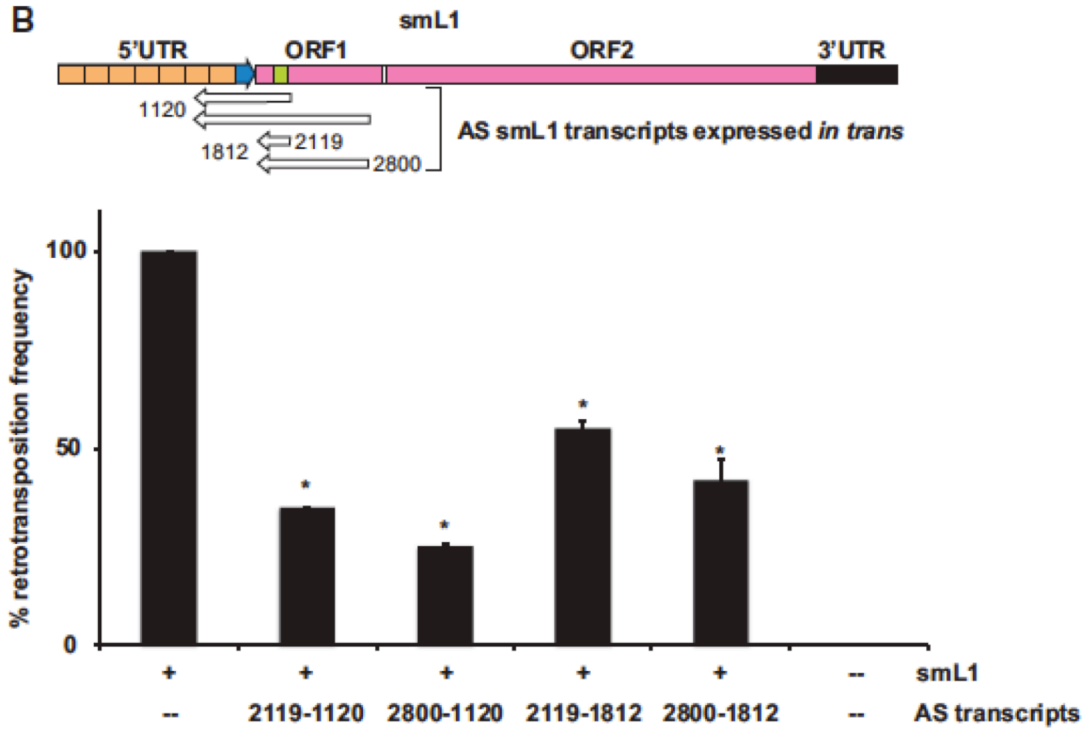


Figure 20 AS L1 transcription helps limit retrotransposition in trans

To measure the suppressive effects of overexpressed AS smL1 RIFTs on retrotransposition by smL1, first we directionally cloned four AS fragments from smL1, i.e. coordinates 2119-1120 (PCR amplicons DES3820xDES3818); 2800–1120 (DES3821xDES3818); 2119–1812 (DES3820xDES3819); and 2800–1812 (DES3821xDES3819) into pCEP4 downstream of its strong CMV promoter. Each cloned construct was co-transfected into HeLa cells with the smL1 retrotransposition donor plasmid, pCEP4/smL1/Neo. As positive and negative controls, smL1 donor alone and pCEP4 alone were transfected into HeLa cells, respectively. After transfection, cells were plated at various dilutions, selected on G418 for 2 weeks and Neo^R colonies were stained and counted. The mean and range of duplicate counts were determined, and retrotransposition frequencies were normalized relative to that of the smL1 positive control (defined as 100%).

Asterisks: significantly different from control retrotransposition frequency (two-tailed t-test, $p<0.05$ in all pairwise comparisons).

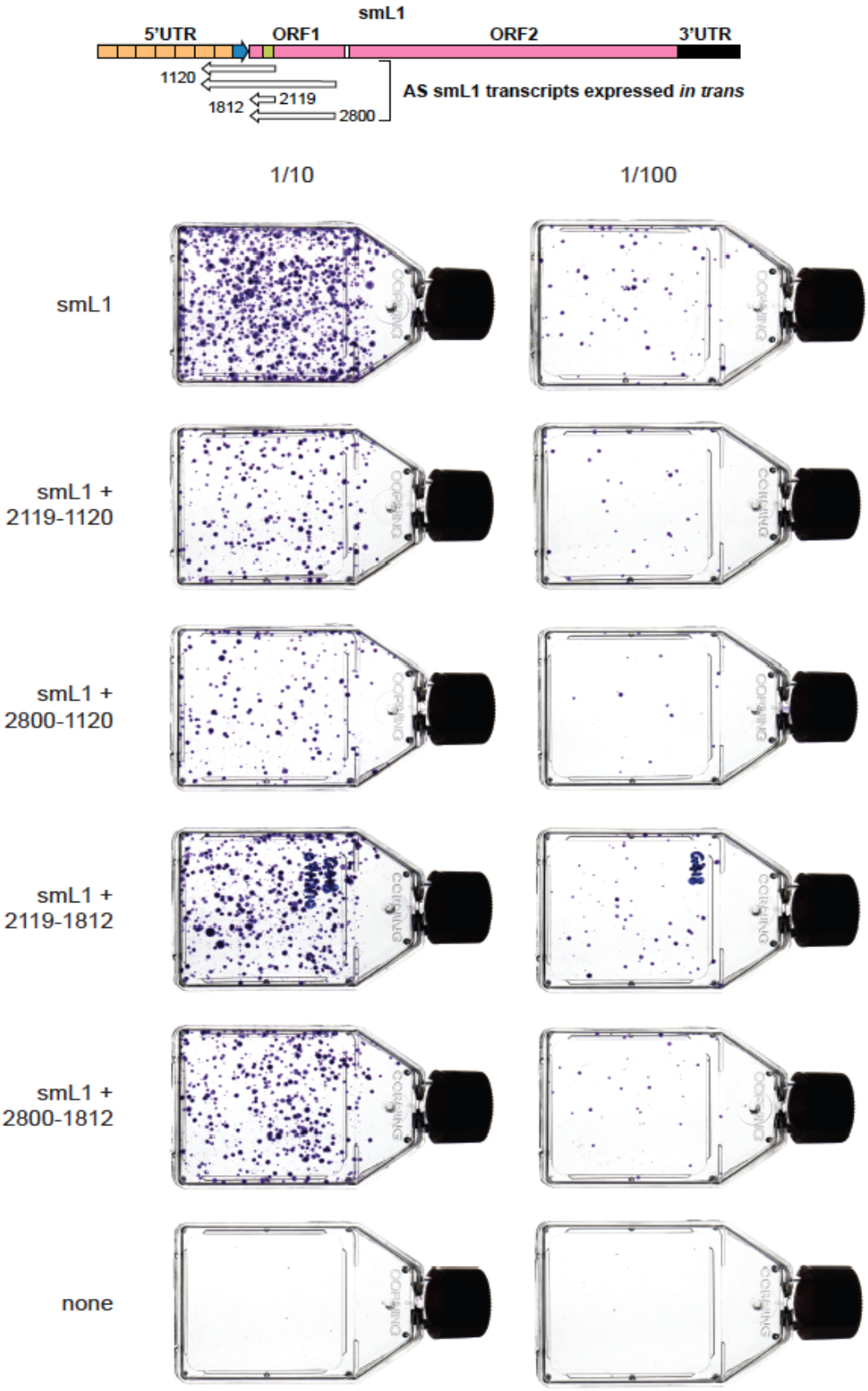


Figure 21 AS L1 transcription helps to limit retrotransposition of synthetic L1

We studied the effects of AS RIFTs (expressed in trans) on smL1 L1 retrotransposition frequencies. To generate AS smL1 RIFTs, first we directionally cloned four AS fragments derived from smL1 (schematic at top), i.e. mapping to corresponding L1spa coordinates 2119-1120 (PCR amplicon DES3820 x DES3818, Table S1); 2800-1120 (DES3821xDES3818); 2119-1812 (DES3820 x DES3819); and 2800-1812 (DES3821 x DES3819). These fragments were cloned into pCEP4 downstream of its strong CMV promoter. Each cloned construct was co-transfected into HeLa cells with the smL1 retrotransposition donor plasmid, pCEP4/smL1/Neo-AI. As positive and negative controls, the smL1 donor alone and empty pCEP4 alone were transfected into HeLa cells, respectively. After transfections, cells were plated at dilutions (i.e. 1x, 1/10, 1/100), selected on G418 for 2 weeks in a transient retrotransposition assay, and Neo^R colonies were stained and counted. We observed non-linear saturation of colony numbers at the 1x plating density (not shown), so only the 1/10 and 1/100 dilutions are shown.

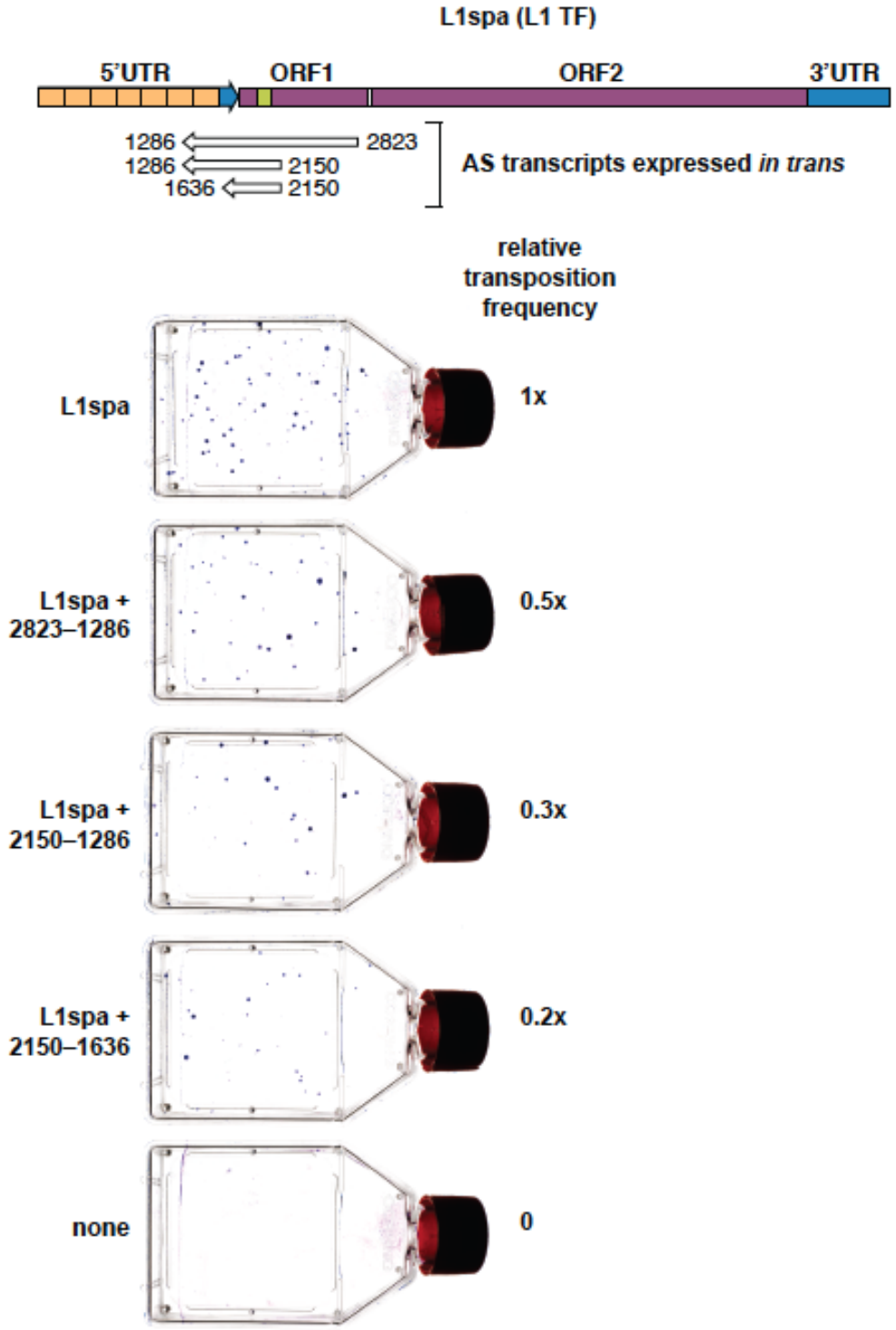


Figure 22 AS L1 transcripts limit retrotransposition of native L1 elements in trans.

We studied the effects of AS RIFTs (expressed in trans) on native L1 retrotransposition frequencies.

(Schematic at top) Three different AS native L1 T_F fragments (derived from L1spa template in pTN201) i.e. mapping to L1spa coordinates 2823-1286, 2150-1286 and 2150-1636, were amplified from L1spa using primers DES2879xDES2882). These fragments were directionally cloned into

pCEP4 downstream of its strong CMV promoter. Each cloned construct was co-transfected into HeLa cells with the native L1spa (L1 T_F) donor plasmid, pTN201. As positive and negative controls, L1spa and empty vector pCEP4 were transfected alone, respectively. After transfection, equal numbers of HeLa cells were plated, selected on G418 for 2 weeks, and Neo^R colonies were stained and counted.

3.3.6 Modest role of Dicer in limiting native L1 retrotransposition

We hypothesized that AS transcripts initiated from the AS promoter, expressed together with sense transcripts initiated from the conventional 5' promoter of mouse L1s, could result in the formation of double-stranded (ds) RNAs. In turn, these dsRNAs could trigger formation of short interfering RNAs or microRNAs through a Dicer-dependent pathway (Tam et al., 2008; Watanabe et al., 2008), thereby reducing sense strand L1 transcripts and limiting retrotransposition. We tested this possibility by using Dicer knockout cells in a retrotransposition assay. Because Dicer ex5 –/– HCT116 human colorectal cells are Neo^R (Cummins et al., 2006), we engineered novel L1 donors, marked with the β-lactamase TEM1 reporter interrupted by an artificial intron. Either native or hybrid recoded L1s were transfected into HCT116 Dicer ex5 –/– cells and control wild-type Dicer cells. After selection on donor plasmids, retrotransposition was assayed by qRT-PCR analysis of spliced TEM1 transcripts, expressed from new L1 insertions (Raiz et al., 2012).

The retrotransposition rate of L1spa, which contains an active AS promoter, increased slightly, i.e. <2-fold, in Dicer–/– cells compared with control cells. By contrast, retrotransposition by recoded elements lacking AS promoter activity, i.e. pJL3/TEM1, pMK28/TEM1 and pCEP4/smL1/TEM1, was essentially unchanged in Dicer–/– cells as compared control cells (Figure 23). Thus, Dicer played a modest role in suppressing native L1 retrotransposition, mediated by AS L1 transcription; most of the suppression by AS L1 transcripts occurred independent of Dicer. Previous experiments showed a similar ~2-fold level of suppression of human L1 retrotransposition on knockdown of Dicer in cultured cells. That result was interpreted as showing the role for Dicer-dependent RNA interference in regulating human retrotransposition (Yang and Kazazian, 2006).

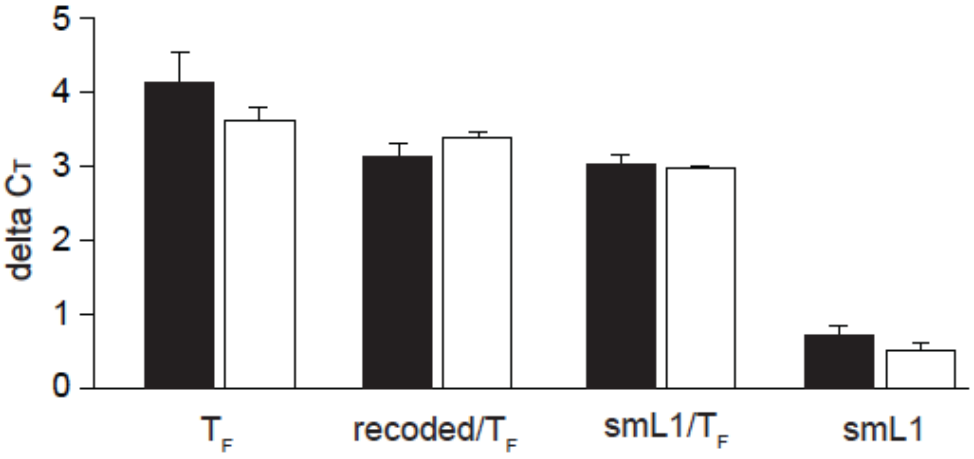


Figure 23 Minimal role of Dicer in regulating L1 retrotransposition

Retrotransposition of various L1 elements was assayed in the presence or absence of Dicer. We quantified spliced TEM1, expressed from retrotransposed genomic integrants, using qRT-PCR of spliced TEM1 transcripts. Y-axis: delta Ct = Ct (TEM1) – Ct (GAPDH). Higher Ct values correspond to lower transcript template concentrations. X-axis categories: TF, native L1spa from pTN201; recoded/TF, neutral changes in A/T content of recoded ORF1 in pJL3; smL1/TF, low A/T content in ORF1 swapped from smL1 into L1spa (L1 TF), resulting in a hybrid L1 donor, pMK28; smL1, fully synthetic L1 ORF1 and ORF2 in pCEP4/smL1.

Values, mean of duplicates; error bars, range; black bars, HCT116 wildtype cells; white bars, HCT116 Dicer ex 5 -/- cells.

3.4 DISCUSSION

A recent analysis of human and mouse transcriptomes suggested that 6–30% of all transcripts are initiated from repetitive elements (Faulkner et al., 2009). Here, we have identified and experimentally characterized an active initiator of such transcripts, i.e., an AS promoter within ORF1 of mouse L1 retrotransposons, present in thousands of full-length copies genome-wide, more than its human counterpart. It initiated a diverse range of fusion transcripts, as shown by > 100 distinct AS L1 RIFTs identified here and elsewhere (Akagi et al., 2008; Zemojtel et al., 2007). AS L1 RIFTs included spliced, unspliced and/or noncoding RNAs, and were readily detected in various mouse cell lines, tissues, developmental stages and strains (Figure 16). In addition to adding significantly to transcriptional diversity, AS L1 transcription helped to limit L1 retrotransposition (Figures 14–16).

3.4.1 Characterization of an AS L1 promoter and AS L1 RIFTs

The co-existence of a protein-coding sequence together with an antiparallel promoter activity in opposite overlapping orientations is unusual, but is not unprecedented, in mammalian genomes (Kampa et al., 2004; Lehner et al., 2002; Vanhee-Brossollet and Vaquero, 1998). Many sequence differences, particularly in the 5' UTR and within ORF1, distinguished the three active mouse L1 subfamilies, i.e., T_F, G_F and A elements (Figure 11). Several putative transcription factor binding sites in the AS promoter sequence of L1spa (Naas et al., 1998) and other T_F subfamily elements could be altered by natural sequence variants occurring in other L1 subfamilies (Figure 11). Although members of each subfamily retrotransposed recently, (Goodier et al., 2001; Naas et al., 1998; Saxton and Martin, 1998), these sequence differences simultaneously could affect both their distinct retrotransposition rates, by affecting ORF1p structures, and their AS promoter activities. We note that a single amino acid substitution in mouse ORF1p can affect L1 retrotransposition (Martin et al., 2008). In addition, the recoded synonymous sequences in ORF1 of pMK28 and pJL3 disrupted numerous predicted transcription factor binding sites in the AS promoter (Figure 11), consistent with a complete lack of AS promoter activity observed in those elements.

The various AS promoter activities associated with each L1 subfamily (Figure 9) were roughly proportional to the number of RIFTs initiated by them *in vivo*. Thus, we concluded that AS L1 promoter activities ranked as L1 T_F >> G_F ~ A > F. Notably, the latter subfamilies possessed modest, but detectable, AS promoter activities.

Estimated ages, counts and retrotransposition frequencies of L1 subfamily members have varied considerably. The average ages of L1 T_F elements range from 0.25 to 1.23 million years, and numbers of full-length insertions range from 3400 to ~4800, whereas active and/or polymorphic T_F elements ranged from ~1900 to 3000 (DeBerardinis et al., 1998; Sookdeo et al., 2013). The average ages of L1 G_F subfamily members have been estimated at 0.75 to 2.16 million years. Full-length G_F element counts have varied from 704 to 1500 (Goodier et al., 2001). There are ~400 to 535 active and/or polymorphic L1 G_F elements. The average ages of the youngest L1 A subfamily members have been estimated to range from 0.21 to 2.15 million years, and older A subfamilies have also been identified. Full-length A elements have ranged in number from 3400 to 6500 (Saxton and Martin, 1998). There are ~900 to 1600 active and/or polymorphic L1 A insertions. Individual elements of all three subfamilies have been shown to retrotranspose at comparable frequencies.

These findings prompted us to consider an apparent paradox. How might T_F subfamily elements harbor the strongest AS promoter activity, even though they have accumulated to some of the highest copy numbers of full-length L1 integrants in the genome (Akagi et al., 2008; Sookdeo et al., 2013)? We speculate that more robust host defenses might be necessitated by elements with increased retrotransposition potential, thereby resulting in relatively equivalent mobilization frequencies of distinct subfamily elements (Goodier et al., 2001). This paradox could also be explained by comparing the long evolutionary times over which different subfamilies have accumulated, moving in germ line tissues under negative selection (Sookdeo et al., 2013), versus the expression of AS L1 RIFTs in germ line and somatic tissues, measured in real time.

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. Considered together with results from 5' RACE analysis (Figure 17) and in vitro promoter assays (Figure 9), these findings clearly established that many diverse RIFTs were expressed from AS promoters located in L1 ORF1 *in vivo*.

In this study, although we detected both sense and AS L1 transcripts expressed in the same tissues, including testis and thymus, we have not tested whether sense and AS L1 promoters may be active simultaneously in single cells. If they are not, the resulting unbalanced expression of sense versus AS L1 transcripts in distinct cells or tissues could allow

particular L1 elements to evade this putative defense mechanism. Moreover, individual mouse and human L1 elements can mobilize over a wide range of frequencies, despite similar ORF sequences shared by 'hot' versus 'cool' elements (Brouha et al., 2003; Lavie et al., 2004; Seleme et al., 2006). Although we found many diverse AS L1 RIFTs expressed, many were expressed at low levels, and many other potentially active, distinct L1 elements had no detectable AS RIFT expression.

Many additional AS L1 RIFTs might have been missed in our study, owing to a lack of saturation of our screens: a limited range of mouse tissues and lineages used in the various screens, low expression levels, and/or strict criteria imposed in our RIFT assay. Even so, after summing up all AS L1 RIFTs observed by various methods, we conclude that the robust AS L1 promoter activity characterized here still does not account for most of the 6–30% of all transcripts initiated from transposons in mouse (Faulkner et al., 2009). A possible explanation is that there exist other promoters inside or outside of TEs that initiate such abundant transcription. We are currently working to identify such potential promoters, but to date, no experimental evidence for them has been reported. Alternatively, this reported range of 6–30% could dramatically overestimate actual TE-initiated transcription. Our phage library screens revealed ~0.03 to 0.1% of all transcripts hybridized with an L1 ORF2 probe (Figure 9), far less than identified from CAGE tags (Faulkner et al., 2009). In addition, recent studies in mouse embryonic stem cells identified most L1-specific small RNAs mapping to both strands of the L1 5' UTR and proximal ORF1, but not ORF2 or the 3' UTR (Chow et al., 2010).

The presence of a particular full-length L1 element was necessary, but not sufficient, to initiate a locus-specific AS L1 RIFT. We found that only 13 (19%) of 68 polymorphic full-length L1s initiated AS L1 RIFTs in testis, as assayed by RT-PCR. Moreover, some RIFTs only were expressed in embryonic, newborn or adult mouse testis, whereas smaller numbers were expressed in other organs such as brain and kidney (Figure 16). A few AS L1 RIFTs were expressed in several tissues. We speculate that the determinants of variable initiation of RIFTs by various L1s across the genome may include position effects, neighboring transcription units, other nearby genomic features, tissue-specific factors and/or variable chromatin marks (Akagi et al., 2013). Alternatively, certain L1 integrants could undergo differential, transcriptional gene silencing in situ (Garcia-Perez et al., 2010).

3.4.2 Biological roles of AS L1 RIFTs

L1 elements move in germline cells and brain during normal development, and in aberrant disease states such as cancer. Because antisense promoters (including many polymorphisms) are inherently part of many such L1 integrants mobilized, they could contribute substantially to natural transcriptional variation distinguishing between lineages, individuals and even cells (Akagi et al., 2008). In addition to the robust level of AS L1 RIFT expression at Arhgap15 (Figure 15), we previously reported comparably robust levels of AS L1 RIFT and native transcripts at *Drosha*, as shown by northern blot (Akagi et al., 2008). However, aside from these cases, most other mouse AS L1 RIFTs appear to be expressed at low levels, as in human (Rangwala et al., 2009). Further experiments are needed to quantify and compare RIFT expression levels versus long noncoding RNAs (Kung et al., 2013), microRNAs and other biologically significant transcripts.

Antisense L1 RIFTs frequently can be expressed from non-polymorphic L1 integrants in diverse mouse lineages (Figure 16), implying that at least some may share a conserved, albeit unknown, biological function. Certain expressed RIFTs may play several distinct biological roles including possible protein translation. In some cases, the predicted protein-coding ORF sequences of AS L1 RIFTs match the cognate ORF in transcripts from the associated native genes, suggesting that although they may encode identical proteins, their expression patterns may be added to, or modified by, the AS L1 promoter. Other AS L1 RIFTs may modify or replace cognate protein structures or expression, generate novel proteins or long noncoding RNAs (Guttman et al., 2010; Mourier and Willerslev, 2009; van de Lagemaat et al., 2003b) or introduce different 5' UTR sequences that could alter translational regulation. Transcripts that are AS to canonical sense transcripts could play other roles including degradation of sense strand transcripts through RNA interference or Dicer-independent mechanisms, variable compartmentalization and/or effects on transcript splicing and termination, RNA editing and translation (Conley et al., 2008; Lehner et al., 2002; Yu et al., 2008).

We also found that AS L1 transcription limited L1 retrotransposition, as demonstrated both by altered L1 transcript levels (Figure 18) and mobilization on synonymous recoding of the AS L1 promoter in ORF1 in cis and upon overexpression of AS L1 RIFTs in trans (Figure 18 and Figure 19). Hybrid L1s, containing either a recoded synonymous ORF1 segment from smL1 with decreased A/T content (Han and Boeke, 2004) or a second recoded ORF1 segment with neutral changes in A/T content, exhibited higher rates of retrotransposition than that of

native L1spa (Figure 19). The native AS L1 promoter could inhibit L1 retrotransposition in cis by triggering transcriptional interference, i.e. convergent, bidirectional transcription (Eszterhas et al., 2002). Expression of AS L1 transcripts alternatively could result in formation of double-stranded (ds) RNA molecules that could affect chromatinization and silencing of the L1 template (Yu et al., 2008) or trigger an interferon response (Daly and Reich, 1993). Such dsRNAs could form substrates for processing to small inhibitory RNAs through Dicer-dependent (Watanabe et al., 2008) or -independent mechanisms (Yu et al., 2008).

Interestingly, a modest number of ~23-nt small RNAs that map to the mouse L1 5' UTR region recently were identified in testis and in full-grown and meiosis I oocytes (Watanabe et al., 2006). In addition, both sense and AS small RNAs, mapping to the 5' end of mouse L1 elements, have been identified in mouse ES cells (Chow et al., 2010). Thus, both human and mouse L1 retrotransposition can be inhibited by RNAi in various cellular contexts.

We showed that Dicer played a modest <2-fold role in suppression of endogenous mouse L1 elements (Figure 23), similar to the results reported earlier for human L1s (Yang and Kazazian, 2006). We found that retrotransposition of pJL3/TEM1 was higher than that of pTN201/TEM1, even without Dicer. For this reason, we believe that the RNAi pathway is not likely to be the predominant suppressive mechanism of mouse L1 elements, and that other suppressive mechanisms are involved, at least in the differentiated somatic cells tested here. Thus, we conclude that AS L1 transcripts act mostly independent of Dicer in decreasing L1 expression and retrotransposition.

In summary, based on the work described in this Chapter 3, we conclude that mouse L1s encode a built-in mechanism that regulates them and alters expression of neighboring genes. We note a similar organization of bidirectional promoters exists in most other classes of autonomous mammalian retrotransposons, including human L1s and mouse and human LTR retrotransposons (Domansky et al., 2000; Ferrigno et al., 2001; Lunyak et al., 2007; Medstrand et al., 2001; Watanabe et al., 2008). Interestingly, bidirectional transcription at a particular mouse SINE B2 element was found to help establish an insulator or boundary element that, in turn, is critical to the developmental regulation of a neighboring gene (Lunyak et al., 2007). The evolutionary implications of such self-antagonizing promoters may be that transposons, including mouse L1 retrotransposons, can thereby limit their own expression. This would reduce their deleterious effects and costs to the fitness of their host (Boissinot et al., 2006), while modifying and diversifying the structure, expression and control of many other genes (Conley et al., 2008).

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

4.1 BACKGROUND OF THE STUDY

There are three active classes of retrotransposons in the human genome – LINE L1, *Alu* and SVA elements. The most abundant class of autonomous transposons in mammals – L1, encodes proteins that are not only able to mobilize its own self, but also help other classes of non-autonomous transposons (*Alu* and SVA elments in humans, and B1 and B2 elements in mice) and processed pseudogenes mobilize in the mammalian genome.

While L1-derived sequences constitute almost 17% of the mammalian genome (Richardson et al., 2015) and there are active L1s found in mouse and man, the mobility of L1 is tightly regulated by various cellular mechanisms, depending on the type and developmental context of the cell. LINEs can move only in those contexts when these controls are relaxed – in the germline (Ostertag et al., 2002), during the early stages of embryonic development (Kano et al., 2009; van den Hurk et al., 2007), in embryonal carcinoma cells (Martin and Branciforte, 1993), and in neural progenitor cells (NPCs) in the brain (Muotri et al., 2005). Detecting such *de novo* transposition events in the genome is like the proverbial finding the needle (new insertion) in the haystack (endogenous pool of genomic L1s), facilitated partly by the presence of the reporter that tags the L1. However, use of high-throughput techniques is required to detect and accurately determine the retrotransposition frequencies of natural transposition events *in vivo*. While some of the early attempts to determine the frequency of retrotransposition estimated it as 1 in 20 live births (Kazazian, 1999), more recent efforts, based on whole-genome sequencing have estimated it to be 1 in 100 to 1 in 150 births (Ewing and Kazazian, 2011).

As mentioned in Chapter 1, L1 mobilization is kept under control by various cellular mechanisms, including genetic and epigenetic mechanisms (Ostertag and Kazazian, 2001a; Richardson et al., 2015). Cytosine methylation is a key epigenetic regulatory mark that is localized predominantly within endogenous L1 retrotransposons and other transposable elements in mammalian genomes. It has been strongly associated with their transcriptional silencing and regulation, and may contribute to gene disruption (Yoder et al., 1997). These repressive marks are mitotically and meiotically heritable, and in general are stably maintained. In normal somatic cells, L1 retrotransposons are heavily methylated at CpG

dinucleotides, but in most cancers they become hypomethylated, potentially resulting in increased transcription and mobilization. Whether their increased activity directly leads to tumorigenesis is not yet clear.

In classic examples of epigenetic silencing of TEs in mammalian tissues, variable epigenetic marks (e.g., density of methylcytosine residues) at pre-existing, integrated endogenous retroviruses (ERV) can result in a high degree of variability in gene expression and the resulting phenotypes in genetically identical siblings (Ekram et al., 2012; Morgan et al., 1999; Rakyan et al., 2003). The classic concept of variegation involves the relatively unstable transmission of chromatin modifications and the "epigenotype" of a gene from a single cell to its daughter cells. Thus a differential pattern of gene expression can arise from the initial status. Variable expression of integrated transgenes has been well studied, and has been attributed to position effects and strain-specific modifiers.

Retrovirally transduced reporter genes typically are methylated rapidly upon integration in mammalian cells (Pannell et al., 2000; Rival-Gervier et al., 2013). Such silencing was linked to the source and sequence content of the reporter genes themselves. The expression of transgenic genes such as bacterial lacZ or green fluorescent protein (GFP) can be silenced by repressive epigenetic controls including histone deacetylation, histone methylation and/or cytosine methylation (Goll et al., 2009; Rakyan et al., 2002). The order of establishment of these marks has remained unclear, although several possibilities exist (Lund and van Lohuizen, 2004). Unlike retrovirally integrated reporters, whose target sites are likely determined by the mechanistic process of viral integration, transgenic reporters probably are "randomly" integrated in genomes. The latter integrants are frequently observed in concatemers. This tandem pattern may influence the silencing marks established at such transgene integrants by facilitating pairing interactions between adjacent elements, as described in plants (Luff et al., 1999; Melquist et al., 1999).

Recently, epigenetic transcriptional silencing based on histone tail modifications, of a marked LINE L1 in human embryonal carcinoma (EC) cells was reported (Garcia-Perez et al., 2010). In this study, the authors show that a full-length L1 de novo integrant (mobilized from a plasmid construct that had L1.3 marked with GFP) that was silenced in PA-1 EC cells was reactivated when treated with a histone deacetylase inhibitor (HDACi). They ruled out the involvement of DNA cytosine methylation because treatment with 5-aza-dC had no effect in reactivating the L1 reporter integrants, and further established the role of histone modification in epigenetic silencing by chromatin immunoprecipitation (ChIP) experiments.

The authors then correlated this silencing/reactivation phenomena with differentiation pathway, showing a ~40-fold reactivation of the L1 integrants in cells made to differentiate, as compared to controls. The authors concluded that silencing of L1 integrants is more efficient in EC cells than in differentiating cells. What is interesting about this experiment is that the L1-integrants behaved like viral sequences in terms of the silencing being attenuated in differentiating cells, as the authors note. However, the kinetics of L1-reporter silencing was different from the kinetics of retroviral-based silencing. This observation somewhat supports the idea that marked L1 reporter cassettes may not accurately reflect the actual dynamics of native L1 silencing, because one would expect native L1s to be active in the undifferentiated state and turned off as differentiation proceeds, unlike in viruses wherein reactivation of silenced elements in differentiated cells seems to be a more logical consequence after the virus has invaded.

In our study, reported in this chapter, 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. To assess the possibility that new reporter integrants could be differentially expressed or silenced, we constructed novel, real-time reporters for L1 retrotransposition, whose expression levels would not be subjected to positive or negative selective pressures imposed on the cells. 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). This greatly exceeds sensitivity of other "real time" reporters used in retrotransposition assays. As a second, convenient reporter, we chose green fluorescent protein (GFP) (Ostertag et al., 2000) for particular assays. Mimicking the design of other retrotransposition reporter constructs, we introduced the AI into donor cassettes to disrupt the *TEM1* or GFP open reading frames (ORFs), respectively. We report variable expression of the L1 reporters in clonal populations of cells, a phenomenon we term "oscillation", owing to differential expression of the marker gene due to variable silencing.

Using this reporter system, here we show how different cellular contexts and mechanisms of transposition may impact epigenetic silencing of new insertions. We observed variable repressive epigenetic controls established at L1 reporter genes that were newly mobilized in different genomic, cellular and developmental contexts. 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. 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 hypothesize 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*, a DNA transposon, 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 cytosine methylation was experimentally reduced. We conclude that the host cellular and developmental contexts of L1 retrotransposition are significant determinants of the epigenetic controls established at new somatic integrants.

4.2 MATERIALS AND METHODS

4.2.1 Plasmids used in this study

The donor plasmid pDES46 was constructed by Dr. David Symer as follows: the BamHI site in ORF2 of L1.3 sequence was deleted by site-directed mutagenesis and was moved into pBSII-KS using NotI and BamHI. Bst1107I -HSVTKpolyA-MluI oligo was then introduced to remove neo-intron and HSVtK promoter fragment. BseRI sites were introduced to flank both TEM1 β-lactamase gene (that was obtained from vector pBLAK-b and then codon optimized) and β-globin artificial intron (AI) by fusion PCR. Both these constructs were cut using BseRI and ligated together seamlessly. The entire L1.3/TEM1-AI construct was then dropped out using NotI and BamHI and cloned into pCEP4 (Invitrogen) to give pDES46 (Figure 24). The *piggyBac* donor plasmids for mobilizing reporters were kind gifts from Dr. Allan Bradley (Sanger Institute, UK) and Dr. Tian Xu (Yale University, USA) (Ding et al., 2005). The plasmids were later modified to accommodate the TEM-1 β-lactamase as a reporter, instead of GFP flavours they came with.

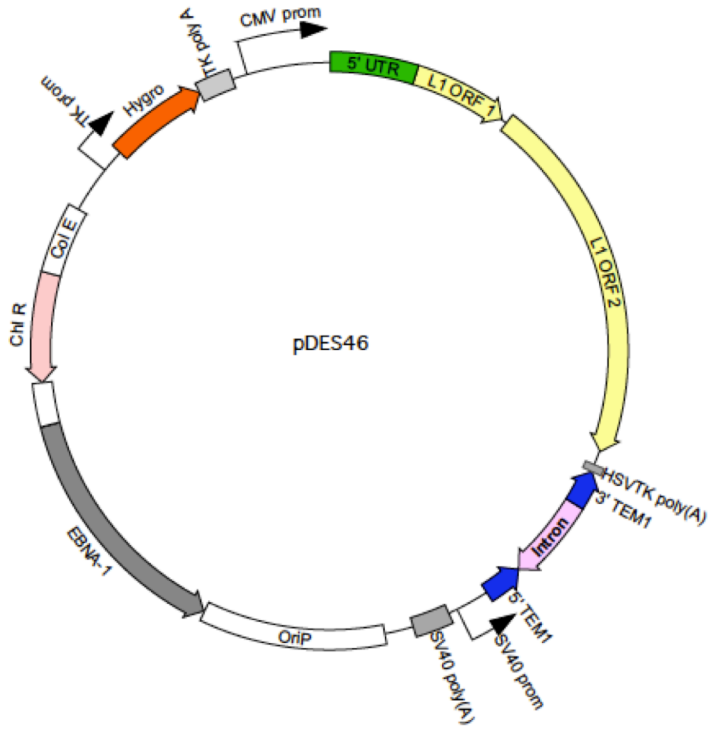


Figure 24 Schematic of a human L1 retrotransposon donor plasmid, pDES46

Human LINE L1.3 was tagged at its 3' end with a highly sensitive reporter gene, β-lactamase (TEM1; blue open read frames) (Li et al., 2014), interrupted by an artificial intron (AI; pink). This L1 donor construct, based on the pCEP4 episomal plasmid, was stably maintained on Hygro selection. Upon L1 mobilization, expression of real-time β-lactamase reporter (encoded by the spliced, integrated TEM1 gene) was assayed for.

4.2.2 Cell lines and mouse tissues

HCT116 (human colorectal carcinoma) cells, kindly provided by Drs. Ina Rhee, Christoph Lengauer and Bert Vogelstein (Johns Hopkins University), were cultured in McCoy's 5A modified medium (Gibco, Life Technologies) supplemented with 10% heat inactivated fetal bovine serum and 1% penicillin-streptomycin, at 37 °C and 5% CO_2 in a humidified chamber. HeLa.JVM (a subclone of human cervical carcinoma) cells, provided by Dr. John V. Moran (University of Michigan), were cultured in Dulbecco's Modified Eagle Medium (DMEM) with the same supplements as those used for HCT116 cells.

4.2.3 Isolation of PMEFs and making feeders for mouse embryonic stem cells

Mouse ES cells need feeder cells to grow on. Fibroblasts, from day 12-13 mouse embryos, are cultured and growth-arrested (using Mitomycin C or γ-irradiation) to obtain feeders. It is best to start growing ES cells on the feeder plates within 2-3 days after the feeders have been plated. The feeders start detaching from the surface 8-10 days after being plated; this was kept in mind when planning for the experiment.

Pregnant mice were obtained from the animal house when the embryos were day 12 of pregnancy. They were sacrificed and the uterus with the embryos dissected out. Each embryo was removed from the uterus and separated from the yolk sac and placed in sterile 1X PBS in a 100mm dish. The embryos were washed one more time with PBS to remove any traces of blood. Under a dissection microscope, the head of each embryo was cut and discarded. As far as possible, the heart and liver tissues were also removed (these are two organs that wouldn't contribute fibroblasts to culture). It was made sure that any red tissue in the embryo was removed. The embryos in PBS were passed through a syringe exactly three times. The suspension of cells and tissues were cultured in a 150 mm dishes, using one dish per 5 embryos in DMEM+10%FBS at 37°C in an incubator with 5% $C0_2$. After the cells reached confluence, they were trypsinized and split 1:5. These were passage 1 (P1) PMEFs. P1 PMEFs were passaged one more time, up to P2 PMEFs. They were allowed to grow for 2-3 days to reach complete confluence. It is good to keep the passage number of PMEFs when making feeders between 2 and 3. Later passages may result in poor quality of feeders. Use of PMEFs after P5 was strictly avoided. Media from the confluent 150 mm dishes was removed and replaced with DMEM with 10μg/ml of Mitomycin C. The dishes were incubated in the cell culture incubator for 4 hours. Incubation time between 3.5-4 hours is optimal to give complete inactivation of the PMEFs. The medium was removed from the dishes. The cells

were rinsed with 1X sterile PBS 3-4 times to remove traces of the medium. Required numbers of feeder dishes were made, based on the cell growth surface area. Typically, to thaw out a vial of ES cells, based on the pellet size, either a 60mm dish (small pellet) or a 100mm dish (large) was used. Since the concentration of feeders is important for successful culturing of ES cells (if it is very high, the feeders may "curl up" as a sheet from the plate after a few days of growth, and if too sparse, they may not provide sufficient growth substratum for the ES cells, and may not allow them to be maintained in undifferentiated state), the table provided here was used to plate the appropriate numbers of feeder cells. Although some protocols call for treating the dishes with 0.1% gelatin prior to plating feeders, it does not seem to be absolutely necessary.

Size of dish/plate	**Approximate area* (in cm^2)**	**Number of PMEFs***
150mm dish	148	N/A$^{\xi}$
100mm dish	55	2.5-3 x 10^6
60mm dish	21	1 x 10^6
6-well plate	9.5	4 x 10^5
12-well plate	3.8	2 x 10^5
24-well plate	1.9	1 x 10^5
48-well plate	0.95	5 x 10^4
96-well plate (flat bottom)	0.32	2 x 10^4

* Indicates the number per well, in case of multi-well dishes

$^{\xi}$ Feeders are usually plated in dishes smaller than the 150 mm dish

When plating cells in a multi-well dish, the total number of cells for the whole dish was diluted in the appropriate volume of media before plating onto individual wells Notwithstanding the actual cell numbers, calculations based on growth area conversion was used for plating. For instance, one confluent 150mm dish of feeders when plated onto three 100mm dishes gave optimal surface growth. Given below are the calculations that were commonly used:

1 x 150mm dish = 3 x 100mm dishes 7 x 60mm dishes

1 x 100mm dish = 3 x 60mm dishes 2 x 6-well plates

1 x 6-well plate = 1.25 x 12-well plates

1 x 12-well plate = 1 x 24-well plate 1 x 48-well plate

The feeder plates were incubated at 37°C, 5% CO_2 for two days and observed for any signs of growth (multiplication). Incomplete treatment with mitomycin C may result in growth of the MEFs and concomitant change in color of the culture medium. If this is the case, then the plates were discarded, since active PMEFs would have easily outgrown the ES cells.

4.2.4 Mouse Embryonic Stem Cells and their culture

The following ES cells and feeders were used during the course of the study:

Cells	Parent	Source
E14Tg2a.4	129P2 ES cells	Bay Genomics
Bruce4 ES cells	C57BL/6	Dr. Lino Tessarollo's lab, NCI-Frederick
Truck_305 ES cells	Bruce4 ES cells	Dr. Jef Boeke's lab, Johns Hopkins University
neo2 Primary Mouse Embryonic Fibroblasts (PMEFs)	N/A	Ms. Terry Sullivan at NCI-Frederick; also isolated from mice

E14Tg2a.4 mouse embryonic stem (ES) cells were cultured without feeder cells in Glasgow's Modified Eagle's Medium (Sigma Aldrich) supplemented with 10% ES cell-qualified FBS, 1% non-essential amino acids, 1% L-glutamine, 1% sodium pyruvate, 0.1mM 2-mercaptoethanol and ESGRO (Millipore) at 1000U/mL in 7% CO_2 atmosphere. Bruce4 and Truck_305 ES cells were grown in high-glucose DMEM supplemented with 15% ES-cell-qualified FBS, 1% L-glutamine, 1% non-essential amino acids, 0.1mM 2-mercaptoethanol, 50U/ml of Penicillin-streptomycin and ESGRO in 10% CO_2. *Neo2* mouse embryonic fibroblasts (MEFs), used as feeder cells for Bruce 4 and derivative ES cells, were arrested (using Mitomycin C or γ-irradiation) cultured in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin in 5% CO_2.

Chimeric founder and pseudofounder mice were generated from Truck_305 ES cells as described. Briefly, mES cells were microinjected into developing blastocysts that were delivered into the uterine horns of pseudopregnant female mice. Chimeric founder mice were identified by PCR amplification of the L1 donor construct. Pseudofounders were identified by PCR amplification of spliced reporter integrants.

A frozen vial of ES cell line was dipped into a 37°C water bath till the cells were seen almost thawed. In the laminar flow hood, the cells were transferred into a sterile 15ml conical tube. About 10ml of pre-warmed M15 medium (without LIF) was added drop by drop, swirling the tube every few seconds (Note: this method of thawing results in increased viability of

cells). The cells were spun down at 1,000 rpm for 5 min. After removing the supernatant medium, the cell pellet was briefly flicked to loosen up. Then, depending on the dish to be plated on, the pellet was re-suspended in either 5ml (for 60mm dish) or 10ml (100mm dish) of pre-warmed M15+LIF. The suspension was plated and the dish incubated at 37°C, 10% CO2. The culture medium was changed once every 2 days or when discoloration was seen. Although most ES cell culture protocols require medium be changed every day, for Bruce4 cells, this does not seem to be necessary. After the colonies reached about 80% confluence, the cells were split 1:3 or 1:4. Usually, the cells are ready for splitting 3 days after plating. If left longer in the same dish, many of them start dying and some differentiate. This was avoided.

4.2.5 Cell transfection

Fugene 6 (Roche Applied Science) was used to transfect cells. Cells were transfected with pDES46 were selected on hygromycin (Invitrogen) at 0.3 mg/ml for 2 weeks. Cells were cloned by limiting dilution into 96-well plates and observed for single colonies. Colonies were assayed for TEM1 β-lactamase using CCF2/AM assay. Selected colonies were propagated in larger plates and frozen down (by trypsinizing and spinning down cells from typically one T-75 or two T-25 flasks, resuspending the pellet in 0.5 ml of fetal bovine serum and then adding in a drop wise manner 0.5 ml of freshly made FBS+20% DMSO (Sigma Aldrich, USA) mixture; in other words, cells are resuspended in FBS with 10% DMSO), as well cloned again by limiting dilution.

4.2.6 Cell Cloning by limiting dilution

Cells were trypsinized, neutralized and counted using a hemocytometer. Then, for making limiting dilution of cells, the original cell suspension was diluted appropriately in complete medium to give cell concentrations desired. These cell dilutions were then plated from a sterile reagent reservoir onto the corresponding 96-well plates, each well receiving 100μl of the diluted cells. The usual concentrations of dilution were: 0.1, 0.3, 1, 3.3, 10, 33, 100, 300 and 1000 cells per well. In some cases where the cell line had been cloned earlier, not all these dilutions were used. After the plates were incubated at optimal conditions in the cell culture incubator for 2-3 days, the wells were inspected for colony growth. Colonies became visible usually after 5-7 days of plating cells and were subsequently expanded into either 48-well dish, or two well of 96-well dish, and later into 24-well dish, 12-well dish, 6-well dish and then to cell culture flasks (T-25 and upward).

4.2.7 Reporter assay and flow cytometry

Cells were stained, visualized and analyzed in flow cytometer by the same procedures described in section 3.2.3. Each flow cytometry session included controls for blue and green cells – cell populations that were all green (untransfected cells stained with CCF2/AM) or all blue (variegating cells treated with HDAC inhibitor Trichostatin-A, 24 hrs. prior to the flow analysis).

4.2.8 Bisulfite treatment and DNA methylation analysis

About 1 μg of genomic DNA was subjected to bisulfite conversion, which deaminates unmethylated cytosines to uracils, while leaving the methylated cytosines intact. For this, the EZ DNA Methylation kit (Zymo Research, USA), the Qiagen Epitect kit (Qiagen), or a published protocol (Cheng et al., 2004) was used. The manufacturer's protocol was followed when using commercial kits. In the other procedure, the following protocol was followed: DNA to be modified (0.2-1.0 μg in total volume of 30 μl in water) was heated at 94°C for 1 min. and overlaid with 35 μl of mineral oil. To this, 15 μl of 0.9 M NaOH (final concentration, 0.3 M) was added and incubated at 42°C for 10 min. for further denaturation. Next, 2.25 μl of 100 mM hydroquinone (final concentration, 0.5 mM) and 402.75 μl of 3.35 M sodium bisulfite solution (final concentration, 3 M) were added, and the mixture incubated under mineral oil (~ 65 μl) at 55°C for 16-18hours. Afterward, the mineral oil was removed the rest of the product pipetted into Microcon 30 spin column (Amicon, Millipore). The mix was spun at 14,000 rpm for 10 min. The modified DNA was washed with 500μl of deionized water to remove excess sodium bisulfite by spinning at 14,000 rpm for 10 min. For eluting the DNA, 50μl of deionized water was add to the spin column and incubate at room temperature for 10 min. The column was then inverted into a fresh receiving tube and spun at 14,000 rpm for 10 min. to collect the DNA. The converted DNA was subjected to PCR amplification using stand-specific primers suitably designed for the changed base sequence in the target region. PCR bands were cut out of the gel, DNA extracted and cloned into TOPO TA cloning vector system (Invitrogen). After picking bacterial clones, each representing a strand of bisulfite converted DNA, and size-verifying the clones by PCR (using M13 forward and reverse primers present on the TOPO vector), the clones were sequenced by Sanger sequencing method using automated capillary sequencing, with M13 forward and reverse primers. Sequence reads were later analyzed using Sequencher program (GeneCodes Corporation).

4.2.9 Treatment with histone deacetylase inhibitors (HDACi)

Variegating cells at around 50% confluency, plated in 96-well plates, were treated with various concentrations of Trichostatin A (Sigma-Aldrich,USA). The growth media was removed, and cells fed with media containing TSA. After 24 h treatment, CCF2/AM assay was performed to look for phenotype changes. Other HDAC inhibitors reported in literature were titrated at various concentrations to determine any effect on the cells (all chemicals from Sigma-Aldrich): Apicidin (1.0-316.2nM); Scriptaid (0.5-40.0μM); Sodium Butyrate (0.1-100mM); Niacinamide (0.1-100mM). These concentrations of HDACi were determined after reviewing literature (Abbas et al., 2001; Ishihara et al., 2004; Shao et al., 2004) and performing kill curve experiments. For generating blue control cells for cytometric analysis, the oscillating HeLa cell line clone 5B 0.3c/w C9 was treated with 250nM TSA for 24 hours.

4.2.10 Mapping genomic integration sites by inverse PCR (iPCR) and Linear Amplification-Mediated (LAM) PCR

Genomic DNA was purified from single-cell clones, and digested with a restriction enzyme (RE) such as XbaI, EcoRI or HindIII. Upon heat inactivation of the RE, digested products were diluted to 1 ng/μL in a total of 500 μL, and incubated with T4 DNA ligase overnight at 16°C to allow intra-molecular ligation. After ethanol precipitation, DNA was resuspended and used in iPCR reactions using divergent primers DES682 and DES209 annealing to the 3' end of the retrotransposed cassette. In order to be able to pick even small integrants that are 5' truncated, a primer set that is located at the farthest end of the cassette was chosen (Figure 29). PCR reactions consisted of 40 cycles at 94°C for 30", 55°C for 30" and 72°C for 2'20". Each of the several bands obtained after PCR were cloned into pCR2.1 using TOPO cloning kit (Invitrogen) and transformed into bacteria. Bacterial clones that contained an insert were identified by colony PCR using M13 forward and reverse primers. The PCR products were cleaned up and sequenced. To map the de novo insertion sites, Sanger sequence reads were analyzed by alignment using Blat against the reference human genome (hg19).

LAM-PCR was used to map new L1 integrants in ES cells, using three clones (i.e. 1B6, 1C6 and 2B2) in which the retrotransposed, spliced reporter gene was present as shown by PCR assays. We set up linear amplification reactions including 50 ng gDNA from each ES cell clone, 2 nM dNTPs, 5 nM 5'-biotinylated primers DES3171 or DES3174, 1X Advantage 2 buffer (Clontech) and 1ul of Advantage 2 enzyme, for 50 cycles (20 s at 95C, 45 s at 60C, and 90 s at

68C). Streptavidin-coated magnetic beads (200 mcg) were washed twice in 100 μL of binding buffer (1 M NaCl, 5mM Tris, pH 7.5, 0.5 mM EDTA) using a magnetic separation stand, resuspended in 50 μL of 2X binding buffer, and mixed with the linear PCR reaction. The suspension was incubated for 60 min at RT under constant agitation, and then washed three times in 200 μL of wash buffer (10 mM NaCl, 5mM Tris pH7.5, 0.5 mM EDTA and 0.01% Triton X-100). For second-strand synthesis, the matrix-bound DNA was resuspended in 20 μL of a reaction mixture containing 500 nM dNTPs, 100 ng/μL random hexamers, 5 U Klenow enzyme (New England Biolabs) and 1X NEB Buffer 2, and incubated at 37C for 60 min. After washing first with wash buffer and then twice with 1X reaction buffer, the dsDNA was restricted using HaeIII or Sau3AI (NEB) at 37C for 2 hr, washed again in wash buffer followed by two washes in 1X ligation buffer, and ligated with either the HaeIII adapter (DES3177 and DES3178) or Sau3AI adapter (DES3177 and DES3179) using T4 DNA ligase at 16C overnight. To elute the products, beads were resuspended in 5ul of 0.1N NaOH and incubated at RT for 10 min. The eluate was separated from the matrix using the magnetic stand, and was neutralized by adding 5 μL of TrisHCl, pH 7.0.

To perform nested PCR, one microliter of the eluate was added as template in a 50 μL PCR reaction. After the first round of PCR, products were diluted 100-fold, from which 1 μL was used in the second (nested) round of PCR. Primers for the adapter (DES3181) and nested adapter (DES3182) were paired with DES3172 and nested primer DES3173 (HaeIII), or DES3175 and nested primer DES3176 (Sau3AI), respectively, in the donor plasmid. Products were cloned into the Topo-TA pCR2.1 backbone (Invitrogen). Sanger sequencing was performed on clones containing nested PCR products to map integration junctions.

4.2.11 Southern blotting analysis and probing using a TEM-1 probe

Southern blotting was done using capillary transfer of DNA to nylon membrane (Biorad), after digesting 10μg of DNA (per sample) with an appropriate restriction enzyme and running the digests on a 0.8% agarose gel at < 5 V/cm. The transfer was set up as described in standard protocol manuals and left undisturbed for ~24 hours.

After the transfer, the membrane containing the target DNA was floated on the surface of a tray of 6x SSC (or 6x SSPE) until the membrane became thoroughly wetted from beneath. The membrane was then submerged for 2 minutes. For pre-hybridizing the membrane by ‘roller-bottle’ method, it was gently rolled into the shape of a cylinder and placed inside a hybridization roller bottle. 0.1 ml of pre-hybridization solution was added for

each square centimeter of membrane and the bottle closed tightly. The hybridization tube was placed inside a pre-warmed hybridization oven at the appropriate temperature (68°C for aqueous solvents; 42°C for solvents containing 50% formamide; 65°C for phosphate-SDS solvents). The double-stranded DNA probe for the TEM1 gene, labeled with ^{32}P using random primer labeling method employed in a commercial kit (Amersham Megaprime DNA Labeling Systems, GE Healthcare), was denatured by heating for 5 minutes at 100°C and chilling the probe rapidly on ice.

Next, the pre-hybridization solution was poured out from the bottle and replaced with fresh hybridization solution containing probe. The bottle was sealed and replaced in hybridization oven. Incubate for the required period of hybridization, usually 6 hours. After hybridization, the membrane was washed twice with 2X SSC 0.1% SDS solution (15 min. per wash), followed by 1X SSC 0.1% SDS solution (twice @ 15 min. per wash), and lastly with 0.1X SSC 0.1% SDS solution (twice, 15 min each wash). In each wash, the roller bottle was filled with each of the buffers and allowed to rotate at 68°C for the specified time. After the washes, most of the liquid from the membrane was removed by placing it on a pad of paper towels. The damp membrane was placed and covered entirely using cellophane paper. The membrane was exposed to X-ray film for 16-24 hours at –70°C with an intensifying screen and then taken for developing to obtain an autoradiographic image.

4.3 RESULTS

4.3.1 A sensitive, real-time reporter for L1 retrotransposition reveals dynamic silencing of new genomic insertions

We first marked the full-length, retrotransposition competent, human L1.3 retrotransposon (RC-L1) with the novel *TEM1*-AI reporter cassette, as described earlier. The resulting construct was subcloned into a stable episomal vector, pCEP4, resulting in pDES46 - the donor plasmid (Figure 24). Cultured human cancer cells, namely, HCT116 and HeLa, were transfected with this plasmid. To assure stable episomal maintenance, the transfected cells were selected for Hygromycin resistance, the gene for which is expressed from pDES46. As a control to measure transfection efficiency, the same cell lines were transfected, in parallel, with pBLAK-c, a vector expressing GFP transiently. HCT116 cells routinely exhibited less GFP fluorescence when compared to HeLa cells.

To measure expression levels as a surrogate for active retrotransposition, we stained transfected cells with CCF2 fluorescent substrate and examined them by fluorescence microscopy (Zlokarnik et al., 1998). Some of the cells showed blue fluorescence, indicating the reporter gene was functional. We confirmed that retrotransposition was the mechanism of reporter integration, using PCR amplification across the spliced AI.

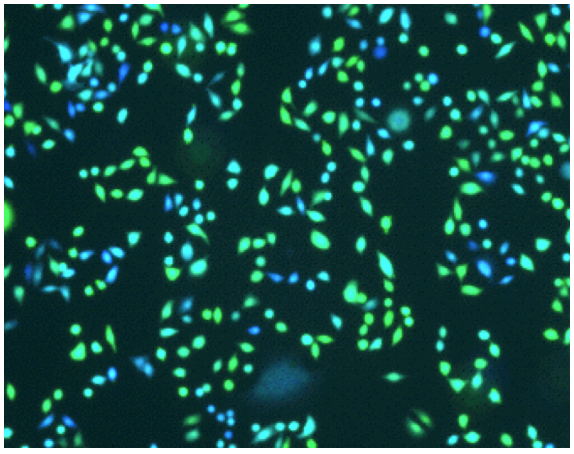


Figure 25 HeLa cells transfected with pDES46 show variegated phenotype

Fluorescence microscopy reveals wide-ranging levels of TEM1 expression, ranging from zero or low (green cells) to high (blue) levels. HeLa cells were transfected with pDES46, subcloned so they all contained de novo L1 reporter integrants, and stained using a fluorescent substrate for the beta-lactamase reporter, CCF2.

After selection on hygromycin for 10 days, bulk cell populations were screened for expression by treating them with CCF2. Fluorescence microscopy revealed that they appeared "variegated" (Figure 25). Some individual cells fluoresced bright blue, indicating robust expression of TEM-1 β-lactamase as the reporter for the occurrence of at least one retrotransposition event. Other adjacent cells appeared green, suggesting that they lacked an integrated reporter, or that any integrated reporter genes were silenced. Many other cells displayed intermediate shades of blue and green, implying partial silencing or less-than-maximal expression of the reporter. To quantify TEM1 reporter expression on a cell-by-cell basis, we used flow cytometry to analyze cells incubated with CCF2 (Knapp et al., 2003). As was the case with fluorescence microscopy, wide-ranging fluorescence emissions were observed in the cell population, ranging from green to blue (Figure 26). Blue/ green ratios were calculated to enhance the use of individual fluorescence signals as a surrogate for activity (Zlokarnik et al., 1998). These ratios ranged from <10 to well over 150, thereby demonstrating a large dynamic range over which was differentially expressed in individual cells.

To test heritability of reporter expression or silencing, we isolated individual cells from these mixed populations by limiting dilution, and grew up their progeny as subclones. Many such cellular clones were stained with the dye CCF2 and then visualized by fluorescence microscopy. Presumably, all the cells of one clone arose from a single cell that was obtained by limiting dilution. Very frequently, the resulting subcloned daughter cells still showed variegation of various shades of blue and green, indicating continued variability in reporter expression. Occasional colonies contained mostly blue cells, indicating high levels of reporter expression amongst most of the cloned daughter cells. However, even in such predominantly blue subclones, occasional green cells inevitably arose, indicating the stochastic and dynamic establishment of reporter silencing.

To check for the presence of a full-length TEM-1 β-lactamase reporter in the various clones, PCR was performed using primers that spanned the entire length of the reporter gene (Figure 27). While every clone that had blue cells showed a clear band corresponding to the full-length reporter (~800 bp) and the green clones did not show the presence of this band, all clones were positive for the presence of the donor plasmid used for transfection, as indicated by a faint band exactly matching the size of the donor plasmid control (Figure 27). (An extra band smaller than the full-length reporter band, as seen in all blue clones, was found to be a ~100 bp truncation in the reporter gene, as revealed after sequencing the integrants.

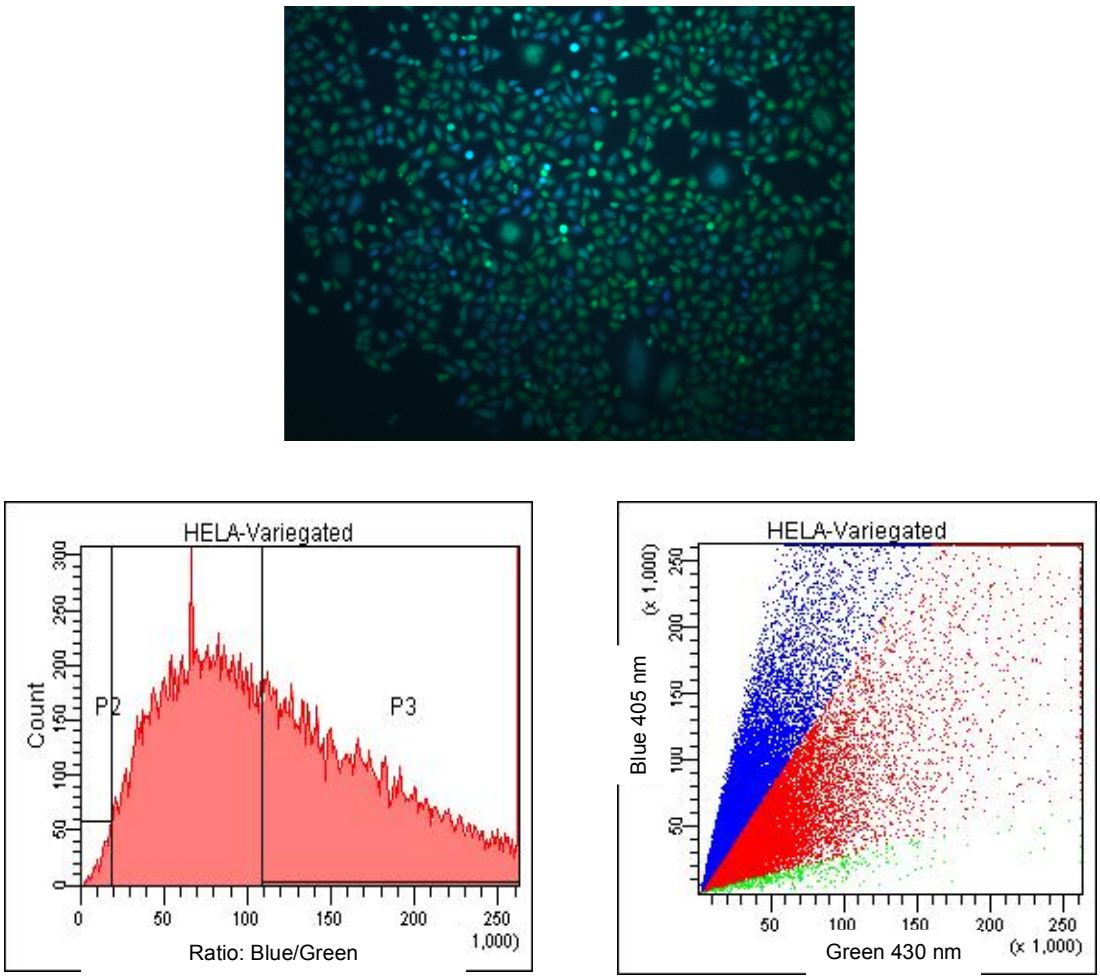


Figure 26 FACS results of the variegating cell line clone

The top panel shows the micrograph of a variegating cell population, viewed after staining with CCF2. Scatter plot from flow cytometry, performed on a subclone of cells harboring L1 reporter integrants. Fluorescence emissions were detected for (y-axis, 405 nm emission) blue and (x-axis, 430 nm) green individual CCF2-stained cells, as well all intermediate expression levels (red).

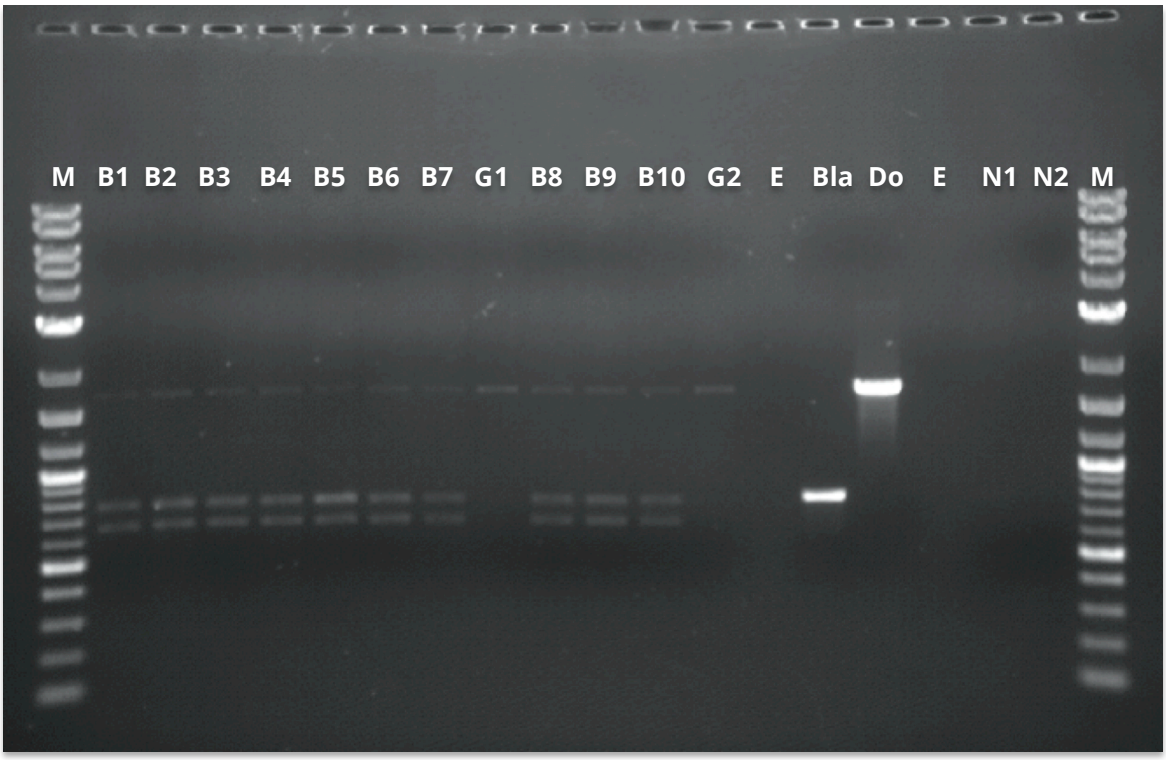


Figure 27 Clones that show β-lactamase activity (blue) have a full-length copy of the gene

In order to check if a given clone harbored an L1-mediated insertion containing the full-length β-lactamase reporter gene, PCR was done using the primer set DES657 x DES658 that spans the entire gene. Presence of full-length, spliced reporter gave a band of ~0.8kb, while the presence of a full-length unspliced reporter was indicated by a band ~ 1.7 kb.

B – clones with blue cells

G – clones with green cells

Bla – positive control for spliced reporter (pUC19 vector)

Do – positive control for unspliced reporter (pDES46)

N1 – HeLa genomic DNA

N2 – water (negative control)

E – empty well

M – marker (2-log ladder)

To assess the transmission of stable or variable reporter expression or silencing through many additional mitotic cell divisions, we conducted a second round of cell subcloning by sequential limiting dilution. As before, we observed mitotically heritable patterns of reporter expression, revealing mostly stable (blue) or variable (mixed) expression of L1 reporter integrants (Figure 28). Again this suggested that while the reporter expression or silencing states were mostly heritable, they also could oscillate. These features are characteristic of reporter variegation/oscillation, indicating epigenetic regulation (Feng et al., 1999).

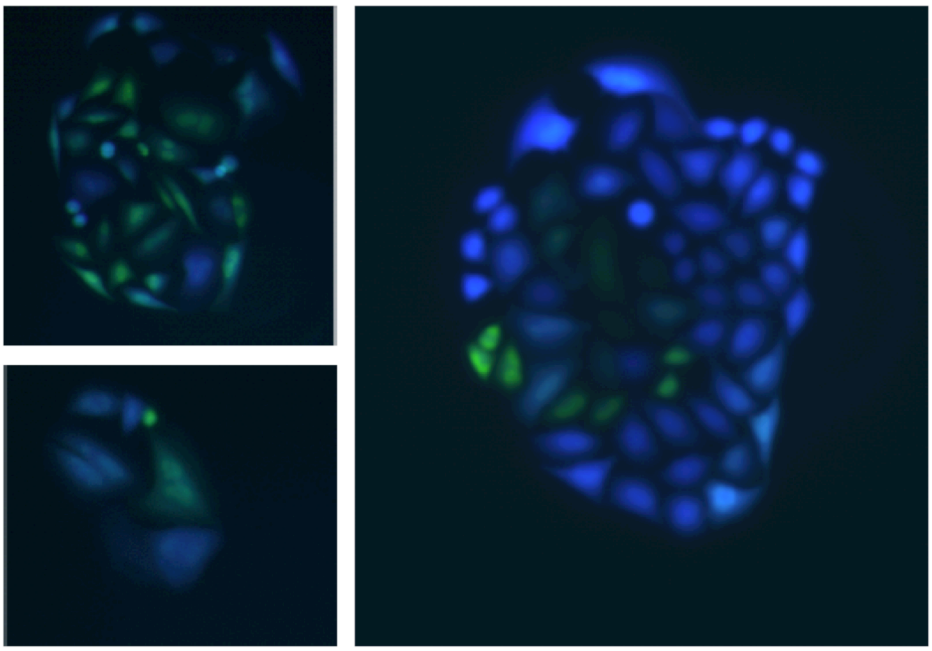


Figure 28 Variegation of L1 reporter expression in subcloned populations of HeLa cells
Individual clones of HeLa cells transfected with pDES46 were further subcloned by limiting dilution. The resultant subclones also showed variegated pattern of expression of the reporter. Each of the three panel shows a colony that presumably arose from a single cell.

To map L1 integrants in various cell clones, including some that expressed high levels of TEM1 reporter (i.e. they were predominantly blue), as well as others that revealed mixed levels of reporter expression, we employed two PCR-based assays. The first assay was inverse PCR using diverging primers located within the 3' UTR of retrotransposons cassette (Figure 29). We chose the farthest possible end of the 3' UTR to maximize the recovery of even small integrants (since, most L1 integrants are 5' truncated). After performing nested PCR, we obtained bands (Figure 29) that were cloned, sequenced and aligned to the human genome sequence (BLAT) in order to determine the chromosomal location of integration events. In

the second assay, genomic DNA was cut using common 4 bp-cutting restriction endonucleases. Appropriate adaptors were ligated onto compatible overhangs, and PCR primers annealing to the L1 and to the adaptor were used (Li et al., 2012; Pornthanakasem and Mutirangura, 2004). PCR products were assessed by Sanger sequencing, resulting in the recovery of 9 independent integrant sites, mapped as insertion-host genome junctions. Although several integrants were not long enough to include any L1 sequence, they nevertheless all exhibited features that characterize bona fide L1 retrotransposition events. These included target site duplications, a short stretch of poly(A) tail, and occasional 5' inversions (Lee et al., 2012; Ostertag and Kazazian, 2001b; Symer et al., 2002). An integrant on chromosome 2 inserted an intact, spliced TEM1 gene. Its spliced structure and the presence of target site duplications confirmed that it had been retrotransposed. Additional independent insertions that were also recovered are summarized in Table 1.

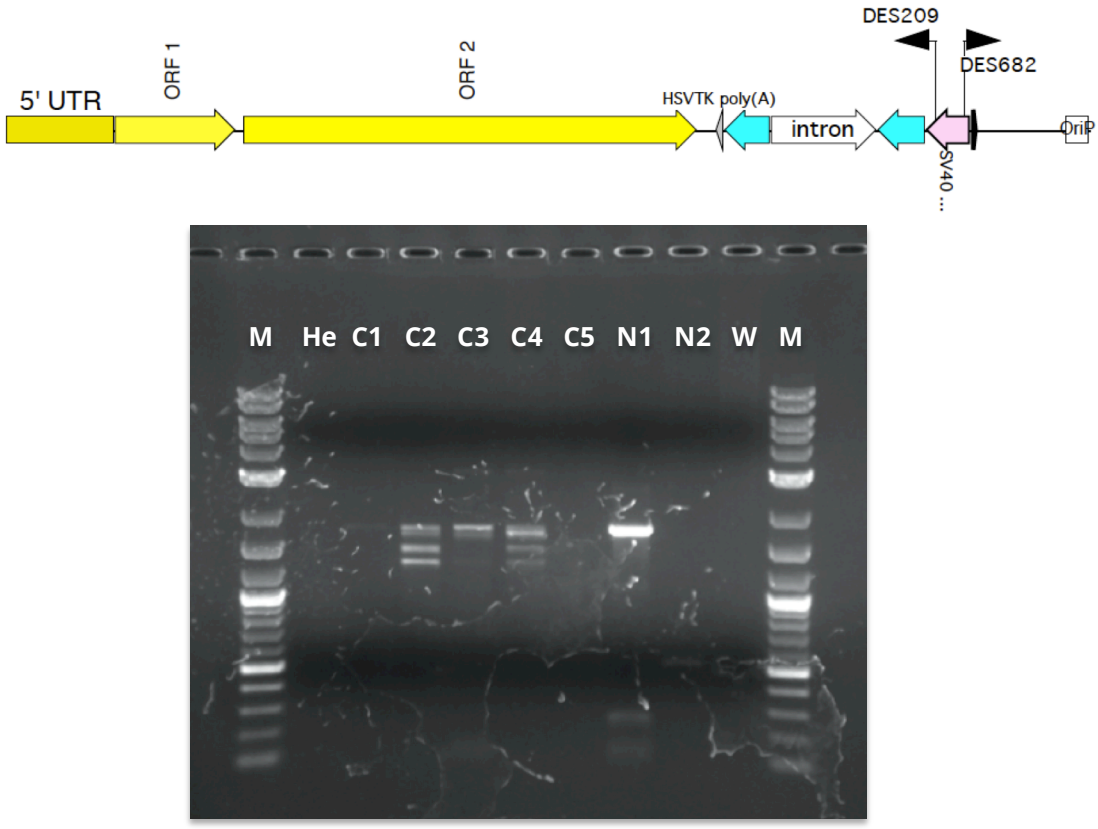


Figure 29 Inverse PCR to recover L1 reporter integrants

Inverse PCR was performed using primers DES682 and DES209 as shown in the top panel, on genomic DNA samples obtained from selected HeLa cell clones C1 through C4 as templates. Negative controls: He, HeLa cell line DNA; N1, Fish sperm DNA (0.1ng/µl concentration); N2, Fish sperm DNA (1ng/µl concentration); W, water control. M, 2-log ladder (MW marker)

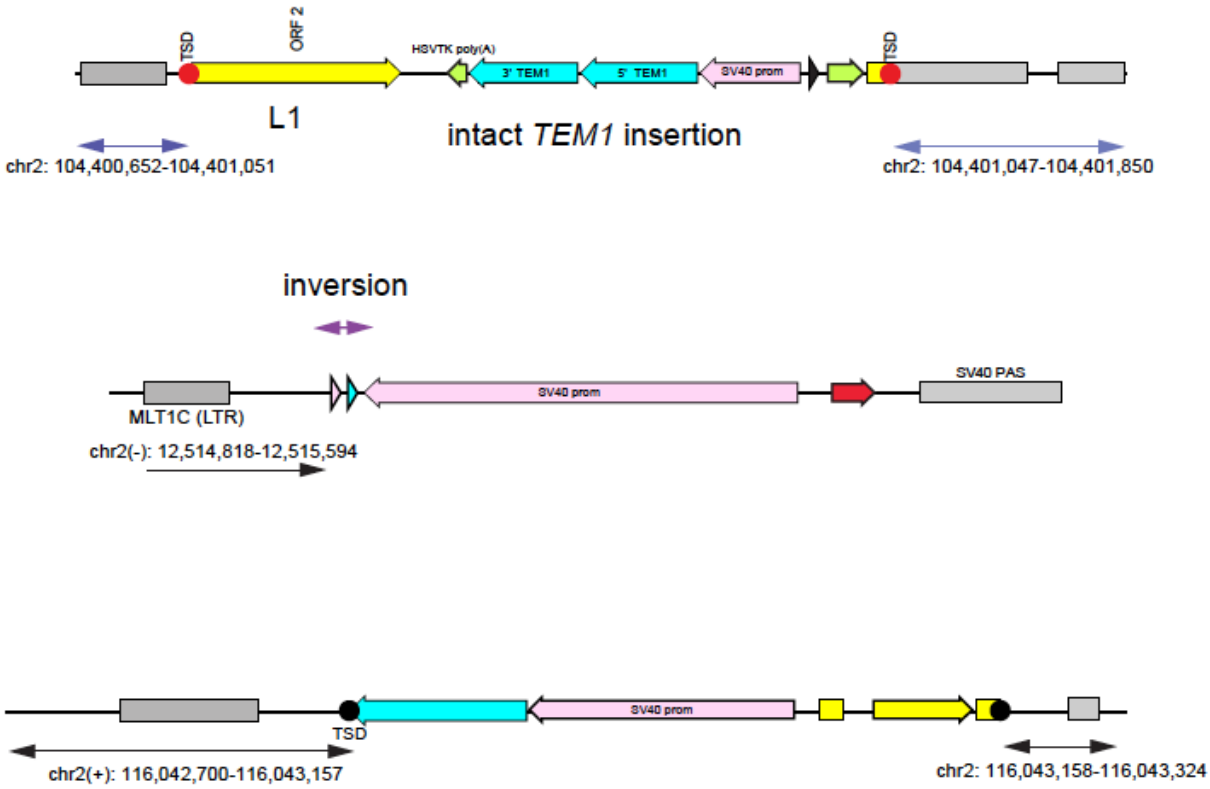


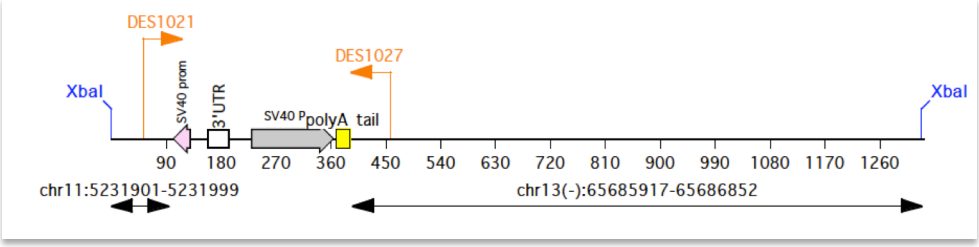
Figure 30 Schematic representations of L1 integrant structures recovered from cultured cells.

All integrants obtained showed characteristics of bona fide L1 retrotranposition integration event – poly A tail at the 3' end, target-site duplications at both the ends, and the rare inversion. One integrant recovered had the entire reporter gene intact (the first panel), while the others had truncations at the 5' end (second and third panels). Chromosome coordinates of the integration sites are also specified alongside.

Table 2 Summary of L1 integrants obtained in HeLa cells

Chromosome locus (strand orientation)	Coordinates	Enzyme used for recovery	Genomic context
2q14.1 (+)	116,043,157	EcoRI	5th intron of DPP10 (dipeptidyl peptidase)
17q24.1 (+)	60,690,241	EcoRI	SINE element MIRb
1q31.2 (-)	189,998,816	XbaI	LINE L1ME2
2p24.3 (-)	12,515,594	XbaI	Between LTR1B and MLT1C
3q13.31 (+)	118,798,663	XbaI	Region 20bp upstream of GA repeat
6q22.31 (+)	121,035,080	XbaI	Region 55bp upstream of AT repeat
11/13 putative translocation 11p15.4 (+) 13q21.32 (-)	11: 5,231,999 13: 65,685,917	XbaI	chr11: intron of gamma globin gene HBG2 chr13: 150bp downstream of L1PB4
10q21.1		Sau3AI	Near L1MCa transposon
2q12.1 (+)	104,401,046	BclI	Between Alu and LTR repeat regions

Based on the inverse PCR sequencing data obtained, there was a plausibility of a translocation between chromosomes 11 and 13. Because L1s are known to cause genomic instability in vivo (Symer et al., 2002), we speculated this could be an L1-mediated translocation. We constructed a schematic map for the plausible translocation event (top panel of Figure 31) and designed PCR primers across the putative chromosome 11–chromosome 13 junction to verify the integrant structure. One set of PCR primers gave positive results to lend some support to our speculation. Hence, we decided to perform FISH to confirm if indeed there had been a translocation that occurred in the HeLa cell clones, mediated by a de novo L1 retrotransposition event. However, the chromosome painting results showed no evidence for a translocation event (Figure 32).



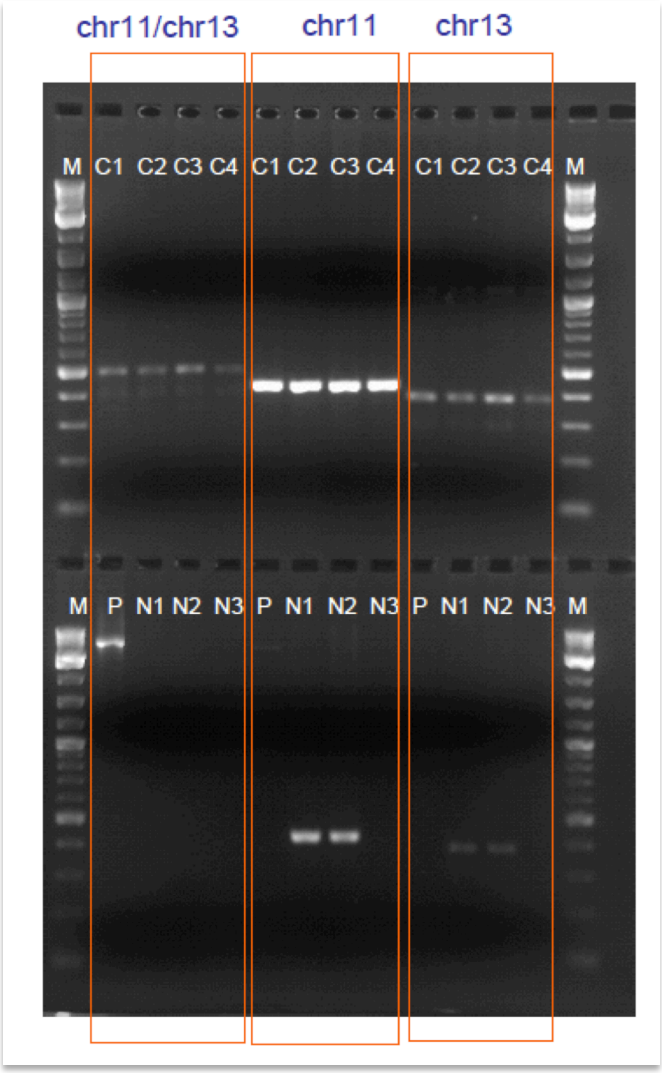
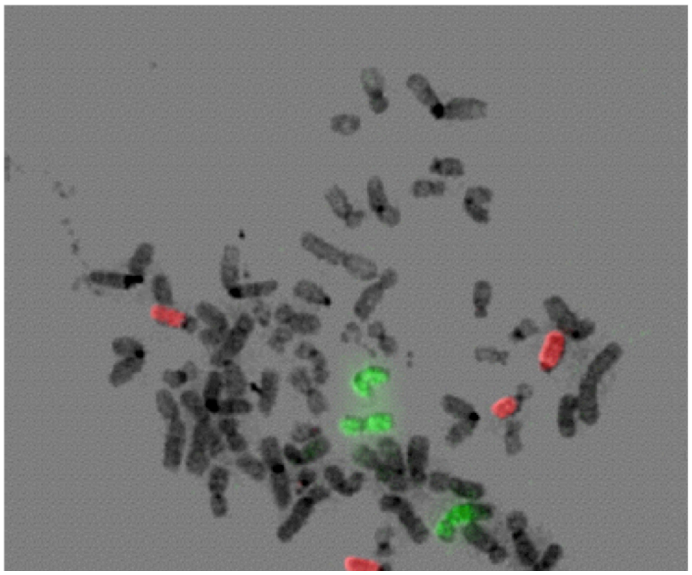


Figure 31 PCR assay for testing the plausible chromosomal translocation between chromosomes 11 and 13 mediated by an L1 transposition event

The top panel shows a schematic of the putative structure of such an integrant, while the bottom panel shows the PCR evidence to support this structure. C1, blue clone; C2-C4, variegating clones; P, positive control (plasmid); N1, negative control (HeLa); N2, negative control (HCT116); N3, negative control (water); M, marker (2-log ladder)

A



B

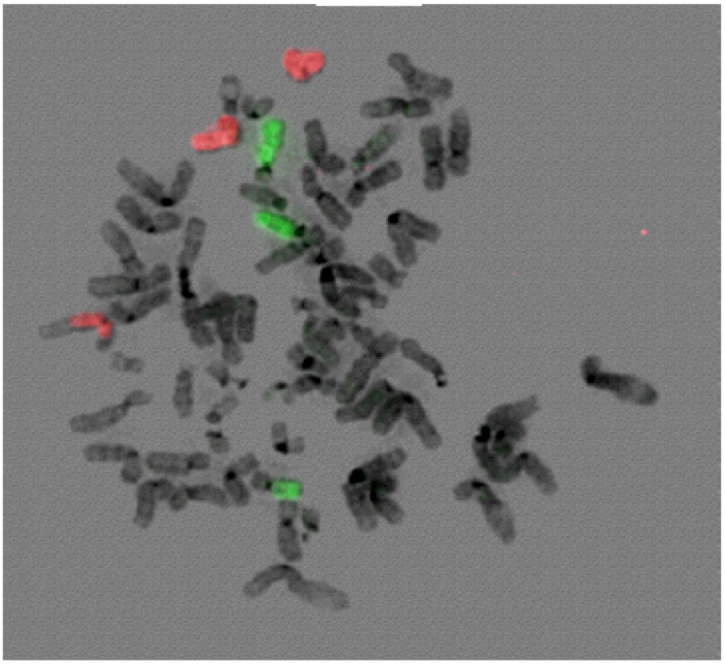


Figure 32 Chromosome painting analysis to verify the plausible chromosomal translocation

Reciprocal staining was done on the HeLa cell line to detect possible translocation between chromosomes 11 and 13. In panel (A), chromosome 11 is stained green and chromosome 13 red; in panel (B), the staining order reversed. There was no indication of a possible translocation between the two chromosomes.

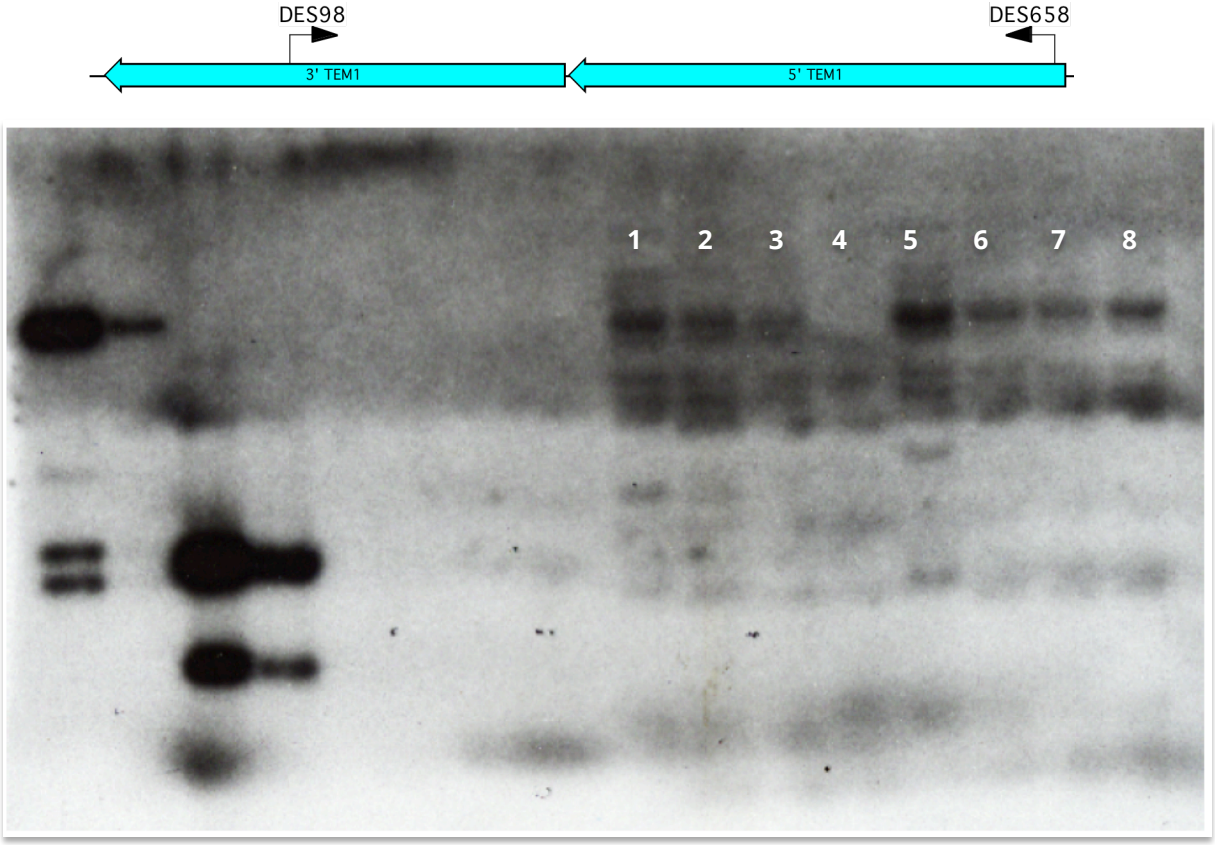


Figure 33 Southern blot reveals clonality of L1 insertions.

Using a probe for TEM-1 β-lactamase gene (using primers as shown in the top panel of the figure, several integrants were detected among various clones of HeLa cells transfected with pDES46. Lanes 1 and 2 correspond to clones that has predominantly blue cells; lanes 3-5 are clones and subclones of variegating phenotype; and lanes 6-8 are clones and subclones of another variegating cell line.

Lane 1 : 1 #5H5

Lane 2 : 1 1c/w B11 (parent of sub-clone in lane 1)

Lane 3 : 6I #9F8

Lane 4 : 6I 8G2

Lane 5 : 6I 0.3c/w C8 (parent of sub-clones corresponding to lanes 3 and 4)

Lane 6 : 5B #4D9

Lane 7 : 5B 1G10

Lane 8 : 5B 0.3c/w C9 (parent of sub-clones corresponding to lanes 6 and 7)

The first six lanes have positive and negative controls.

To assess genetic relationships between L1 integrants in single cell clones, and measure the copy number of new integrants, we conducted Southern blotting on eight representative subclones, each with predominantly blue or variegated phenotypes, which had been derived from the same initial population of transfectants. A radiolabeled probe specific

for the β-lactamase reporter was used to detect spliced integrants that had undergone retrotransposition. Around ~7-10 bands of distinct molecular weights were visible recurrently in each of the clones. However, unique bands also were detected in individual clones (Figure 33). Interestingly, although levels of reporter expression diverged amongst the different cell clones, the pattern of insertion bands was largely similar amongst the cell clones. For instance, the patterns of bands for the variegating cell line named "5B 0.3c/w C9" shown in lane 8 and its subclones shown in lanes 6 and 7 of the blot do not show any noticeable dissimilarity. This further suggests a role for epigenetic variation and not just genetic variation in differential reporter expression.

4.3.2 Lack of de novo cytosine methylation at new L1 integrants in HCT116 cells

In previous studies of human L1 retrotransposition in cultured cancer cells (HCT116 cell line), we applied positive selection on expression of the integrated *Neo*^{R} reporter (Moran et al., 1996; Symer et al., 2002). Such drug selection presumably would impose a requirement for strong expression of the resistance gene, as any cells lacking its robust expression would be killed by the drug. Thus the epigenetic marks observed at newly retrotransposed *Neo*^{R} reporters would be expected to be biased in favor of active, euchromatic marks. In addition to finding many truncated *de novo* L1 insertions, we mapped two full-length L1 insertions (named 2A2 and 7H2) on chromosomes 2 and 14 (Symer et al., 2002). Their identification provided us with a unique opportunity to study *de novo* cytosine methylation established both at the inserted reporter and several kilo bases upstream at the L1 5' untranslated region (5' UTR). We measured DNA methylation using conventional bisulfite treatment followed by PCR amplification and sequencing. We found virtually no DNA methylation at the 5' UTR of both insertions, as only 2.9% and 0.4% of all CpG dinucleotides at those locations were methylated, respectively (Figure 34). In addition, as expected, the spliced reporter integrants at the 3' ends of these full-length L1 integrants also were almost entirely unmethylated; only 0.4% of all their CpG dinucleotides were methylated (Figure 34).

In a control experiment, we verified that the host cells (HCT116) still harbored effective maintenance methyltransferase activity. Thus we measured cytosine methylation at the 5' UTRs of pre-existing genomic L1-Hs elements. We found that the endogenous L1s were heavily methylated (~64% on average), confirming that methylated CpG dinucleotides are maintained in the cultured cancer cells (Figure 35). The bisulfite-sequencing assay of methylation frequency may underestimate actual methylcytosine content at CpG

dinucleotides, since many such nucleotides can undergo spontaneous deamination over time in the cellular genome, resulting in TpG dinucleotides. Such natural mutations are indistinguishable from unmethylated CpG dinucleotides upon treatment with bisulfite. In mutant double knockout (DKO) HCT116 cells, lacking both the maintenance DNA methyltransferase gene *DNMT1* and the *de novo* methyltransferase *DNMT3b* (Rhee et al., 2002), cytosine methylation at L1 insertions was markedly reduced. In these mutant DKO cells, we found that only ~6.5% of all CpG dinucleotides were methylated (Figure 35).

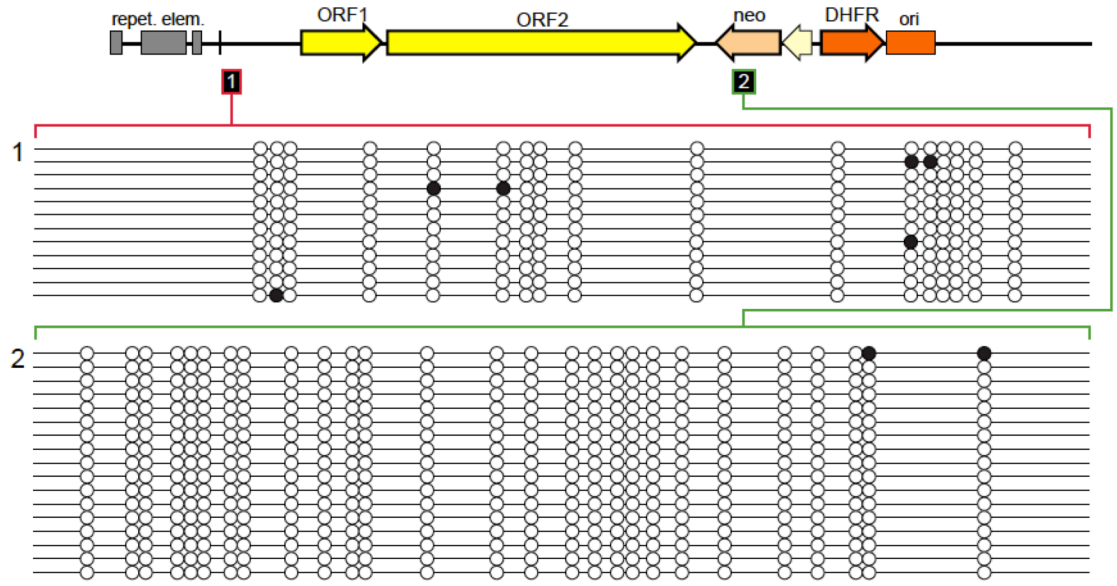


Figure 34 Schematic of a full-length de novo human L1.3 insertion in HCT116 cell line.

This full-length integrant (referred to as 7H2) was recovered in HCT116 cells and reported earlier (Symer et al., 2002).

Minimal de novo cytosine methylation was observed at the (1) L1 5' untranslated region (UTR), and (2) spliced Neo reporter gene of the de novo integrant 7H2.

Black bars: segments analyzed by bisulfite sequencing.

Each circle represents a CpG dinucleotide. Empty – unmethylated; shaded – methylated.

Each row represents a sequencing read (a cloned DNA molecule).

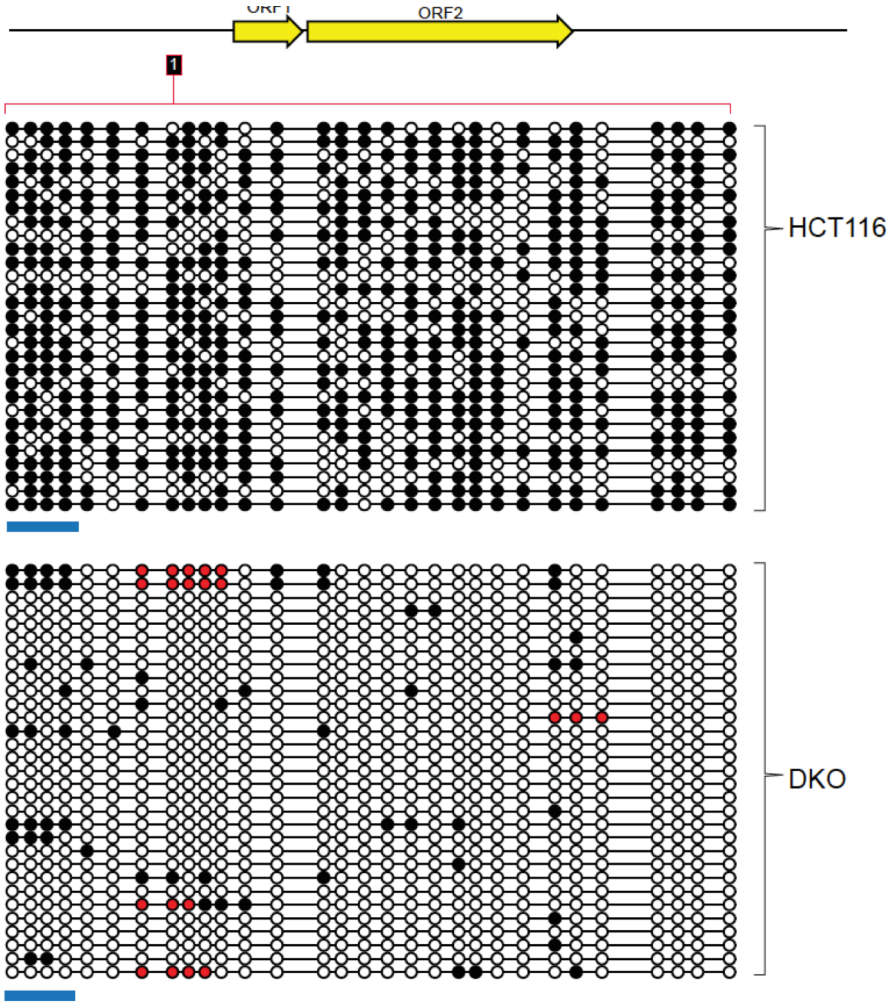


Figure 35 Dense methylation of pre-existing L1 retrotransposons in cultured human cells.

Top: Schematic of pre-existing, full-length L1 integrants genome-wide. Black bar: segment within the 5' L1 untranslated region (UTR), analyzed by bisulfite sequencing.

Relatively dense cytosine methylation was observed in the L1 5' UTR of (A) cultured HCT116 cells but not in (B) cells which lack both DNMT1 and DNMT3b methyltransferases (double knockout, DKO cells).

Bottom bars: position of PCR amplicon to quantify methylation at CpG dinucleotides.

4.3.3 DNA methylation does not silence newly integrated TEM-1 β-lactamase reporters in HeLa cells

We conducted bisulfite sequencing to examine cytosine methylation levels at several independently retrotransposed, spliced lactamase L1 reporter integrants. After their retrotransposition, these L1 reporter sequences were retained in the host cell genomes in the absence of imposed positive selection. We measured the CpG methylation of L1 reporters across two regions (Figure 36): first, in the body of the reporter gene that included 18 CpG dinucleotides, and second, a region encompassing the SV40 promoter and the reporter that included 20 CpG sites. As was the case with L1-NeoR integrants (Figure 34), both these regions of L1-TEM1 integrants were almost completely unmethylated in HeLa clonal cell lines that showed β-lactamase expression. This indicated that methylation played no role in silencing or oscillating expression of the reporter in these cells, and suggested that even without selection, only minimal methylation is established at new integrants. This implies that selective pressure may not be an important factor influencing the epigenetic controls present or absent at the new insertions.

Many of the integrants recovered from HeLa cells had inserted into repetitive elements in the host genome (Table 1). Nevertheless, bisulfite-sequencing analysis of the reporter insertions in bulk showed that most were unmethylated (data not shown). This result suggests that the epigenetic controls established at *de novo* insertions do not necessarily reflect the repressive marks at neighboring, existing copies of repetitive elements.

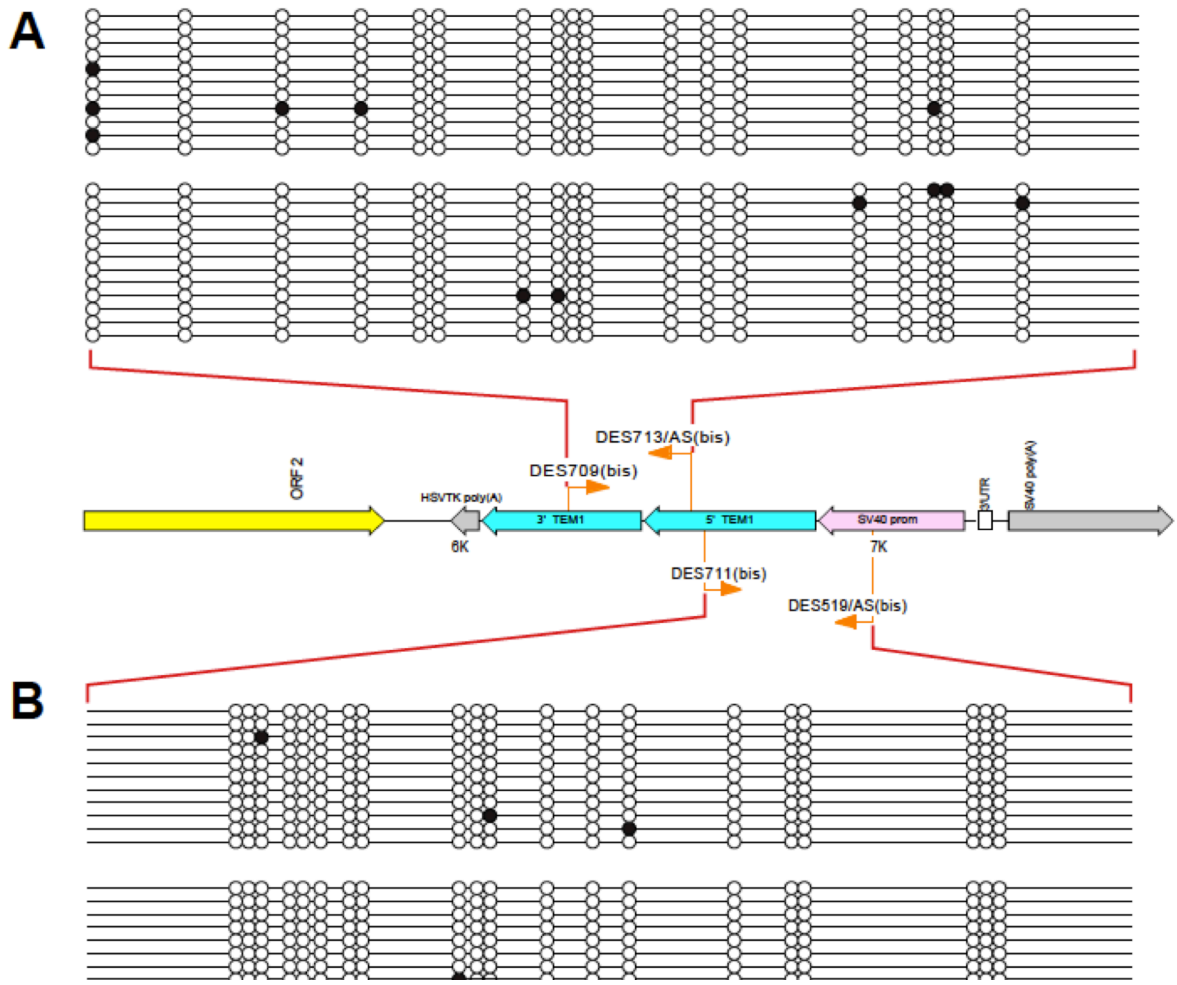


Figure 36 Lack of cytosine methylation at silenced L1 reporter loci in cultured cells.

Cytosine methylation was assessed (using bisulfite sequencing) at new L1 integrants in cloned HeLa cells, both (A) in the body of the TEM-1 beta-lactamase reporter gene, and (B) in the SV40 promoter that drives the reporter. Each circle represents a CpG dinucleotide. A solid circle represents a methylated dinucleotide, while an open circle represents an unmethylated one. Each row represents a clone that was sequenced.

4.3.4 Histone deacetylation is strongly associated with L1 reporter silencing in cultured cancer cells

Both the heritability and the variability in L1 reporter expression in cultured cancer cells suggested that they are epigenetically mediated. However, as described above, we ruled out a role for *de novo* cytosine methylation in regulating new L1 insertions, even though it is maintained at existing elements. To evaluate the possibility that histone tail modifications (i.e. changes in lysine acetylation) could be involved in L1 reporter silencing, we investigated several subcloned cell lines harboring reporter integrants whose expression was variegated or mostly repressed. We treated these cell populations with various histone deacetylase inhibitors (HDACi) including 100 nM trichostatin A (TsA), 10 mM sodium butyrate, 1 μM scriptaid, 1 nM apicin, and 5 mM nicotinamide. Each of these agents was added in standard growth medium to the cultured cells. Upon incubation for 24 h, expression of the reporter gene was reactivated in all cells. Drug-treated cells showed consistently high levels of *TEM1* reporter expression, as demonstrated by uniform blue fluorescence in the CCF2 assay. Thus, a range of HDAC inhibitors, from different mechanistic categories, was effective in de-repressing the silenced reporters.

Upon withdrawal of the HDAC inhibitor TsA from reactivated cells, silencing of the *TEM1* reporter was rapidly re-established (Figure 38). This rapid resetting of TEM1 silencing demonstrated that it can be dynamically established and is reversible. In addition, the state of reporter expression generally appears to be heritable (Figure 28). Thus we conclude that the establishment and maintenance of L1 reporter silencing in cultured human cancer cells is consistent with an epigenetic mechanism involving dynamic changes in histone deacetylation, but not cytosine methylation.

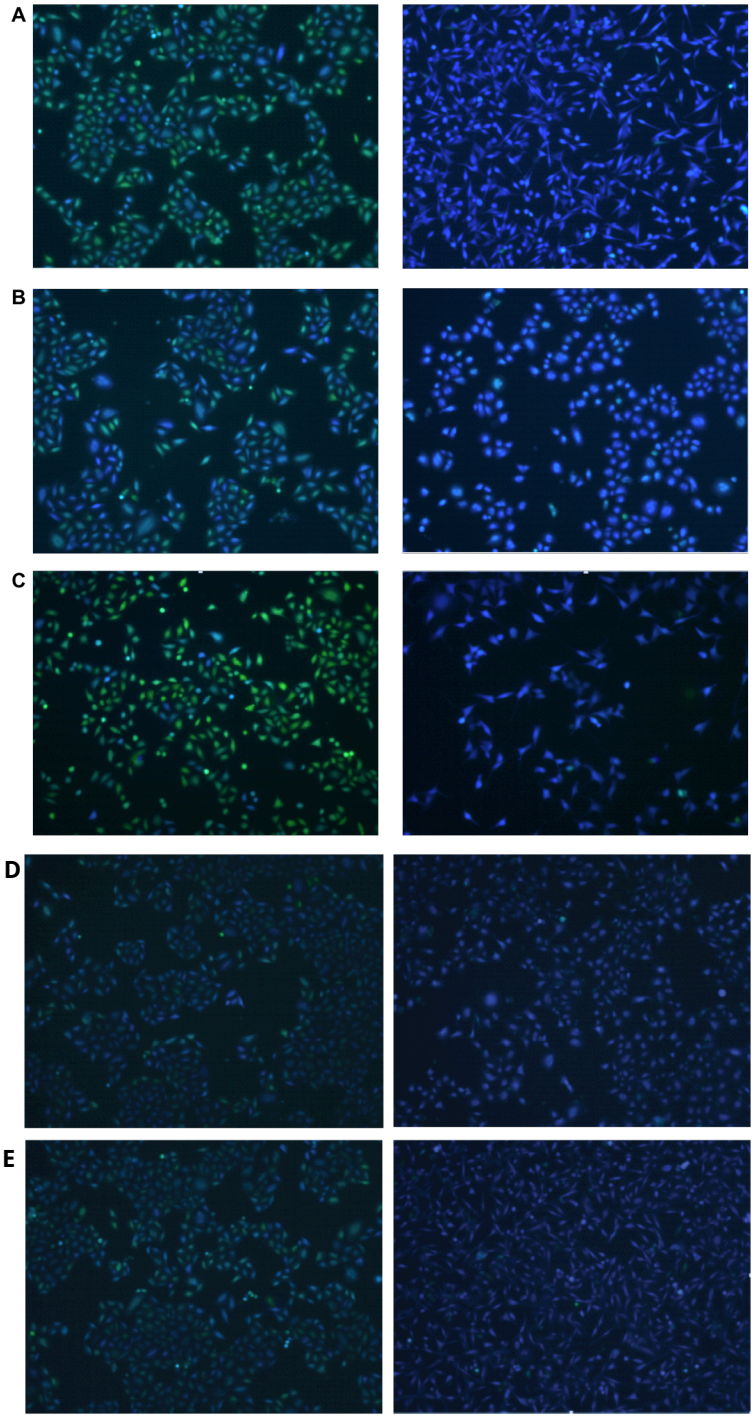


Figure 37 Variable L1 reporter expression in cultured cancer cells is associated with changes in histone acetylation.

Cultured human cancer cells (HeLa cells) harboring de novo L1 reporter integrants were assayed for reporter β-lactamase expression by incubating them with the fluorescent substrate, CCF2. Left: before, and right: after incubation for 24 h with various histone deacetylase inhibitors including: (A) 10 mM butyrate, (B) 1 nM apicidin, (C) 100 nM trichostatin A (i.e. TsA), (D) 1μM scriptaid and (E) 5 mM nicotinamide. In every case, the reporter was de-repressed.

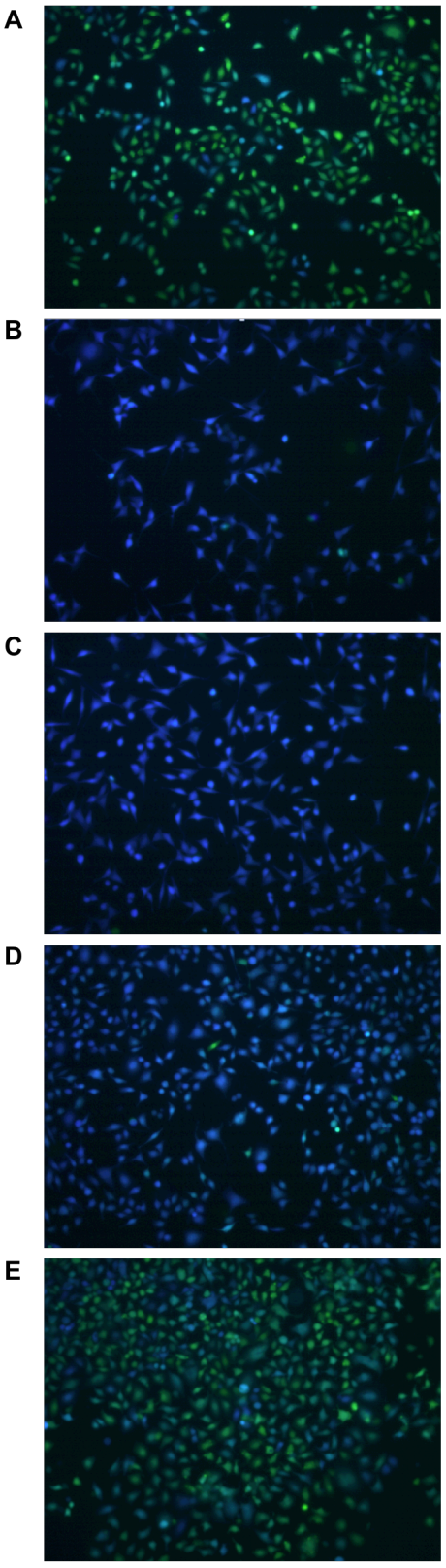


Figure 38 De-repression and re-repression of reporter gene upon treatment and removal of HDAC inhibitor.

Variegating HeLa cells (A) were treated with TSA (100nM concentration), stained and visualized after 24 hrs (B), when the population was entirely blue. Subsequently, TSA was washed and cells visualized by staining after 8 hrs (C), 26 hrs (D) and 56 hrs (E), when variegation was established in the cell population again.

4.3.5 Newly transposed *piggyBac* (*PB*) reporters are robustly expressed in cultured cancer cells

We wanted to know whether the oscillation of reporter expression was due to L1 being used as a mobilizing vehicle, or whether it was independent of the means of mobilization. Hence, as a control, the same reporter genes, viz., green fluorescent protein (GFP) and TEM1 β-lactamase, were mobilized as cargo by *piggyBac* DNA transposons. A large majority of HeLa cells harboring newly transposed integrants displayed stable, robust expression of these reporters (Figure 16). As a negative control, we transfected the reporter without PB transposase. In these transfectants, no integration events occurred, the transient donor plasmid harboring the unintegrated reporter gene was gradually lost, and no reporter gene expression was detected after several days in culture. In cells that had been transfected with *PB* transposase and the reporter donor, and then subcloned by limiting dilution, we occasionally observed a fraction of the cells as expressing stably diminished levels of the reporter protein.

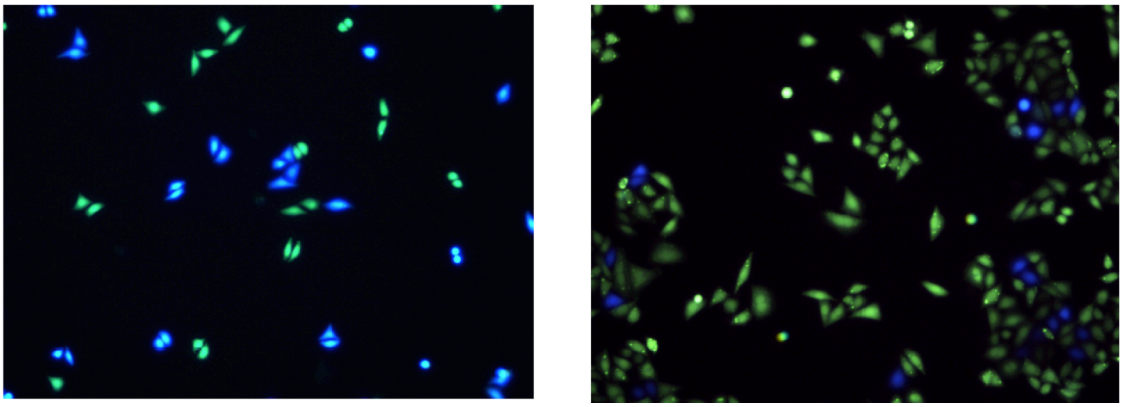


Figure 39 Stable expression of TEM-1 β-lactamase reporter mobilized by piggyBac transposon in HeLa cells

The top two panels show the pattern of expression of the TEM-1 reporter, when it was mobilized using a piggyBac vector into HeLa cells.

4.3.6 New L1 integrants undergo rapid and dense DNA methylation in mES cells

To study epigenetic control of de novo L1 integrants in other cellular and developmental contexts, we induced new mobilization of a highly active mouse L1 retrotransposon ("ORFeus") in mouse ES cells called Truck_305 cells. The parental cell cline for Truck_305, named Bruce4 (Kontgen et al., 1993) had been transfected previously with

linearized pJH435, thereby incorporating a regulatable, codon-optimized mouse L1 (ORFeus) donor in its genome (An et al., 2008). In Truck_305 cells, the conditionally activated ORFeus transgene cassette consists of these elements: (i) a composite cytomegalovirus (CMV) IE enhancer/modified chicken β-actin promoter, (ii) a floxed β-geo/stop cassette comprising a β-galactosidase/neomycin phosphotransferase fusion gene (Friedrich and Soriano, 1991) and triple tandem copies of SV40 late polyadenylation signal (Lobe et al., 1999), (iii) ORFeus ORF1 and ORF2 (Han and Boeke, 2004), (iv) a GFP-based retrotransposition indicator cassette with its own promoter (inverted LTR) and polyadenylation signal, and (v) β-globin polyadenylation signal (An et al., 2008). In this setup, the L1 retrotransposition cassette is activated in cells only when the floxed β-geo gene is removed from the donor construct, which results in the juxtaposition of OFReus ORF1 and ORF2 directly downstream of the CAG promoter, thereby activating L1 expression. This was done by exposing the Truck_305 mES cells to Cre recombinase (transfected into the cells using an adenovirus), thereby inducing L1 expression and retrotransposition.

We picked individual cells to derive subclonal populations. Genomic DNA was isolated from several ES cell clones, and linear amplification mediated-PCR (LAM-PCR) was performed to recover new L1 ORFeus integrants. They were sequenced and mapped, and custom bisulfite sequencing primers were designed. The genomic DNA was modified with sodium bisulfite, and then PCR amplification was performed using primers internal to the reporter (Figure 40A,B), or to specific integrants (Figure 40B). Amplicons were cloned and sequenced. The results showed heavy methylation of the retrotransposed reporter gene both in bulk and when individual integrants were investigated.

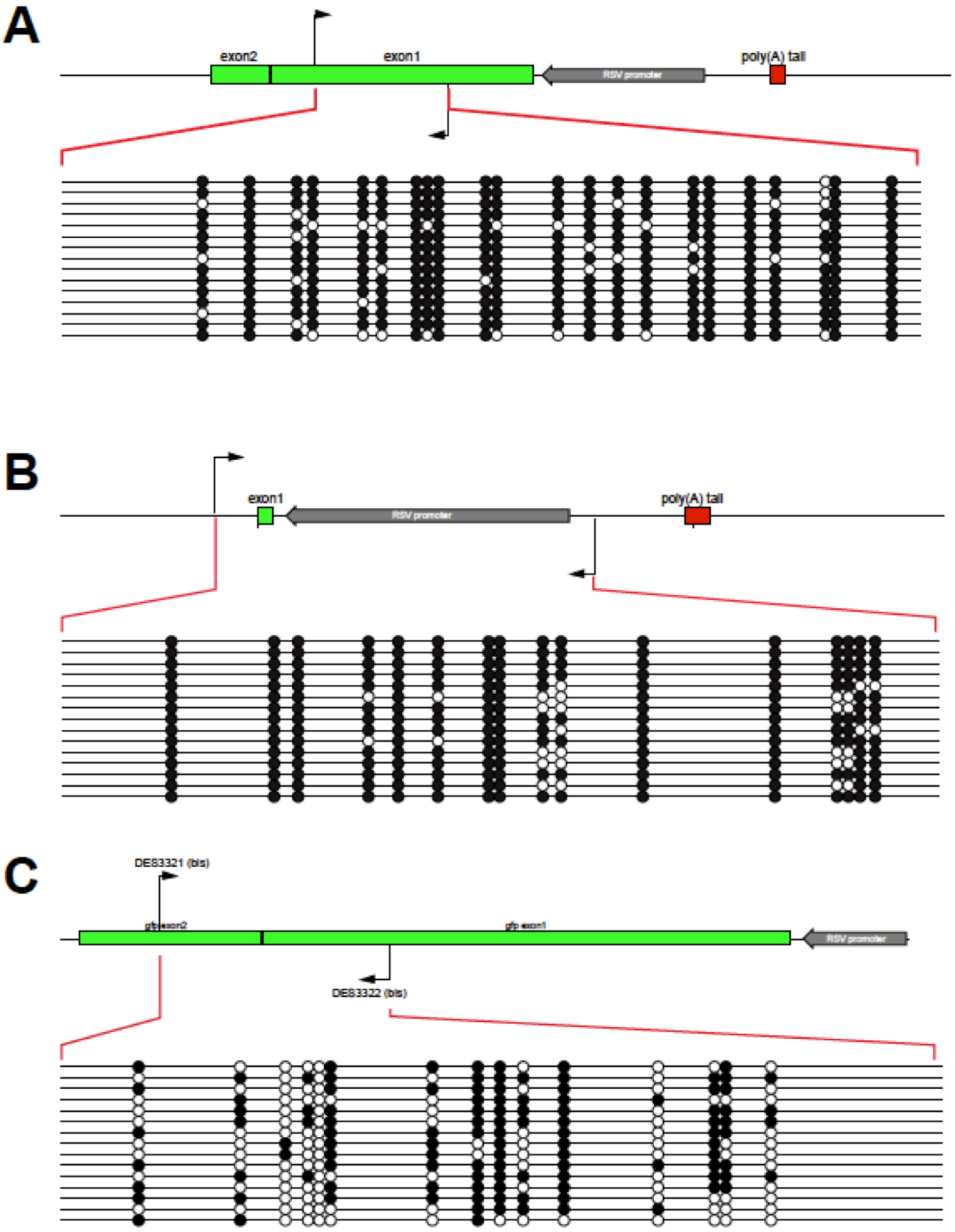


Figure 40 New L1 integrants in mouse embryonic stem (mES) cells undergo dense cytosine methylation.

Dense cytosine methylation at new L1 integrants in mouse ES cells was revealed by bisulfite sequencing. Initially, Bruce4 ES cells were transfected with pJH435, encoding an inactivated L1

ORFeus donor element marked with GFPuv-AI reporter (An et al., 2008). Upon activation of L1 donor expression by transient infection of the culture using adenoviral Cre recombinase, individual colonies were picked and cultured on feeder cells for > 2 months. The cytosine methylation status of these ES clones that harbored new integrants was assessed either in bulk or at individual loci using bisulfite sequencing.

(A) For mES subclone 1B6-A07, we used primers DES3301 x DES3314, which anneal within GFPuv. This PCR amplicon does not cross the AI splice site, so unspliced donor template DNA also can be amplified.

(B) For mES subclone 1B6-A08, primers DES3298 x DES3299 were used.

(C) For mES clones 1B06/B02, 1C6 and 2D4, primers DES3321 x DES3322 were used to assay 15 CpG dinucleotides in a 234 nt amplicon across the GFPuv-AI splice junction.

As a control, we transfected into mES cells, a donor plasmid encoding L1 ORFeus marked by TEM1 β-lactamase in its 3' UTR (pJL5). We observed very little expression of β-lactamase reporter integrants, consistent with the idea that dense cytosine methylation led to their strong silencing. Interestingly, upon limiting dilution of the transfectants, rare cellular subclones were observed that harbored derepressed L1 integrants exhibiting stable, robust reporter expression. However, the DNA methylation status of the newly integrated L1s in these subclones was not investigated further.

4.3.7 Newly transposed PB reporter integrants are not silenced in mouse ES cells

To compare reporter integrants mobilized by different mechanisms into distinct genomic targets, we also launched *PB* transposons carrying comparable reporter genes in mES cells (E14Tg2a.4 cells, which were Bruce4 cells). Just as observed in HeLa cell line, the expression levels of integrated *PB* reporters in mES cells remained robust even after many days of culture (Figure 41). No oscillation or decreases in reporter expression were detected by visual inspection. Therefore, we conclude they underwent none or only minimal epigenetic silencing.

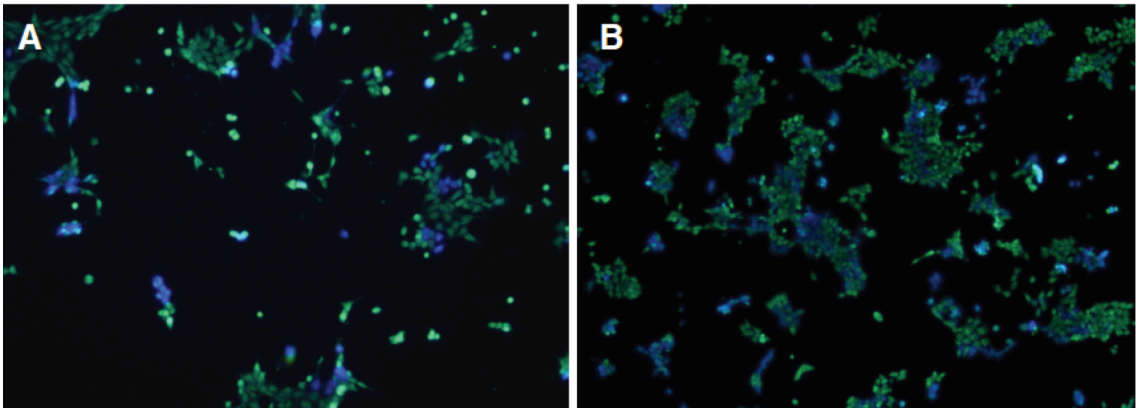


Figure 41 TEM-1 β-lactamase reporter mobilized by piggyBac vector into mES is expressed persistently, without being silenced

E14Tg2a.4 cells were transfected with piggyBac vectors carrying the TEM-1 reporter as cargo. Resultant transfectant cells showed either blue (stable expression of reporter) or green (no reporter present), but no intermediate expression.

4.3.8 New L1 integrants undergo rapid, dense DNA methylation in various tissues in vivo

To study the epigenetic modifications established at new L1 integrants *in vivo*, we obtained several tissues from a transgenic mouse model, in which L1 ORFeus mobilized initially in a "pseudofounder" animal (An et al., 2006). While that individual mouse lacked the donor element in its differentiated tissues, it nevertheless had transmissible new insertions. Its progeny also harbored some of the same new L1 ORFeus insertions as were present in the pseudofounder itself, revealing the transmission of these genomic L1 insertions through the germ line to its offspring. The authors had concluded that new L1 insertions had retrotransposed from the donor episome, immediately after its injection as a transgene and before loss of the episome due to cell division during early embryogenesis.

We measured de novo cytosine methylation established at newly mobilized L1 ORFeus integrants in the pseudofounder, as well as at some of the same integrants transmitted to its offspring. Genomic DNAs isolated from various tissues and members of the pedigree were treated with sodium bisulfite and sequenced. The results showed that almost all of the CpGs in de novo L1 integrants were methylated in mouse B386 (Figure 42). This showed that a variety of somatic tissues including tail and other organs (data not shown) were densely methylated. In another pedigree, integrated L1 ORF2 sequences in a second, independent pseudofounder mouse F234 again were found to be fully methylated (Figures 43 and 44).

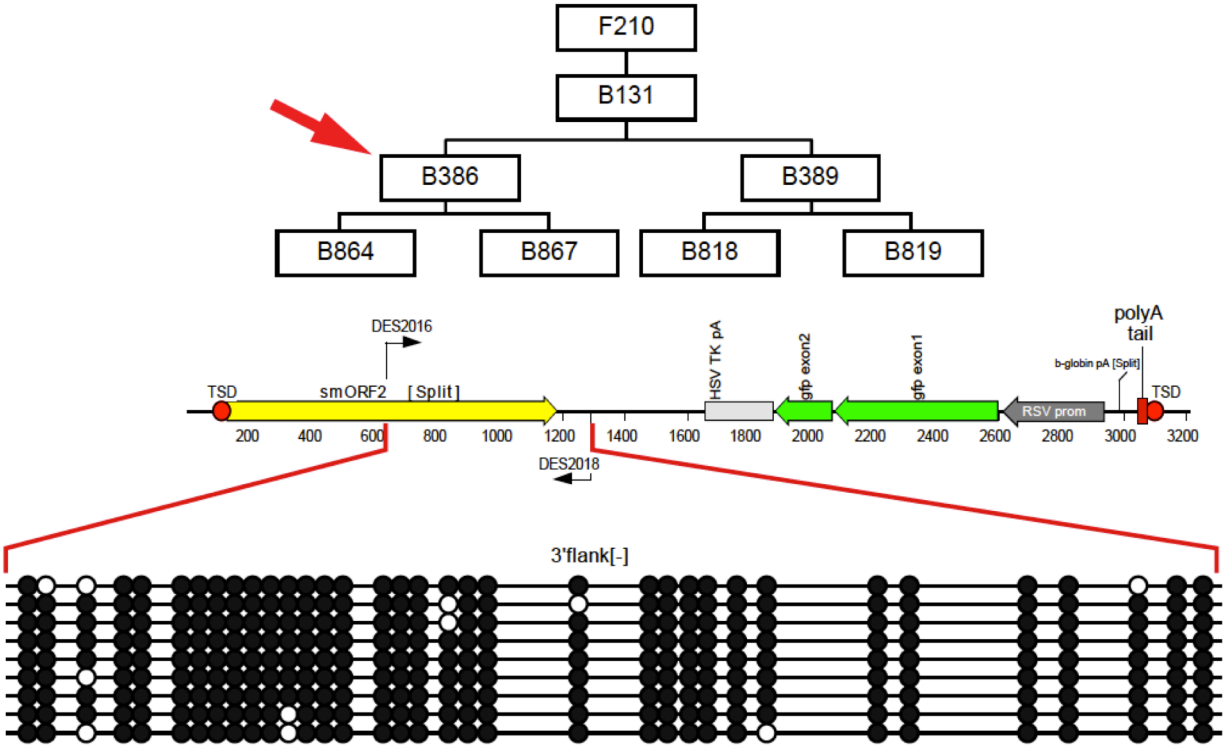


Figure 42 Silencing of somatic de novo L1 insertions by dense cytosine methylation in mouse B386.

Thirty five CpG dinucleotides were studied in the region amplified using PCR primers DES2016 and DES2018. Each of the nine rows represents a separate read (clone). The pedigree of B386 mouse is also shown.

Since this analysis was performed using primers designed to amplify any reporter integrants without specificity for a particular genomic target, we also used locus-specific primers to conduct bisulfite-sequencing PCR. This was done so that more specific detail about individual L1 integrants in a variety of genomic loci could be obtained. Tail DNA from N2 generation mouse B389 and its progeny B818 and B819 (N3 generation) were used for this analysis. Almost all of the 51 CpG dinucleotides in the region studied were methylated (data not shown).

Taken together, these results demonstrated several points: a) that new insertions occurring early in embryogenesis (i.e. somatic insertions) underwent dense, de novo methylation through the course of development; b) that methylation was maintained through differentiation into a range of tissues in the developing organism; and c) that methylation at such new insertions was maintained and/or re-established upon transmission through the germline.

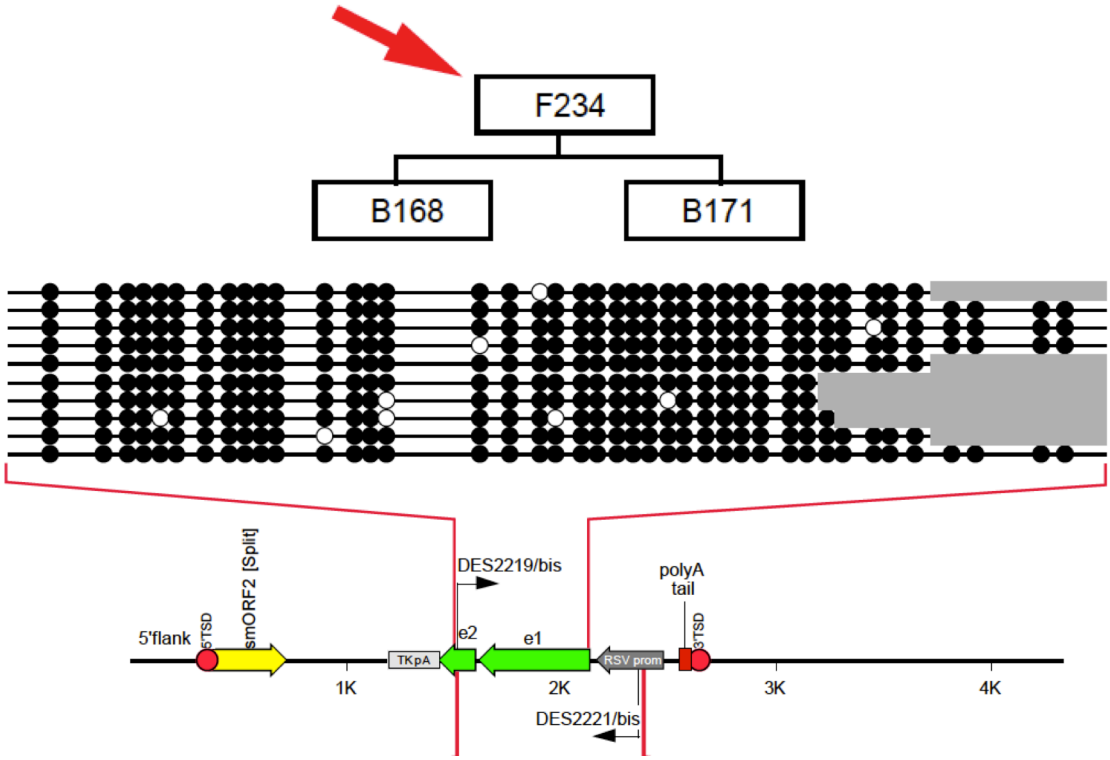


Figure 43 Silencing of somatic de novo L1 insertions by dense cytosine methylation in mice. (tail tissue).

We evaluated cytosine methylation at new L1 insertions genome-wide by performing bisulfite sequencing of amplicons within integrated sequences. Pedigree of pseudofounder mouse F234, demonstrating that at least some of its de novo L1 integrants were transmitted to progeny. The schematic of de novo L1 integrant structure after retrotransposition in vivo is also shown.

Dense methylation at 41 CpG dinucleotides in de novo L1 integrants was revealed by bisulfite sequencing of mouse tail DNA, using the amplicon DES2219 X DES2221.Each row represents an individual sequence read.

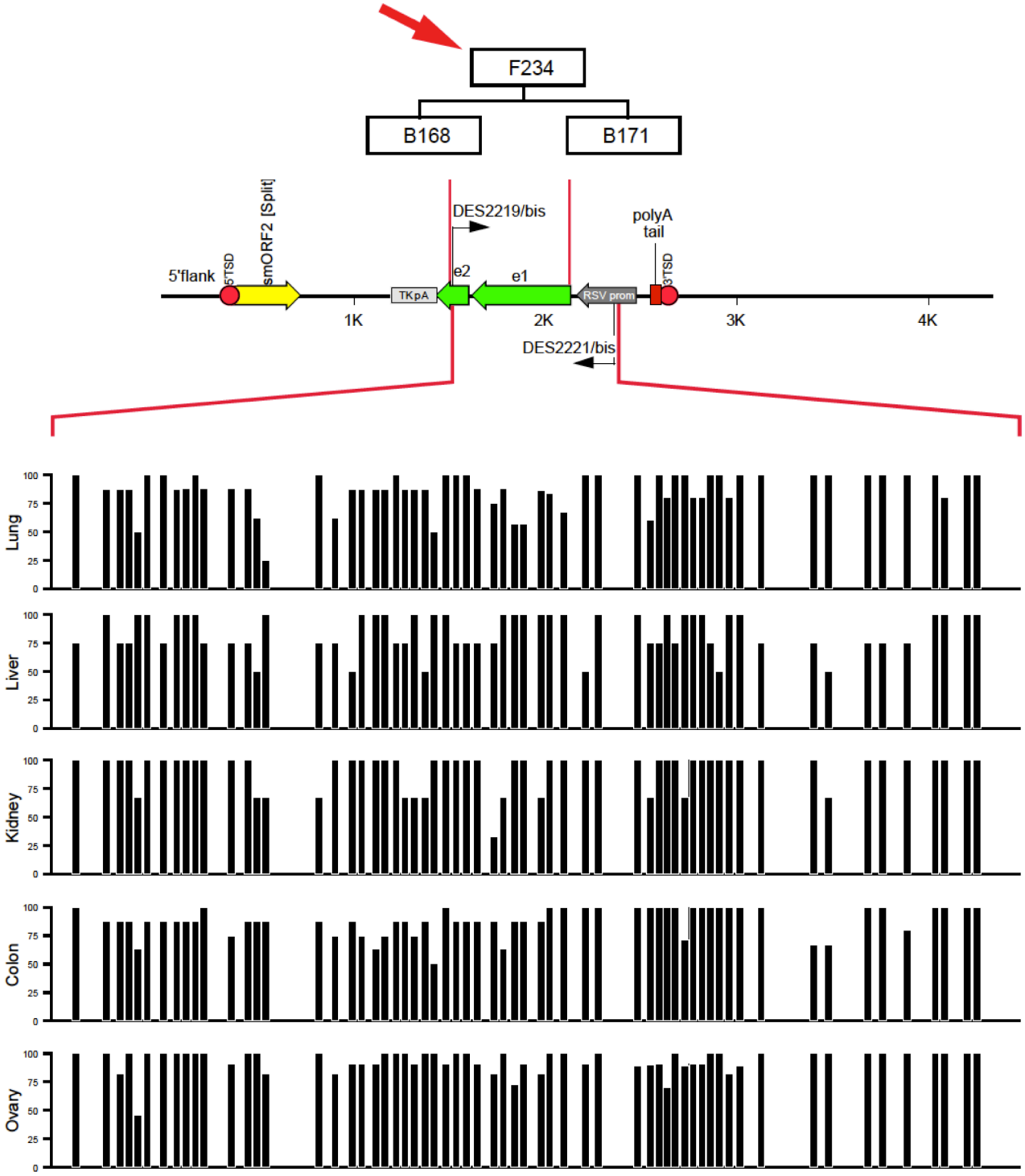


Figure 44 De novo silencing of somatic L1 insertions by dense cytosine methylation in mice (other tissues).

De novo L1 integrants occurring in various tissues of the pseudofounder mouse F234 (shown in the pedigree of the earlier figure) become densely methylated during development.

4.4 DISCUSSION

Endogenous retrotransposons comprise a substantial portion of the mouse and human genomes. Several distinct TE families have accumulated in and modified the mammalian genome profoundly over evolutionary time (Ostertag and Kazazian, 2001a; Symer and Boeke, 2010). The genetic and genomic changes caused by endogenous mobilization of human or mouse L1 retrotransposons have been well studied. By contrast, the regulation of *de novo* L1 integrants has not been evaluated fully in the wide-ranging biological contexts in which retrotransposition can occur, but it is believed that typically epigenetic control underlies such regulation (Whitelaw and Martin, 2001). A recent study of endogenous L1 expression in human ES cells revealed that predominantly those elements localized in expressed genes were expressed, while others located outside of such genes were not, thereby supporting this view (Macia et al., 2011). In addition, the activation of endogenous L1 expression, upon reprogramming somatic cells into induced pluripotent stem (iPS) cells, also implied that epigenetic derepression of silenced elements can occur (Wissing et al., 2012). By contrast, another group reported that epigenetic silencing of TEs is stable through reprogramming of somatic cells to iPS cells (Quinlan et al., 2011).

Here we investigated the epigenetic silencing of newly integrated L1 reporters in cultured human cancer cells, mouse ES cells, and in several tissues of pseudofounder transgenic mice and their progeny. 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 (Figure 28). Regardless of their expression levels, the genomic L1 reporter integrants remained almost completely unmethylated, even after many cell divisions (Figure 36). L1 reporters were silenced rapidly by histone tail deacetylation, as demonstrated by strong, uniform reactivation of reporter expression upon treatment with diverse HDAC inhibitors (Figure 37). This histone deacetylation-mediated silencing was re-established rapidly in most cells upon removal of those inhibitors (Figure 38). A recent analysis of mouse ES cells lacking HDAC1 revealed that transcription of RTLR45 elements, a particular

subfamily of mouse LTR retrotransposons, was upregulated (Reichmann et al., 2012). This variable epigenetic silencing in cultured cancer cells may reflect the timing of L1 integration, i.e. later stages of somatic development, when cellular de novo methyltransferases are expressed at very low levels. It has been proposed that retrotransposon silencing is both incomplete and stochastic (Whitelaw and Martin, 2001). 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. Variegation is essentially clonal, i.e., mitotically stable. However, a recent study raises the interesting possibility that the inheritance of heterochromatin state may be subjected to very quick changes – from a repressed chromatin state to an open state (Janicki et al., 2004). 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 (Figure 40). 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 mice, also were stably silenced by dense cytosine methylation (Figure 44). We surmise that L1 reporter silencing (via dense cytosine methylation) that was observed in somatic tissues of pseudofounder mice 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. A plausible explanation is that de novo DNA methyltransferases, which are highly expressed in early embryogenesis and in ES cells, could target the newly inserted L1 cDNA sequences, which initially were unmethylated. These enzymes normally re-establish DNA methyation after a wave of hypomethylation erases most of the methylcytosine marks in developing embryos (Bodak et al., 2014; Jachowicz and Torres-Padilla, 2015).

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 (Figure 45).

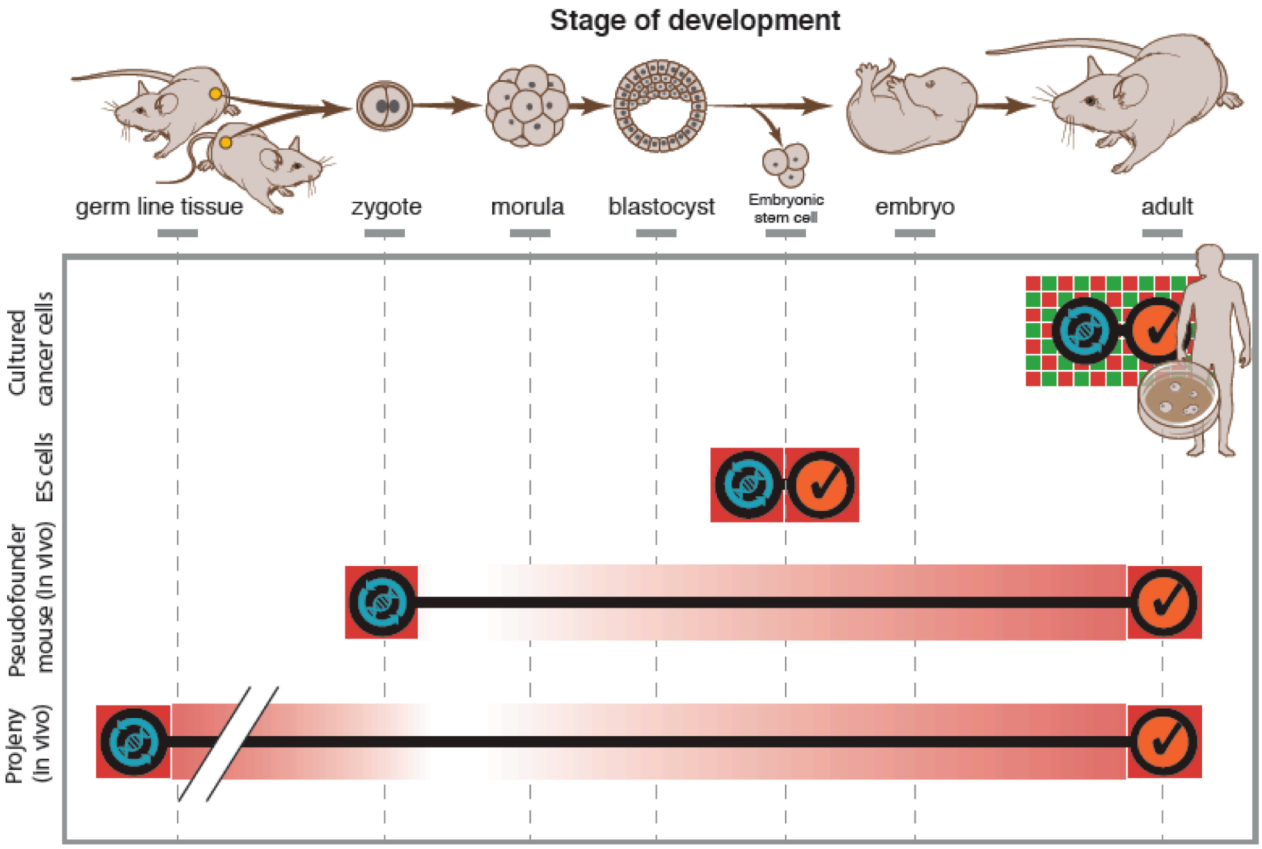


Figure 45 Model for L1 integrant silencing in various cellular and developmental contexts

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 infer that the observed epigenetic marks at de novo L1 integrants could reflect the timing during organismal development at which they integrated (Figure 45). We acknowledge that our measurements of epigenetic silencing were performed much later after the time of retrotransposition per se, i.e. after many cell divisions. This was done for practical reasons, since initially the L1 donor elements had retrotransposed in single cells. Those new L1 integrants initially would be extremely difficult to identify and study, until they could be enriched by subcloning or upon embryonic development and tissue differentiation. This has been an unsolved problem in studying new hops of L1 in somatic tissues such as the brain as

well. We surmise that our results reflect the developmental stage at which retrotransposition occurs, by defining the epigenetic marks established and then maintained at new integrants.

We also observed dense cytosine methylation and silencing of new L1 insertions that had been transmitted in heterozygosity from one generation to the next, i.e. from the pseudofounder mice to some of their progeny. This occurred despite the genome-wide demethylation and remethylation that occur widely in developing germline tissues, as well as during early embryogenesis. We did not evaluate whether the methylation marks that were established in the first generation in which L1 retrotransposition occurred were erased and then re-established in offspring, or alternatively if they were stably maintained in germline tissues and then transmitted without modification through embryonic development in the subsequent generations (Figure 45).

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. This is evident considering the fact that while multiple copies of a transgene frequently can be integrated into a single genomic locus (Garrick et al., 1998; Janicki et al., 2004), individual copies of L1-reporter integrated at multiple sites in the genome. 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 (Figure 39 and Figure 41). In the unusual cases where *PB* reporters were silenced, we found that they were de-repressed in response to HDAC inhibition, as we found for variegated or silenced L1 integrants (data not shown). Second, the percentage of cells in which *PB* reporter integrants were silenced was much lower than that with L1 integrants. These results also are consistent with and extend previous reports documenting a lack of *PB* integrant silencing in vivo (Ding et al., 2005; Nakanishi et al., 2010). 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. Thus, in comparison with the target sites of new L1 integrants, which are enriched slightly in intergenic genomic regions (Babushok et al., 2006) or are distributed randomly (Gasior et al., 2007), more than half of *PB* integrants are enriched inside expressed genes (Meir et al., 2011). Similar, stable expression from *Sleeping Beauty* (*SB*) DNA transposon reporter integrants has been described (Yant et al., 2000). However, more recently, substantial post-integrative silencing of SB insertions has been found, involving DNA methylation and histone deacetylation.

In a recent study of epigenetic silencing of new L1 insertions in human embryonic carcinoma (hEC) cells, histone deacetylation was identified as the silencing mechanism (Garcia-Perez et al., 2010). We note important similarities and differences between that study's results and our data presented here. First, we confirmed that histone deacetylation occurs at new L1 insertions, in cultured cancer cells. Second, we found that new L1 insertions were densely methylated in mES cells. By contrast, new insertions in hEC cells were not silenced by cytosine methylation. This discrepancy may be due to differences in the host cells' species of origin, differences in epigenetic mechanisms operative in hEC vs. mES cells, differences between the mobilized elements' structures or sequences themselves, or potentially differences in the extent to which the cultured cells had differentiated in vitro. Interestingly, reprogramming of somatic cells to form induced pluripotent stem (iPS) cells led to L1 reactivation and mobilization in human but not in mouse (Quinlan et al., 2011; Wissing et al., 2012). Mouse EC cells have been shown to express L1 full length RNA and ORF1p (Martin and Branciforte, 1993), but how the de-repressed state of chromatin in those cells may relate to epigenetic controls in mES cells or hEC cells has remained unclear. Third, here we studied silencing of newly mobilized L1 insertions in vivo, both in differentiated tissues of pseudofounder mice and in their offspring. However, the prior study did not include an analysis of silencing of new integrants in vivo. This was due to their focus on mobilization in human cells, precluding experiments in vivo. By contrast, our inclusion of mouse models facilitated such analysis. Fourth, the controls that were compared with L1 mobilization in the two studies were very different. In the prior study, HIV-like retroviruses mobilized the transgenic reporter genes, By contrast, here we used *PB*, a DNA transposon, as a basis for comparison with L1 retrotransposition. These control vectors differ in their mechanisms of integration, in their genomic target sites, and in the frequency of insertions generated per

host genome (Chen et al., 2015). Each of these factors could play significant roles in determining the downstream epigenetic silencing marks established at the new insertions.

Despite their similarities in terms of expression of certain markers like Oct4, and their state of being undifferentiated, EC cells differ from ES cells considerably, and hence are thought to "present only a caricature of ES cells" (Andrews et al., 2005). Not only do they differ in their origin (ESC from the ICM, while EC from germ cell tumors), but also in the states they represent and their differentiation dynamics. In particular, PA-1 cell line corresponds to a later stage of embryogenesis and differs morphologically from the earlier stage equivalents of EC cells (Andrews, 2013). EC cells and ES cells, therefore, can be thought of as representing very different points in the spectrum of differentiation. Hence, our study that was done in mouse ES cells very well complements the study that was done using the human EC line (Garcia-Perez et al., 2010).

While DNA methylation was found to be not involved in reporter silencing in EC cells, a previous study reported methylation of L1 reporters in mouse, through repeat-induced gene silencing (Rosser and An, 2010). The same group had studied and recently reported methylation and silencing of germline insertions in a transgenic mouse model (Grandi et al., 2015). They did not focus on somatic insertions, however. While our work corroborates theirs findings so far as methylation of germline insertions are concerned, we have also studied the silencing of L1 reporters that have not gone through the germline, i.e. somatic insertions, thereby providing a more comprehensive view of L1 reporter silencing in mouse.

We acknowledge the likely limitations in this study. First, to facilitate their recovery, we artificially marked the L1 and *piggyBac* donor elements using engineered, heterologous reporters including a strong promoter and terminator. In comparison with native, unmarked elements, the reporter genes incorporated into donor TEs could potentially interfere with the transposon mobilization. Moreover, upon integration they could trigger antisense transcripts or otherwise artificially trigger or disrupt silencing by mimicking actively transcribed, protein-coding genes. However, we observed comparable results in evaluating epigenetic control of a range of newly mobilized, diverse reporter cassettes.

Second, we did not investigate L1 insertions that were newly mobilized in germ line tissues. Extensive research has been conducted recently on control of extant TEs in germ tissues during embryonic development. They appear to undergo a wave of demethylation followed by two distinct waves of de novo methylation (Molaro et al., 2014). PIWI-interacting small RNAs (piRNAs), whose transcription is frequently initiated from TEs in germ tissues,

mediate their regulation and silencing. The epigenetic control of full-length L1s is associated with establishment of trimethylation at H3K9 residues, resulting in repressive heterochromatin (Zamudio and Bourc'his, 2010). Recently, it has been shown that KAP1 protein in enriched in full-length L1 elements (in their 5' regions) in human ES cells (Castro-Diaz et al., 2014), which in turn induces heterochromatinization (H3K9me3) induced by KAP1-mediated events, in line with the observation that DNA methylation in hES cells is induced by PIWI/ piRNA-mediated silencing (Marchetto et al., 2013).

Third, we did not recover new insertions immediately after their integration in single cells. Until very recently, this approach has been virtually impossible, as we would have to identify and characterize new insertions in individual cells or very small subclonal populations within a few cell divisions of integration, without a method to enrich for them. The new insertions' minimal allelic fractions would require use of ultra-deep sequencing, resulting in further analysis extremely difficult.

Fourth, we did not conduct chromatin immunoprecipitation experiments to assess enrichment of particular epigenetic marks that are associated with various forms of epigenetic silencing at or near the new L1 integrants. In their previous paper describing L1 reporter silencing in hEC cells (Garcia-Perez et al., 2010), the authors published ChIP data confirming that histone deacetylation indeed accompanied L1 silencing.

And finally, although *in vivo* mouse models harboring control *PB* donor elements have been developed, whereby we could compare their silencing through development and in diverse tissues, such strains were unavailable to us. However, abundant published literature has indicated that *PB* insertions tend not to be silenced even in the absence of imposed selection.

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 (Figure 45). 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.

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. For example, substantial recent evidence suggests that most new somatic L1 insertions mobilized during human cancer formation can mediate only minimal, if any, impacts on neighboring gene expression (Tubio et al., 2014). By contrast, 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.

Finally, the pattern of oscillation of gene expression (and silencing) uncovered through the use of TEM1 β-lactamase, may provide a basis for linking epigenetics to disease states, in addition to, and complementing what genome-wide association studies (GWAS) have contributed to the understanding of individual phenotypic variations (Armstrong, 2014; Kazazian, 2011). It has been proposed that individual variation between identical twins may be a result of differences due to somatic insertions between them. Over and above this, differential epigenetic statuses of existing L1s may also contribute epigenetic mosaicism in somatic cells. These areas remain to be further explored.

5. CONCLUSIONS AND FUTURE DIRECTIONS

The raison d'être and the biological impacts of retrotransposons in mammalian genomes, chiefly L1 elements, still are not completely understood. While some view L1s as evolutionary relics that have shaped mammalian genome evolution and which are now difficult to be gotten rid of, others strongly believe they have functional roles to play in an organism, a attractive hypothesis based on reports such as these: retrotransposons regulating gene expression in preimplantation embryos (Peaston et al., 2004); their polymorphic presence in mice strains (Akagi et al., 2008) and humans (Seleme et al., 2006) leading to genetic variability; involvement of L1s in X-chromosome involvement (Chow et al., 2010); and their movement in the nervous system resulting in somatic mosaicism (Thomas et al., 2012).

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.

The recent decade has also witnessed the emergence of epigenetics as an important field of biology and medicine. A historic event called the “Dutch Hunger Winter” of 1944 can be cited to bring out the relevance of epigenetic phenomena to human health and disease (Heijmans et al., 2008; Schulz, 2010). When the Nazi rulers cut off food and fuel supplies in September 1944 to western Holland, the resultant famine and starvation, combined with a harsh winter resulted in several deaths. Intriguingly, the children that were born from the

starving mothers who survived the Dutch hunger winter were more prone to health conditions such as diabetes, heart disorders, mental health illnesses. Such health impacts of pre-natal starvation continue to be felt even in the third generation, that is, in the grandchildren of the hunger winter mothers. Surprisingly, the underlying molecular basis of this trans-generational inheritance of an "acquired trait" was not a genetic mutation to the DNA, but what could be called as "epigenetic" mutation ("epimutation") in the genomes of the children and grandchildren of the affected mothers. The DNA of the affected patients was methylated differently that the healthy counterparts born of normally fed mothers (Heijmans et al., 2008; Schulz, 2010).

Transposons make a huge target for epigenetic regulation. Epigenetic phenomena such as DNA methylation have been long thought to be associated with control of transposable elements in the genome (Yoder et al., 1997), based on observations in both the embryo during normal development, as well as in systems deficient in DNA methylation machinery wherein transposons get unleashed. L1s also have been shown play a role in epigenetic silencing of the X-chromosome (Chow et al., 2010) and epigenetic silencing in cancers (Cruickshanks et al., 2013; Miousse and Koturbash, 2015).

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*, a DNA transposon, 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. We conclude that the host cellular and developmental contexts of L1 retrotransposition are significant determinants of the epigenetic controls established at new somatic integrants. We have proposed a model whereby the host epigenetic responses to new TE integrants reflect the timing, the molecular mechanism, and the genomic, cellular and developmental contexts of their mobilization.

In the future, this work can be extended in the following ways:

1. The biological importance of the mouse antisense promoter is yet to be ascertained. While studies have implicated human L1 antisense transcription with influencing gene expression in different cancers, a definitive role for antisense transcription in the development of an organism is yet to be ascertained. Recently, several long noncoding RNAs (ncRNAs) have been reported to be transcribed in mouse (Bergmann et al., 2015; Luo et al., 2013). Whether any of these ncRNAs have resulted due to antisense transcription can be verified; and if they do, it would be interesting to know what function those ncRNAs play. Similarly, from the study which reported L1's role in silencing X-chromosome (Chow et al., 2010) we can check if the transcripts originate from the ASP region.
2. We found that the antisense promoter in mouse L1 affects retrotransposition through a Dicer-independent mechanism. Investigations can be done to understand the alternative mechanism behind how L1 transposition is repressed by the ASP, such as those involving the RNA surveillance pathways (Gy et al., 2007).
3. The impacts of *de novo* insertions on expression of the neighbouring genes can be ascertained by looking at the epigenetic status of those genes before and after L1 integration events have occurred. A spread in silencing – either by perpetuation of the repressive histone marks, or by an indirect mechanism, such as that involving small RNA, from the new integrants to the neighbouring genes would support the idea of how transposable elements can change the epigenetic landscape of the genome (Cruickshanks et al., 2013).
4. More characterization of the L1 silencing in ES cells has to be done. To begin with, we can treat the transfected ES cells with HDAC inhibitors and check if the phenomenon of de-repression of the reporter we observed in HeLa cells is recapitulated in ES cells also.

5. Though we show it is HDAC-related, the molecular basis for oscillation of expression of β-lactamase reporters we observed in cell culture is not known. Chromatin immunoprecipitation (ChIP) experiments can be carried out to know the exact histone marks present in the cells that express the reporter (blue cells) and those that don't (reporter-silenced green cells) from the same clonal population of variegating cells. Also, the L1 reporter constructs can be transfected into a variety of cells lines that are lacking one or more epigenetic factors, in order to ascertain the epigenetic determinant/s that play crucial roles in mediating silencing.
6. Since ES cells are different from EC cells, it would be interesting to see the changes in epigenetic marks being induced as a result of differentiating the ES cells in various lineages, and compare them with the results obtained in EC cells (Garcia-Perez et al., 2010).

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

Abbas, H.K., Gronwald, J.W., Plaisance, K.L., Paul, R.N., and Lee, Y.W. (2001). Histone Deacetylase Activity and Phytotoxic Effects Following Exposure of Duckweed (Lemna pausicostata L.) to Apicidin and HC-Toxin. Phytopathology *91*, 1141-1148.

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.

Akagi, K., Stephens, R.M., Li, J., Evdokimov, E., Kuehn, M.R., Volfovsky, N., and Symer, D.E. (2010). MouseIndelDB: a database integrating genomic indel polymorphisms that distinguish mouse strains. Nucleic Acids Res *38*, D600-606.

Allis, D.C., Caparros, M., Jenuwein, T., Reinberg, D., Lachner, M. (2015). Epigenetics, Second edn (CSHL Press).

Alves, G., Tatro, A., and Fanning, T. (1996). Differential methylation of human LINE-1 retrotransposons in malignant cells. Gene *176*, 39-44.

An, W., Dai, L., Niewiadomska, A.M., Yetil, A., O'Donnell, K.A., Han, J.S., and Boeke, J.D. (2011). Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. Mob DNA *2*, 2.

An, W., Han, J.S., Schrum, C.M., Maitra, A., Koentgen, F., and Boeke, J.D. (2008). Conditional activation of a single-copy L1 transgene in mice by Cre. Genesis *46*, 373-383.

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.

Andrews, P.W., Matin, M.M., Bahrami, A.R., Damjanov, I., Gokhale, P., and Draper, J.S. (2005). Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. Biochemical Society transactions *33*, 1526-1530.

Andrews, P.W.a.I.D. (2013). Cell Lines from Human Germ-Cell Tumors. In Atlas of Human Tumor Cell Lines, J.-G.P. Robert K.M. Hay, Adi Gazdar ed. (Academic Press).

Aravin, A.A., and Bourc'his, D. (2008). Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes Dev *22*, 970-975.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Armstrong, L. (2014). Epigenetic predisposition to disease and imprinting-based disorders. In Epigenetics (Garland Science), pp. 209-210.

Babushok, D.V., Ostertag, E.M., Courtney, C.E., Choi, J.M., and Kazazian, H.H., Jr. (2006). L1 integration in a transgenic mouse model. Genome Res *16*, 240-250.

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.

Becker, K.G., Swergold, G.D., Ozato, K., and Thayer, R.E. (1993). Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet *2,* 1697-1702.

Belancio, V.P., Roy-Engel, A.M., and Deininger, P. (2008). The impact of multiple splice sites in human L1 elements. Gene *411,* 38-45.

Bergmann, J.H., Li, J., Eckersley-Maslin, M.A., Rigo, F., Freier, S.M., and Spector, D.L. (2015). Regulation of the ESC transcriptome by nuclear long noncoding RNAs. Genome research *25,* 1336-1346.

Bodak, M., Yu, J., and Ciaudo, C. (2014). Regulation of LINE-1 in mammals. Biomol Concepts *5,* 409-428.

Boissinot, S., Davis, J., Entezam, A., Petrov, D., and Furano, A.V. (2006). Fitness cost of LINE-1 (L1) activity in humans. Proc Natl Acad Sci U S A *103,* 9590-9594.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100,* 5280-5285.

Callahan, K.E., Hickman, A.B., Jones, C.E., Ghirlando, R., and Furano, A.V. (2012). Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. Nucleic acids research *40,* 813-827.

Callinan, P.A., and Batzer, M.A. (2006). Retrotransposable elements and human disease. Genome Dyn *1,* 104-115.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12,* 503-514.

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev *28,* 1397-1409.

Chen, J., Rattner, A., and Nathans, J. (2006). Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements. Hum Mol Genet *15,* 2146-2156.

Chen, J.M., Stenson, P.D., Cooper, D.N., and Ferec, C. (2005). A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet *117,* 411-427.

Chen, L., Dahlstrom, J.E., Lee, S.H., and Rangasamy, D. (2012). Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. Epigenetics *7,* 758-771.

Chen, X., Cui, J., Yan, Z., Zhang, H., Wang, N., Shah, P., Deng, F., Zhao, C., Geng, N., Li, M., *et al.* (2015). Sustained high level transgene expression in mammalian cells mediated by the optimized transposon system. Genes Dis *2,* 96-105.

Cheng, R.Y., Hockman, T., Crawford, E., Anderson, L.M., and Shiao, Y.H. (2004). Epigenetic and gene expression changes related to transgenerational carcinogenesis. Mol Carcinog *40,* 1-11.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E.*, et al.* (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956-969.

Conley, A.B., Miller, W.J., and Jordan, I.K. (2008). Human cis natural antisense transcripts initiated by transposable elements. Trends Genet *24*, 53-56.

Cost, G.J., Feng, Q., Jacquier, A., and Boeke, J.D. (2002). Human L1 element target-primed reverse transcription in vitro. Embo J *21*, 5899-5910.

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.

Cruickshanks, H.A., and Tufarelli, C. (2009). Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics *94*, 397-406.

Cruickshanks, H.A., Vafadar-Isfahani, N., Dunican, D.S., Lee, A., Sproul, D., Lund, J.N., Meehan, R.R., and Tufarelli, C. (2013). Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. Nucleic acids research *41*, 6857-6869.

Cummins, J.M., He, Y., Leary, R.J., Pagliarini, R., Diaz, L.A., Jr., Sjoblom, T., Barad, O., Bentwich, Z., Szafranska, A.E., Labourier, E.*, et al.* (2006). The colorectal microRNAome. Proc Natl Acad Sci U S A *103*, 3687-3692.

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.

Dai, L., Taylor, M.S., O'Donnell, K.A., and Boeke, J.D. (2012). Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. Mol Cell Biol *32*, 4323-4336.

Daly, C., and Reich, N.C. (1993). Double-stranded RNA activates novel factors that bind to the interferon-stimulated response element. Mol Cell Biol *13*, 3756-3764.

DeBerardinis, R.J., Goodier, J.L., Ostertag, E.M., and Kazazian, H.H., Jr. (1998). Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. Nat Genet *20*, 288-290.

Denli, A.M., Narvaiza, I., Kerman, B.E., Pena, M., Benner, C., Marchetto, M.C., Diedrich, J.K., Aslanian, A., Ma, J., Moresco, J.J.*, et al.* (2015). Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. Cell *163*, 583-593.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell *122*, 473-483.

Domansky, A.N., Kopantzev, E.P., Snezhkov, E.V., Lebedev, Y.B., Leib-Mosch, C., and Sverdlov, E.D. (2000). Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. FEBS Lett *472*, 191-195.

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.

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.

Ekram, M.B., Kang, K., Kim, H., and Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. Epigenetics : official journal of the DNA Methylation Society *7*, 370-382.

Eszterhas, S.K., Bouhassira, E.E., Martin, D.I., and Fiering, S. (2002). Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Mol Cell Biol *22*, 469-479.

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.

Ewing, A.D., and Kazazian, H.H., Jr. (2011). Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res *21*, 985-990.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., *et al.* (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat Genet *41*, 563-571.

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

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

Ferrigno, O., Virolle, T., Djabari, Z., Ortonne, J.P., White, R.J., and Aberdam, D. (2001). Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. Nat Genet *28*, 77-81.

Florl, A.R., Lower, R., Schmitz-Drager, B.J., and Schulz, W.A. (1999). DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. Br J Cancer *80*, 1312-1321.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev *5*, 1513-1523.

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.

Garrick, D., Fiering, S., Martin, D.I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nat Genet *18*, 56-59.

Gasior, S.L., Preston, G., Hedges, D.J., Gilbert, N., Moran, J.V., and Deininger, P.L. (2007). Characterization of pre-insertion loci of de novo L1 insertions. Gene *390*, 190-198.

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.

Goll, M.G., Anderson, R., Stainier, D.Y., Spradling, A.C., and Halpern, M.E. (2009). Transcriptional silencing and reactivation in transgenic zebrafish. Genetics *182*, 747-755.

Goodier, J.L., Ostertag, E.M., Du, K., and Kazazian, H.H., Jr. (2001). A novel active L1 retrotransposon subfamily in the mouse. Genome Res *11*, 1677-1685.

Goodier, J.L., Zhang, L., Vetter, M.R., and Kazazian, H.H., Jr. (2007). LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol Cell Biol *27*, 6469-6483.

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.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., *et al.* (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol *28*, 503-510.

Gy, I., Gasciolli, V., Lauressergues, D., Morel, J.B., Gombert, J., Proux, F., Proux, C., Vaucheret, H., and Mallory, A.C. (2007). Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. The Plant cell *19*, 3451-3461.

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.

Hancks, D.C., and Kazazian, H.H., Jr. (2012). Active human retrotransposons: variation and disease. Curr Opin Genet Dev *22*, 191-203.

Hedges, D.J., and Batzer, M.A. (2005). From the margins of the genome: mobile elements shape primate evolution. BioEssays : news and reviews in molecular, cellular and developmental biology *27*, 785-794.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences of the United States of America *105*, 17046-17049.

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.

Huang, C.R., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. Annu Rev Genet *46*, 651-675.

Ishihara, K., Hong, J., Zee, O., and Ohuchi, K. (2004). Possible mechanism of action of the histone deacetylase inhibitors for the induction of differentiation of HL-60 clone 15 cells into eosinophils. Br J Pharmacol *142*, 1020-1030.

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.

Jachowicz, J.W., and Torres-Padilla, M.E. (2015). LINEs in mice: features, families, and potential roles in early development. Chromosoma.

Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., *et al.* (2004). From silencing to gene expression: real-time analysis in single cells. Cell *116*, 683-698.

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.

Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., *et al.* (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res *14*, 331-342.

Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev *23*, 1303-1312.

Kazazian, H.H. (2011). Mobile DNA: Finding Treasure in Junk (FT Press Science.).

Kazazian, H.H., Jr. (1999). An estimated frequency of endogenous insertional mutations in humans. Nat Genet *22*, 130.

Kazazian, H.H., Jr., and Goodier, J.L. (2002). LINE drive. retrotransposition and genome instability. Cell *110*, 277-280.

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.

Kim, D.S., and Hahn, Y. (2010). Human-specific antisense transcripts induced by the insertion of transposable element. Int J Mol Med *26*, 151-157.

Kines, K.J., Sokolowski, M., deHaro, D.L., Christian, C.M., and Belancio, V.P. (2014). Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. Nucleic acids research *42*, 10488-10502.

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.

Kontgen, F., Suss, G., Stewart, C., Steinmetz, M., and Bluethmann, H. (1993). Targeted disruption of the MHC class II Aa gene in C57BL/6 mice. Int Immunol *5*, 957-964.

Kung, J.T., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651-669.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., *et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev *22*, 908-917.

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.

Lavie, L., Maldener, E., Brouha, B., Meese, E.U., and Mayer, J. (2004). The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. Genome Res *14*, 2253-2260.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., *et al.* (2012). Landscape of somatic retrotransposition in human cancers. Science *337*, 967-971.

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A., and Copeland, N.G. (2001). A Highly Efficient Escherichia coli-Based Chromosome Engineering System Adapted for Recombinogenic Targeting and Subcloning of BAC DNA. Genomics *73*, 56-65.

Lehner, B., Williams, G., Campbell, R.D., and Sanderson, C.M. (2002). Antisense transcripts in the human genome. Trends Genet *18*, 63-65.

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.

Li, J., Kannan, M., Trivett, A.L., Liao, H., Wu, X., Akagi, K., and Symer, D.E. (2014). An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. Nucleic acids research *42*, 4546-4562.

Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W., Guo, F., Liang, C., and Cen, S. (2013). The MOV10 helicase inhibits LINE-1 mobility. The Journal of biological chemistry *288*, 21148-21160.

Liu, J., Nau, M.M., Zucman-Rossi, J., Powell, J.I., Allegra, C.J., and Wright, J.J. (1997). LINE-I element insertion at the t(11;22) translocation breakpoint of a desmoplastic small round cell tumor. Genes Chromosomes Cancer *18*, 232-239.

Lobe, C.G., Koop, K.E., Kreppner, W., Lomeli, H., Gertsenstein, M., and Nagy, A. (1999). Z/AP, a double reporter for cre-mediated recombination. Developmental biology *208*, 281-292.

Luff, B., Pawlowski, L., and Bender, J. (1999). An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. Mol Cell *3*, 505-511.

Lund, A.H., and van Lohuizen, M. (2004). Epigenetics and cancer. Genes Dev *18*, 2315-2335.

Lunyak, V.V., Prefontaine, G.G., Nunez, E., Cramer, T., Ju, B.G., Ohgi, K.A., Hutt, K., Roy, R., Garcia-Diaz, A., Zhu, X., *et al.* (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science *317*, 248-251.

Luo, H., Sun, S., Li, P., Bu, D., Cao, H., and Zhao, Y. (2013). Comprehensive characterization of 10,571 mouse large intergenic noncoding RNAs from whole transcriptome sequencing. PLoS One *8*, e70835.

Macia, A., Munoz-Lopez, M., Cortes, J.L., Hastings, R.K., Morell, S., Lucena-Aguilar, G., Marchal, J.A., Badge, R.M., and Garcia-Perez, J.L. (2011). Epigenetic control of retrotransposon expression in human embryonic stem cells. Mol Cell Biol *31*, 300-316.

Marchetto, M.C., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C., Desai, K.N., Herai, R.H., Weitzman, M.D., *et al.* (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525-529.

Martin, S.L., and Branciforte, D. (1993). Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. Mol Cell Biol *13*, 5383-5392.

Martin, S.L., Bushman, D., Wang, F., Li, P.W., Walker, A., Cummiskey, J., Branciforte, D., and Williams, M.C. (2008). A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity. Nucleic Acids Res *36,* 5845-5854.

Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S.C., Hodges, R.S., and Williams, M.C. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. Journal of molecular biology *348,* 549-561.

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.

Matlik, K., Redik, K., and Speek, M. (2006). L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol *2006,* 71753.

Medstrand, P., Landry, J.R., and Mager, D.L. (2001). Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J Biol Chem *276,* 1896-1903.

Meir, Y.J., Weirauch, M.T., Yang, H.S., Chung, P.C., Yu, R.K., and Wu, S.C. (2011). Genome-wide target profiling of piggyBac and Tol2 in HEK 293: pros and cons for gene discovery and gene therapy. BMC Biotechnol *11,* 28.

Melquist, S., Luff, B., and Bender, J. (1999). Arabidopsis PAI gene arrangements, cytosine methylation and expression. Genetics *153,* 401-413.

Menendez, L., Benigno, B.B., and McDonald, J.F. (2004). L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. Mol Cancer *3,* 12.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., 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.

Miousse, I.R., and Koturbash, I. (2015). The Fine LINE: Methylation Drawing the Cancer Landscape. Biomed Res Int *2015,* 8.

Molaro, A., Falciatori, I., Hodges, E., Aravin, A.A., Marran, K., Rafii, S., McCombie, W.R., Smith, A.D., and Hannon, G.J. (2014). Two waves of de novo methylation during mouse germ cell development. Genes Dev *28,* 1544-1549.

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.

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nat Genet *23,* 314-318.

Mourier, T., and Willerslev, E. (2009). Retrotransposons and non-protein coding RNAs. Brief Funct Genomic Proteomic *8,* 493-501.

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.

Naas, T.P., DeBerardinis, R.J., Moran, J.V., Ostertag, E.M., Kingsmore, S.F., Seldin, M.F., Hayashizaki, Y., Martin, S.L., and Kazazian, H.H. (1998). An actively retrotransposing, novel subfamily of mouse L1 elements. Embo J *17,* 590-597.

Nakanishi, H., Higuchi, Y., Kawakami, S., Yamashita, F., and Hashida, M. (2010). piggyBac transposon-mediated long-term gene expression in mice. Mol Ther *18,* 707-714.

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.

O'Donnell, K.A., and Boeke, J.D. (2007). Mighty Piwis defend the germline against genome intruders. Cell *129,* 37-44.

Orgel, L.E., and Crick, F.H. (1980). Selfish DNA: the ultimate parasite. Nature *284,* 604-607.

Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., and Kazazian, H.H., Jr. (2002). A mouse model of human L1 retrotransposition. Nat Genet *32,* 655-660.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001a). Biology of mammalian L1 retrotransposons. Annu Rev Genet *35,* 501-538.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001b). Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res *11,* 2059-2065.

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res *28,* 1418-1423.

Pannell, D., Osborne, C.S., Yao, S., Sukonnik, T., Pasceri, P., Karaiskakis, A., Okano, M., Li, E., Lipshitz, H.D., and Ellis, J. (2000). Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. EMBO J *19,* 5884-5894.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell *7,* 597-606.

Peddigari, S., Li, P.W., Rabe, J.L., and Martin, S.L. (2013). hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. Nucleic acids research *41,* 575-585.

Perepelitsa-Belancio, V., and Deininger, P. (2003). RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet *35,* 363-366.

Pornthanakasem, W., and Mutirangura, A. (2004). LINE-1 insertion dimorphisms identification by PCR. Biotechniques *37,* 750, 752.

Quinlan, A.R., Boland, M.J., Leibowitz, M.L., Shumilina, S., Pehrson, S.M., Baldwin, K.K., and Hall, I.M. (2011). Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. Cell Stem Cell *9,* 366-373.

Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Lower, J., Stratling, W.H., Lower, R., and Schumann, G.G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic acids research *40,* 1666-1683.

Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I., and Whitelaw, E. (2002). Metastable epialleles in mammals. Trends Genet *18*, 348-351.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V., and Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci U S A *100*, 2538-2543.

Rangwala, S.H., and Kazazian, H.H., Jr. (2009). The L1 retrotransposition assay: A retrospective and toolkit. Methods.

Rangwala, S.H., Zhang, L., and Kazazian, H.H., Jr. (2009). Many LINE1 elements contribute to the transcriptome of human somatic cells. Genome Biol *10*, R100.

Rebollo, R., Romanish, M.T., and Mager, D.L. (2012). Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annu Rev Genet.

Reichmann, J., Crichton, J.H., Madej, M.J., Taggart, M., Gautier, P., Garcia-Perez, J.L., Meehan, R.R., and Adams, I.R. (2012). Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. PLoS computational biology *8*, e1002486.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., *et al.* (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552-556.

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.

Rival-Gervier, S., Lo, M.Y., Khattak, S., Pasceri, P., Lorincz, M.C., and Ellis, J. (2013). Kinetics and epigenetics of retroviral silencing in mouse embryonic stem cells defined by deletion of the D4Z4 element. Mol Ther *21*, 1536-1550.

Rodic, N., and Burns, K.H. (2013). Long interspersed element-1 (LINE-1): passenger or driver in human neoplasms? PLoS genetics *9*, e1003402.

Rosser, J.M., and An, W. (2010). Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation. Gene *456*, 15-23.

Saxton, J.A., and Martin, S.L. (1998). Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. J Mol Biol *280*, 611-622.

Schulz, L.C. (2010). The Dutch Hunger Winter and the developmental origins of health and disease. Proceedings of the National Academy of Sciences of the United States of America *107*, 16757-16758.

Schumann, G.G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochemical Society transactions *35*, 637-642.

Seleme, M.C., Vetter, M.R., Cordaux, R., Bastone, L., Batzer, M.A., and Kazazian, H.H., Jr. (2006). Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. Proc Natl Acad Sci U S A *103*, 6611-6616.

Shao, Y., Gao, Z., Marks, P.A., and Jiang, X. (2004). Apoptotic and autophagic cell death induced by histone deacetylase inhibitors. Proceedings of the National Academy of Sciences of the United States of America *101*, 18030-18035.

Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S.*, et al.* (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell *153*, 101-111.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*, 103-113.

Soifer, H.S., Zaragoza, A., Peyvan, M., Behlke, M.A., and Rossi, J.J. (2005). A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. Nucleic Acids Res *33*, 846-856.

Solyom, S., Ewing, A.D., Rahrmann, E.P., Doucet, T.T., Nelson, H.H., Burns, M.B., Harris, R.S., Sigmon, D.F., Casella, A., Erlanger, B.*, et al.* (2012). Extensive somatic L1 retrotransposition in colorectal tumors. Genome research.

Sookdeo, A., Hepp, C.M., McClure, M.A., and Boissinot, S. (2013). Revisiting the evolution of mouse LINE-1 in the genomic era. Mob DNA *4*, 3.

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

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell *134*, 587-598.

Suter, C.M., Martin, D.I., and Ward, R.L. (2004). Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int J Colorectal Dis *19*, 95-101.

Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Developmental biology *269*, 276-285.

Symer, D.E., and Bender, J. (2001). Genomic stability. Hip-hopping out of control. Nature *411*, 146-147, 149.

Symer, D.E., and Boeke, J.D. (2010). An everlasting war dance between retrotransposons and their metazoan hosts. In Retroviruses: Molecular BIology, Genomics and Pathogenesis, B.N. Kurth R, ed. (Norwich, U.K., Caister Academic Press), pp. 1-33.

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.

Szak, S.T., Pickeral, O.K., Makalowski, W., Boguski, M.S., Landsman, D., and Boeke, J.D. (2002). Molecular archeology of L1 insertions in the human genome. Genome Biol *3*, research0052.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M.*, et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

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

Thomas, C.A., Paquola, A.C., and Muotri, A.R. (2012). LINE-1 retrotransposition in the nervous system. Annu Rev Cell Dev Biol *28*, 555-573.

Tubio, J.M., Li, Y., Ju, Y.S., Martincorena, I., Cooke, S.L., Tojo, M., Gundem, G., Pipinikas, C.P., Zamora, J., Raine, K.*, et al.* (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science *345*, 1251343.

Van de Lagemaat, L.N., Landery, J.-R., Mager, D.L., and Medstrand, P. (2003a). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends in Genetics *19*, 530-536.

van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. (2003b). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet *19*, 530-536.

van den Hurk, J.A., Meij, I.C., Seleme, M.C., Kano, H., Nikopoulos, K., Hoefsloot, L.H., Sistermans, E.A., de Wijs, I.J., Mukhopadhyay, A., Plomp, A.S.*, et al.* (2007). L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet *16*, 1587-1592.

Vanhee-Brossollet, C., and Vaquero, C. (1998). Do natural antisense transcripts make sense in eukaryotes? Gene *211*, 1-9.

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev *20*, 1732-1743.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

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.

Weber, B., Kimhi, S., Howard, G., Eden, A., and Lyko, F. (2010). Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene *29*, 5775-5784.

Wei, W., Morrish, T.A., Alisch, R.S., and Moran, J.V. (2000). A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal Biochem *284*, 435-438.

Whitelaw, E., and Martin, D.I. (2001). Retrotransposons as epigenetic mediators of phenotypic variation in mammals. Nat Genet *27*, 361-365.

Wissing, S., Munoz-Lopez, M., Macia, A., Yang, Z., Montano, M., Collins, W., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2012). Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. Human molecular genetics *21*, 208-218.

Xie, Y., Rosser, J.M., Thompson, T.L., Boeke, J.D., and An, W. (2011). Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. Nucleic acids research *39*, e16.

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.

Yang, N., Zhang, L., Zhang, Y., and Kazazian, H.H., Jr. (2003). An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res *31,* 4929-4940.

Yant, S.R., Meuse, L., Chiu, W., Ivics, Z., Izsvak, Z., and Kay, M.A. (2000). Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nature genetics *25,* 35-41.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13,* 335-340.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature *451,* 202-206.

Zamudio, N., and Bourc'his, D. (2010). Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity *105,* 92-104.

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.

Zhao, K., Du, J., Han, X., Goodier, J.L., Li, P., Zhou, X., Wei, W., Evans, S.L., Li, L., Zhang, W., *et al.* (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. Cell Rep *4,* 1108-1115.

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.

PUBLICATIONS FROM THIS WORK

Li J, Kannan M, Trivett AL, Liao H, Wu X, Akagi K, Symer DE. *An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition*. Nucleic Acids Res. 2014; 42(7): 4546-62.

Kannan M, Li J, Fritz SE, Halleck KE, Sanford JC, Sullivan TL, Tiwari PK, An W, Boeke JD, Symer DE. *Dynamic silencing of somatic L1 retrotransposon integrants reflects the genomic, cellular and developmental contexts of their mobilization.* 2015 (manuscript under preparation).

BRIEF BIOGRAPHICAL SKETCHES

Mr. Manoj Kannan currently works as a lecturer (visiting) in the Department of Biological Sciences at Birla Institute of Technology and Science (BITS), Pilani. While working for his masters (integrated) degree in Information Systems at BITS, he decided to switch fields and move into biology. Subsequently, he obtained his second masters degree in biotechnology from BITS, before going on to pursue Ph.D. at the National Cancer Institute, Frederick, Maryland (USA) on a Fogarty International Fellowship offered by the National Institutes of Health. He carried out research in the laboratory of Dr. David Symer in the field of L1 transposons. With keen interest in teaching, Mr. Kannan has been handling theory and laboratory courses in general biology, genetics and recombinant DNA technology at BITS Pilani since 2008.

Dr. David Symer is an assistant professor in the Human Cancer Genetics Program, Department of Molecular Virology, Immunology and Medical Genetics and Department of Internal Medicine, Division of Hematology at the Ohio State University Comprehensive Cancer Center – James Cancer Hospital. Dr. Symer is also the director of the core laboratory offering shared resources in genomics at Ohio State, the Nucleic Acids / Microarray Shared Resource. Dr. Symer has ongoing research interests in molecular biology and in cancer genetics and genomics. His recent interests are in studying biological impacts of retrotransposons and human papilloma virus, and diseases associated with aberrant RNA splicing. He graduated with an AB degree in mathematics from Dartmouth College, earned his M.D. and Ph.D. degrees at Johns Hopkins University School of Medicine, did his residency at Brigham and Women's Hospital (Harvard Medical School), and completed his fellowship training at Brigham and Women's Hospital and Johns Hopkins Oncology Center. Dr. Symer was a principal investigator at the National Cancer Institute prior to moving to Ohio State in 2009. In November 2014, he was presented the High Performance Computing Innovation Excellent Award by International Data Corporation at the Supercomputing 2014 meeting in New Orleans, Louisiana. He has journal publications in several high-impact journals such as *Nature*, *Science* and *Cell* to his credit.