KEEPING L1S IN LINE: GENOMIC DEFENSES AGAINST RETROTRANSPOSITION

Evan A. Farkash

A DISSERTATION
in
Immunology

Presented to the Faculties of the University of Pennsylvania
in Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy

2007

[Signature]
Supervisor of Dissertation

[Signature]
Graduate Group Chair
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
Acknowledgements

I thank my mentor, Nina Luning Prak, for her generosity and guidance. She has boundless creativity and curiosity, and the rare gift of being able to impart these traits to her peers and protégés.

I thank my wife, Cherie, for her patience, trust, love, and understanding on this long and interesting road that we travel together.

I thank my family, Mom, Dad, Mathew, Marsha, Jimmy, Allyson, and Jamye, for their love and encouragement. Without you, this would not be possible.

I thank my thesis committee, Avinash Bhandoola, Marisa Bartolomei, Wafik El-Deiry, and Gary Koretzky, for their time and constructive critiques. I also thank Prescott Deininger for being the external reviewer for this dissertation.

I thank my collaborators, Gary Kao, Petr Svoboda, and John Moran, for their scientific expertise and direction. I thank Brad Johnson, Haig Kazaizian, and members of their respective labs for interesting and helpful discussions.

I thank Doug Cines and the National Institutes of Health for providing my predoctoral training support.

Finally, I thank past and present members of the Luning Prak lab, Shane Horman, Peffin Lee, Sara Smith, Zach Faber, Noah Goodman, Yang Zhu Du, Lenka Yunk, Anil Panigrahi, Andrew Fesnak, Mike Frank, Deep Chandra, Sarah Fox, Janet Sallit, Julia Sarino, and Deborah Sekiguchi for their camaraderie and technical assistance. Every day is a new adventure.
ABSTRACT

KEEPING L1S IN LINE: GENOMIC DEFENSES AGAINST RETROTRANSPOSITION

Evan A. Farkash
Nina Luning Prak

Long Interspersed Elements (L1) are mobile elements responsible for shaping as much as 45% of the human genome. Highly active L1 retrotransposons are agents of genomic instability in vivo. Conversely, L1s might be activated by genomic stress. Here I report on one form of DNA damage, gamma radiation, which is able to increase the retrotransposition frequency of a tagged L1 in tissue culture. Gamma radiation likely enhances L1 mobility by inducing a more favorable cellular environment for endonuclease-dependent retrotransposition. Because L1 mobility is toxic to cells, the survival of an organism may be contingent on its ability to regulate L1 elements. One strategy of L1 regulation is to limit the cell types in which it is active; namely, by restricting L1 expression and retrotransposition to germ cells. New L1 insertions also appear to be subject to regulation. CpG methylation appears to play a role in establishing tissue specific silencing of L1 insertions. L1 insertions are differentially methylated in somatic and germ tissues. Moreover, new L1 insertions are not downregulated in cells grown in tissue culture under demethylating conditions. However, maintenance of L1 repression is likely to involve multiple pathways, as even a toxic dose of demethylating agents is not sufficient to re-express a silenced L1 insertion. DNA methylation is likely to be an important weapon in a cell’s arsenal to defend the integrity of the genome against assault by mobile elements.
Table of Contents

Chapter 1: Introduction ...........................................................................................1

1.1 L1 elements are highly successful, autonomous genomic parasites ..................1
1.2 The impact of L1 retrotransposition on the host genome ....................................2
1.3 Transposons are activated by genomic stress ......................................................6
1.4 The cellular response to DNA damage is complex .............................................7
1.5 L1 RNA levels and genotoxic stress .................................................................8
1.6 Altered L1 integration during DNA damage ......................................................10
1.7 Cellular co-factors and inhibitors of retrotransposition .......................................12
1.8 Germline-only retrotransposition maximizes L1 proliferation while minimizing its impact on the host .................................................................15
1.9 Methylation can affect L1 mobility .......................................................................18

Chapter 2: Gamma Radiation Increases L1 Retrotransposition .........................22

Introduction ..........................................................................................................22
Results ..................................................................................................................24
2.1 L1 elements induce γH2AX foci in human osteosarcoma cells .............................24
2.2 Gamma radiation increases L1 retrotransposition in a cultured cell assay ..........26
2.3 Gamma radiation does not increase L1-EGFP RNA levels .................................29
2.4 Other genotoxic agents do not increase L1 retrotransposition .........................29
2.5 Gamma radiation and calicheamicin γ1 generate similar numbers of DSBs ........31
2.6 Endonuclease deficient L1 retrotransposition does not increase with gamma irradiation .................................................................33
2.7 Genomic flanks of novel insertions have endonuclease-dependent features .......35
2.7 Genomic flanks of novel insertions have endonuclease-dependent features .......36
Discussion .........................................................................................................38

Chapter 3: Expression of new L1 insertions is restricted to the testis ...............45

Introduction ..........................................................................................................45
3.1 A transgenic mouse model of L1 retrotransposition expresses L1-EGFP in the testis ....46
3.2 L1 retrotransposition is limited to the testis in a mouse model ............................49
3.3 Insertion of L1-EGFP in the germline and during early embryogenesis ...............52
3.4 L1 insertions express EGFP primarily in testis ..................................................56
3.5 L1-EGFP insertions are expressed in a subset of cells in the testis ......................60
3.6 Expression of L1 insertions is restricted during embryonic development ..........65
Discussion .........................................................................................................69

Chapter 4: Methylation of CpG DNA in Novel L1 Insertions .........................79

Introduction ..........................................................................................................79
Results ..................................................................................................................79
4.1 EGFP is driven by a strong promoter that is active in multiple mouse tissues .......79
4.2 L1 inserts into transcriptionally permissive chromatin .......................................80
4.3 EGFP and its CMV-IE promoter are CpG rich ................................................85
4.4 The L1-EGFP transgene and L1-EGFP insertions are hypomethylated in the testis ........86
4.4 The L1-EGFP transgene and L1-EGFP insertions are hypomethylated in the testis ....87
4.5 Demethylation increases expression from L1-EGFP insertions in cultured cells ....89

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
List of Figures

Figure 1-1. L1 regulation can occur at multiple stages of the L1 lifecycle. 3
Figure 1-2. Potential mechanisms of gamma radiation induced retrotransposition. 11
Figure 1-3. Game theory analysis of retrotransposition outcomes in somatic and germ cells. 16

Figure 2-1. Active L1 elements induce formation of γH2AX foci. 25
Figure 2-2. L1-EGFP is a human L1 element that expresses an EGFP reporter following retrotransposition. 27
Figure 2-3. Gamma radiation increases the L1 retrotransposition frequency. 28
Figure 2-4. L1 RNA is not increased in cells exposed to gamma radiation. 30
Table 2-1. Most genotoxic agents do not increase L1 retrotransposition in a cultured cell assay. 32
Figure 2-5. Gamma radiation and Calicheamicin γ1 generate comparable numbers of γH2AX foci. 33
Figure 2-6. Gamma radiation does not greatly increase endonuclease-independent L1 retrotransposition in CHO-K1 cells. 35
Figure 2-7. L1 insertions from irradiated clones have endonuclease-dependent features. 37

Figure 3-1. A human L1 is used to generate a transgenic model of L1 retrotransposition. 47
Figure 3-2. Strand specific amplification detects L1-EGFP transcripts in the testis. 49
Figure 3-3. Detection of L1-EGFP insertions in the mouse testis by flow cytometry. 51
Figure 3-4. The offspring of L1-EGFP transgenic mice have germine insertions. 53
Table 3-1. Analysis of L1-EGFP insertions in transgenic mice. 55
Figure 3-5. L1-EGFP insertions are polymorphic in 5’ truncation by ruler PCR. 55
Figure 3-6. The L1-EGFP insertion is present in multiple tissues in a mouse with an early embryonic L1-EGFP insertion. 57
Figure 3-7. Mice with L1-EGFP insertions express EGFP in cells of the testis. 58
Figure 3-8. Tubules are chimeric for EGFP expression in a mouse with an early embryonic L1-EGFP insertion. 59
Figure 3-9. EGFP transcripts are detected only in the testis in L1-EGFP insertion mice. 61
Figure 3-10. Heterogeneous EGFP expression within the tubules of mice with a germline L1-EGFP insertion. 62
Figure 3-11. L1-EGFP is expressed at multiple stages of germ cell development. 64
Figure 3-12. Helical EGFP expression along a transverse tubule section. 66
Figure 3-13. Genital ridges from day 14.5 embryos are not fluorescent. 67
Figure 3-14. Genital ridges from D14.5 embryos do not express EGFP. 68
Figure 3-15. L1 insertions are regulated during development. 76

Table 4-1. CMV-IE promoter is expressed in multiple murine tissues. 81
Figure 4-1. Mouse fibroblasts are fluorescent after transfection with an EGFP expression plasmid. 82
Figure 4-2. L1-EGFP inserted into a gene transcribed in multiple somatic tissues. 84
Figure 4-3. L1 5’UTR, CMV-IE promoter, and EGFP coding regions are all CpG rich. 86
Figure 4-4. L1-EGFP insertions are relatively unmethylated in testis. 88
Figure 4-5. The L1-EGFP transgene is relatively unmethylated in testis. 90
Figure 4-6. 5’Aza-2’deoxycytidine increases the fluorescence from L1-EGFP insertions, but not their frequency. 92
Figure 4-7. 5’Aza-2’deoxycytidine increases EGFP RNA in cells transfected with L1-EGFP. 93
Figure 4-8. 5’Aza-2’deoxycytidine does not induce EGFP expression in somatic tissues. 94
Table 4-2. Combination therapy of DNA demethylating and histone acetylating agents does not permit EGFP expression or retrotransposition in primary cells. 97
Figure 4-9. A model of L1-EGFP insertion regulation by methylation during embryogenesis. 105

Figure 5-1. Retrotransposition in chicken lymphocytes deficient in DNA repair enzymes. 112
Figure 5-2. Artemis deficiency increases EGFP+ cells in chicken lymphocytes. 113
Figure 5-3. Artemis inhibits L1 retrotransposition by digesting insertion intermediates. 116
Abbreviations

ATM  Ataxia telangectasia mutation
BLM  Bloom Syndrome
CMV  Cytomegalovirus
NHEJ  Non-homologous end joining
DNA-PKcs  DNA Protein Kinase catalytic subunit
DSB  Double Strand Break
EGFP  Enhanced Green Fluorescent Protein
EN  Endonuclease
H2AX  Histone H2AX
γH2AX  Phosphorylated Histone H2AX
HR  Homologous recombination
HSVtk  Herpes Simplex Virus thymidine kinase
IAP  Intracistemal A Particle
L1  Long Interspersed Element 1
LINE1  Long Interspersed Element 1
LTR  Long terminal repeat
ORF  Open reading frame
pPolII  Polymerase II major subunit promoter
RT  Reverse transcriptase
SINE  Short Interspersed Element
SV40  Simian virus 40
TSA  Trichostatin A
TSD  Target site duplication
UTR  Untranslated region
WRN  Werner syndrome
YY1  YinYang 1
Chapter 1: Introduction

1.1 L1 elements are highly successful, autonomous genomic parasites

Long Interspersed Elements (LINE1 or L1) are a highly successful class of mobile elements that are found in the genomes of all sexually reproducing eukaryotes. The human genome contains some 500,000 L1 elements comprising 17% of the mass of the genome [1]. In addition, L1s are thought to be responsible for the mobilization of Short Interspersed Elements (SINE) and processed pseudogenes [2]. L1s continue to proliferate in the human genome today [3]. Although most of the elements in the human genome have been inactivated by 5' truncation or disabling mutations, it is estimated that the average human genome contains 80 full length L1 elements capable of retrotransposing [4, 5]. However, most new insertions are the result of a few highly active elements [6].

L1s are autonomous, in that a full length L1 element encodes much of the enzymatic machinery required for proliferation. A full length L1 element is 6 kb long and codes for two proteins, ORF1 and ORF2. ORF1 is a novel protein that binds single stranded RNA and DNA, acts as a chaperone for nucleic acids, and participates in strand exchange during L1 integration [7-10]. ORF2 contains an endonuclease domain, a reverse transcriptase domain, and a cystine-rich domain [11]. An ORF1 mutated L1 does not complement an ORF2 mutated L1, suggesting that L1 proteins bind and mobilize the RNA that encoded them in cis [12]. In addition, L1 contains two untranslated regions.
(UTR). The 5' UTR is thought to be an internal promoter and contains several transcription factor binding regions. The 3' UTR contains a weak poly-A sequence [11].

L1 proliferates throughout its host genome by a complex, copy-and-paste lifecycle (Figure 1-1). A full length RNA encoding the ORF1 and ORF2 proteins is transcribed from a retrotransposition-competent L1. L1 RNA is exported to the cytoplasm where its encoded ORF1 and ORF2 proteins are translated. This protein-RNA complex returns to the nucleus, where the endonuclease domain of ORF2 nicks the target site. The reverse transcriptase domain of ORF2 creates a cDNA copy using the target site's 5' overhang as a primer. Subsequent displacement of the mRNA by a complementary strand of cDNA and ligation of the breaks are thought to require host machinery [13].

1.2 The impact of L1 retrotransposition on the host genome

The proliferation of L1 elements is not without risk to the host. Retrotransposition into or near genes can disrupt their expression, causing disease. An L1 insertion into an exon can physically interrupt coding sequences [14]. L1 insertion can also result in alternately spliced gene products, either by skipping exons or creating new donor and acceptor sites [15-17]. An L1 insertion into a gene may decrease full length transcripts by introducing a premature polyadenylation signal [18]. In addition, the presence of L1 sequences near genes downregulates or silences their expression via a poorly understood mechanism [19, 20]. In humans, L1 insertions are a cause of
Figure 1-1. *L1 regulation can occur at multiple stages of the L1 lifecycle.* 1. Transcription of the L1 element is controlled by epigenetic factors and transcription factors. 2. L1 RNA is exported to the cytoplasm, where its copy number influences retrotransposition frequency. 3. Translation of ORF1 and ORF2 proteins. 4. L1 protein and mRNA are imported into the nucleus, where ORF2 endonuclease creates a DNA double strand break. Induced breaks may be able to serve as alternative substrates for insertion. 5. ORF2 reverse transcribes a cDNA copy of L1 at the insertion site. Host factors are thought to inhibit or assist in resolution of the insertion.
hemophilia A, muscular dystrophy, retinitis pigmentosa, colon cancer, and other conditions [21-23]. Many of these conditions are X-linked. L1s have also been infrequently observed to generate large chromosomal deletions when inserting into the genome [24]. Because L1 and other repetitive sequences exist at high copy number in the genome, recombination events between elements can also cause chromosomal deletions or even translocations [14]. The insertion of L1 elements can be hazardous to the integrity of its host genome. Therefore, high levels of L1 retrotransposition are likely to be deleterious to host fitness.

Although controversial, it is generally agreed that L1 elements do provide the occasional benefit for the host. First, telomerase, the enzyme responsible for extending telomeres, was originally derived from a retrotransposon gene and now has the function of enhancing genomic stability [25, 26]. Telomere maintenance has also been described as a novel function of the Drosophila retrotransposons, HeT-A and TART [27]. Second, it is hypothesized that L1 elements have created genetic diversity through a process known as exon shuffling [28]. The L1 has a weak poly-A signal, and read-through transcripts can incorporate exons if they are located in the 3' genomic flank of the element [28]. If this L1 retrotransposes into the genome, the exon will be carried with it in a process termed 3' transduction [29]. Insertion of the exon into a new gene in the sense orientation can create a novel protein product. If the introduction of the new exon contributes added functionality to the protein, it might provide a selective advantage. L1 mediated mobilization of an exon from the Cyclophilin A gene into the TRIM5 gene is thought to be responsible for owl monkey resistance to HIV [30]. Similar events have
been observed in the human genome [31, 32]. Third, L1s may also be able to insert into
damaged DNA under some conditions, acting as an alternate mechanism of double strand
break repair [33].

L1s have the potential to be used as tools in the basic and clinical sciences.
Because L1 retrotransposition disrupts genes, L1s can be used as insertional mutagens.
The main advantage of L1s over chemical or high energy particle mutagenesis is that the
L1 insertion is present in the disrupted gene. An L1 tagged with a unique sequence can
enable identification of the mutated gene. Another potential use for L1 elements is as
vectors for gene therapy. L1 elements carrying reporter genes have been shown to
successfully integrate into the genome and express the reporter [34]. If provided a
delivery mechanism into the cytoplasm, an L1 could be used to integrate a therapeutic
gene into a patient’s genome in single copy [35]. Since L1s are present in all human
cells, they would be recognized as “self” by the patient’s immune system. In contrast,
cells transfected by viral vectors may display “non-self” viral proteins and be targeted for
destruction [36].

A lack of understanding of the mechanism of retrotransposition and how host
cells regulate L1 mobility currently limits the use of L1s in basic or applied research. In
the course of researching L1 elements, I became interested in two questions related to L1
retrotransposition:
• First, how is L1 retrotransposition affected under conditions of genomic stress?
• Second, in what cell types are L1 insertions expressed, and which cellular processes mediate this effect?

1.3 Transposons are activated by genomic stress

Barbara McClintock, in her 1983 Nobel Address, was the first to propose that genomic stress assisted in the mobilization of transposons [37]. The generation of double stranded DNA breaks (DSBs) is a severe form of genotoxic stress that threatens the integrity of the genome, activates cell cycle checkpoints and, in some cases, causes cell death. While direct evidence for the activation of L1 retrotransposition by DNA damage is still sparse, there is a growing body of data that other mobile elements can be activated by DNA damage. For example, Barbara McClintock initially observed Ac/Ds element transposition in response to chromosomal translocations [37, 38]. Indeed, some transposable elements, including P elements in Drosophila and the synthetic Sleeping Beauty element, appear to be activated by DNA damage and repair processes [39-41]. Mobilization is not limited to DNA transposons: various forms of DNA damage activate retrotransposition of long terminal repeat (LTR) and non-LTR retrotransposons including Gypsy and I factor in Drosophila and Ty1 in yeast [42-47]. Even closer to home for L1, transcription and retrotransposition of Alu elements are increased when cells are exposed to etoposide, a topoisomerase II inhibitor that produces DSBs [48, 49]. This is relevant to L1s because Alu elements are thought to co-opt L1 proteins for their mobilization, so an increase Alu retrotransposition may reflect increased L1 mobility [2]. In a genomic
screen of mice exposed to gamma irradiation, new SINE and L1 insertions were detected, but it was unresolved if the frequency of new insertions was significantly different in irradiated compared to unirradiated controls [50].

1.4 The cellular response to DNA damage is complex

There are many chemical agents and natural processes that have the ability to damage DNA. UV light, X-rays, chemotherapeutic drugs, cigarette smoke, and even cell division have the potential to generate DNA lesions [51]. Depending on the source of DNA damage, the structure of the DNA break and its mechanism of repair may be different. Oxidative damage creates DNA double strand breaks that are repaired by non-homologous end joining [52]. Nucleotide base damage and dimer formation induced by UV rays during sun exposure are repaired by base excision repair [53]. Stalled replication forks in dividing cells are repaired by homologous recombination [54].

Shortly after the induction of a DSB, complex signaling pathways are activated [55]. These signaling cascades recruit DNA repair factors to DSBs, alter transcription and trigger cell fate decisions. Significant damage may trigger cell cycle arrest, or even apoptosis. Various cellular events occurring secondary to DNA damage may affect L1 retrotransposition. Because the cellular response can vary depending on the type of lesion and cell type, the effects on L1 retrotransposition could depend on the context of DNA damage. Three likely ways that DNA damage could influence L1 retrotransposition are by modulating L1 RNA levels, altering the mechanism of L1 entry.
into the genome, and affecting the bioavailability of cellular cofactors and inhibitors of retrotransposition.

1.5 L1 RNA levels and genotoxic stress

Several lines of evidence suggest that L1 RNA abundance is critical and rate-limiting for L1 retrotransposition. L1 RNA is required for retrotransposition, not only because it encodes the machinery needed for L1 to retrotranspose, but because the RNA itself serves as a replication intermediate (Figure 1-1). That L1 transcript abundance is rate-limiting for retrotransposition is suggested by studies in cultured cells with tagged L1 elements showing that decreased L1 mRNA levels result in reduced retrotransposition [56]. Conversely, increased L1 RNA levels have been observed for highly active L1 elements [57]. Furthermore, the correlation between RNA levels and retrotransposition frequency is not unique to L1 retrotransposons: Ty1 elements in yeast appear to retrotranspose in direct proportion to the amount of Ty1 mRNA [46, 58, 59].

Given that RNA is important for L1 mobility, DNA damage might influence L1 transcript abundance. While there are no published data that compare L1 RNA levels in irradiated and unirradiated cells, there is evidence that RNA levels of other retrotransposons are influenced by DNA damage. For example, gamma radiation has been shown to increase Ty1 RNA in yeast [46] and IAP RNA in murine myeloid cells [60]. Furthermore, murine and human cell lines expressing the Bcl-2 survival gene...
exhibit an increase in endogenous Alu mRNA levels following exposure to gamma radiation, UV, etoposide, and cisplatin [48].

Since the induction of DNA damage has an extensive effect on the transcriptional profile of a cell [61], it is plausible that L1 RNA levels are differentially regulated following gamma radiation. One way to regulate L1 expression following DNA damage is to alter transcription factor levels or binding activity. The 5’UTR of the L1 contains an internal promoter element [62-64] with putative binding sites for SRY family members [65], YY1 [66], and RUNX3 [56]. DNA damage could modulate L1 activity by acting through factors that bind these sites.

Binding of the SRY family member, SOX11, to the L1 5’UTR was shown to increase L1 retrotransposition, promoter activity, and RNA copy number [65]. More recently, binding of SOX2 has been shown to inhibit L1 promoter activity in rat hippocampal neuronal stem cells [67]. SOX2 and SOX11 possess high mobility group domains, which have been shown to bind to cisplatin-DNA adducts [68]. If SRY family members are differentially recruited to the sites of DNA damage, then this could alter the profile of transcription factors at the L1 5’UTR.

Another L1 transcription factor that may be affected by DNA damage is the ubiquitous YinYang1 (YY1) factor. YY1 is thought to facilitate the production of full length L1 mRNAs [69]. In response to exposure to methyl-N-nitro-N-nitrosoguanidine, YY1 was polyADP ribosylated in HeLa cells, decreasing its ability to bind its consensus
target sequences [70]. YY1 has also been shown to be a negative regulator of p53 activation under conditions of genomic stress in primary and cancer cell lines [71]. This is interesting given that L1 activity is itself thought to be a genomic stressor that induces apoptosis using a p53 dependent mechanism [72]. Under conditions of DNA damage, YY1 could therefore have opposing effects on the retrotransposition frequency: decreased YY1 binding could result in fewer full-length L1 transcripts while YY1’s effects on p53 might enhance the survival of cells that harbor new L1 insertions.

1.6 Altered L1 integration during DNA damage

Insertion of an L1 copy into the genome necessitates the creation and repair of broken DNA. Based on the elegant work from Tom Eickbush’s group on the non-LTR retrotransposon R2Bm and more recent findings using an in vitro L1 system, the L1 endonuclease is believed to nick DNA in a staggered fashion creating overhanging single stranded DNA [73, 74]. After L1 integration, the DNA ends are sealed and filled in, forming the target site duplications that flank a typical L1 insertion (steps 4 and 5, Figure 1-1). On the other hand, if a cell is subjected to DNA damage, the presence of broken DNA may allow L1 to integrate into pre-formed breaks in an endonuclease-independent fashion (Figure 1-2). Alternatively or in addition, enzymes used by the cell to repair damaged DNA may aid (or inhibit) L1 retrotransposition.

Retrotransposons can use artificially induced DNA breaks as substrates for insertion. Yeast deficient in homologous recombination occasionally capture Ty1 cDNA.
Figure 1-2. Potential mechanisms of gamma radiation induced retrotransposition. Left Arrow: Gamma radiation induced double strand breaks could serve as substrates for L1 insertion. Insertions at pre-formed DNA breaks may not require L1 endonuclease; genomic DNA flanking such insertions is expected to lack target site duplications and contain deletions. Right Arrow: Gamma irradiation may make the host environment more amenable to retrotransposition by upregulating cofactors or downregulating repressors for endonuclease-dependent retrotransposition.
during repair of breaks introduced at the MAT locus [75, 76]. Using a plasmid based assay that selected for DNA breaks repaired by captured cDNA, Yu and Gabriel found that 21 out of 37 captured sequences were derived from Ty1 elements [77]. Furthermore, in mouse cells both LTR retrotransposons and SINE elements were able to repair a break induced by the restriction enzyme I-SceI [78].

Collectively, these studies indicate that retrotransposons can integrate into broken DNA. However, these experiments used genetic screens to look for what may have been rare events. Under the conditions of the cell culture L1 retrotransposition assay, mutation of the L1 endonuclease active site reduced the retrotransposition frequency to ~1% of wild type levels [79]. This result suggests that L1 usually uses its own endonuclease to gain entry into the genome. However, in the setting of DNA repair enzyme deficiency (DNA-PKcs or XRCC4 deficiency in particular) L1s lacking endonuclease exhibited greatly increased rates of retrotransposition [33]. L1s lacking endonuclease generated genomic insertions in repair deficient cells with atypical structures (including large deletions at the site of integration), while fully functional L1s generated fewer “atypical” insertions [33, 80-82].

1.7 Cellular co-factors and inhibitors of retrotransposition

Cellular proteins involved in the response to DNA damage, particularly those of the non-homologous end joining cascade (NHEJ), may act as cofactors or inhibitors of retrotransposition. Transcription of NHEJ factors including Ku70 and its partner Ku80
are up-regulated following exposure to gamma radiation [83]. Furthermore, many of these repair factors co-localize at the sites of double strand breaks [84] and have altered bioavailability following DNA damage [85]. Therefore it seems reasonable to propose that modulation and altered subcellular distribution of DNA repair enzymes in the setting of genotoxic stress could influence L1 retrotransposition.

The contribution of DNA repair factors to the mobilization of DNA transposons has been investigated by several groups. In Drosophila, the P element transposase possesses putative phosphorylation sites for the ataxia telangectasia mutation protein (ATM), a master control kinase of the DNA damage response [86]. Mutation of specific ATM sites increased or in some cases decreased excision of these elements. The DNA repair protein Ku70 and the Bloom helicase, both downstream of ATM [87, 88], have been shown to be important for repair of P element excision sites [89]. Ku70 is also important for repair of Sleeping Beauty excision in mammalian cells [40]. In a survey of multiple repair factors, deficiencies in the NHEJ factors Ku80, DNA-PKcs, and XRCC4 and the homologous recombination factors Rad51C and XRCC3 decreased Sleeping Beauty mobility in mammalian cells [41]. Reconstitution of the knockout reversed the phenotype, and even increased transposition above wild type levels for DNA-PKcs [41].

DNA repair factors also influence the mobility of retrotransposons. A mutagenesis screen for inhibitors of Ty1 retrotransposition revealed genes that help maintain genomic integrity including telomerase, a yeast homologue of Bloom, and components of the NBS complex [90]. Rad3 and Ssl2, helicases involved in nucleotide
excision repair, appear to inhibit Ty1 retrotransposition post-translationally [91]. Potential cofactors for Ty1 retrotransposition are the Ku repair factors. Ku70 protein coprecipitates with Ty1 cDNA, cofractionates with Ty1 retrotransposition intermediates, and deficiency in both Ku70 and Ku80 dramatically decreases retrotransposition [92].

There is also evidence linking NHEJ machinery to the regulation of L1s. Ku70/80 binding sites have been identified in murine L1s: L1s make up 19% of the mouse genome, but account for 26% of the Ku70/80 binding sites [93]. Cell lines deficient in DNA-PKcs permit lower rates of endonuclease dependent L1 retrotransposition than their wild type parentals, while XRCC4 mutants permit higher rates of L1 retrotransposition [33]. Repair enzyme deficiency could affect L1 retrotransposition via multiple pathways. Increased persistence of unrepaired double strand breaks could serve as substrates for insertion and increase endonuclease independent insertion. On the other hand, a dearth of DNA repair enzymes might hinder the resolution of L1 insertions. The loss or altered availability of inhibitors could, conversely, promote retrotransposition.

Therefore, exposure of host cells to ongoing DNA damage could affect L1 retrotransposition in a variety of ways. Subsequent changes in insertion frequency and levels of retrotransposition intermediates in damaged cells could identify factors that regulate L1s or assist in their mobilization. The effects of DNA damaging agents on L1 retrotransposition will be discussed in chapter 2.
Germline-only retrotransposition maximizes L1 proliferation while minimizing its impact on the host

L1 retrotransposition is potentially disruptive to the genomes of the organisms that it inhabits. However, L1s have successfully coexisted with their eukaryotic hosts for hundreds of millions of years. Mechanisms that ameliorate or avert harmful L1 insertions include negative selection and selective or reduced L1 expression. Highly active or disruptive L1 elements will decrease the fitness of their host. Mutation or deletion of these elements may allow "less toxic" elements to strike a more favorable balance between successful colonization and injury to the host. There is evidence that deletion of harmful elements occurs frequently, as sex chromosomes, which cannot recombine in the male, carry a heavier burden of L1 elements than the somatic chromosomes [94].

In economics, evolution, and life, negotiations between parties with competing interests can often be resolved through game theory. Game theory assesses outcomes of actions in an attempt to maximize returns [95]. Since the co-evolution of L1 elements with their hosts is a form of extended negotiation, game theory may provide a teleologic rationale for restriction of retrotransposition to germ cells.

From the perspective of the host, retrotransposition is mostly neutral, but is occasionally harmful (Figure 1-3). An insertion that disrupts a tumor suppressor can be lethal, even if it occurs only in a single cell. Rare beneficial insertions only have an effect in germ cells, as a beneficial insertion in a single cell will have little to no effect on survival.
### Host cell

<table>
<thead>
<tr>
<th></th>
<th>Somatic</th>
<th>Germ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive</td>
<td>Apoptosis of single cell or death of organism due to cancer</td>
<td>Decreased fertility and/or decreased fitness of next generation</td>
</tr>
<tr>
<td>Neutral</td>
<td>No significant effect</td>
<td>No significant effect</td>
</tr>
<tr>
<td>Creative</td>
<td>No significant effect</td>
<td>Survival advantage for offspring if fertilized</td>
</tr>
</tbody>
</table>

### L1 element

<table>
<thead>
<tr>
<th></th>
<th>Somatic</th>
<th>Germ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive</td>
<td>Is not passed on, may negatively affect host</td>
<td>May not be passed on, due to death of germ cell or decreased fitness of next generation</td>
</tr>
<tr>
<td>Neutral</td>
<td>Is not passed on</td>
<td>Successful transmission to next generation</td>
</tr>
<tr>
<td>Creative</td>
<td>Is not passed on</td>
<td>Successful transmission to next generation, increased fitness of next generation</td>
</tr>
</tbody>
</table>

Figure 1-3. *Game theory analysis of retrotransposition outcomes in somatic and germ cells.* Disruptive L1 insertions occur infrequently (medium box) and often negatively affect host survival by mutating genes. Neutral L1 insertions occur frequently (large box) and do not affect host. Creative insertions occur rarely (small box) and may generate new proteins by exon shuffling. Outcomes which result in positive selection are shaded in green. Outcomes which result in negative selection are shaded in red. Outcomes with no net selective pressure are not shaded. Overall selective pressures for L1s and their hosts are the weighted sum of all possible outcomes of retrotransposition. In somatic tissues, retrotransposition is not favored for the L1 or the host. In germ cells, L1s face strong selective pressure to retrotranspose at high levels. However, hosts have a selective disincentive to permit high levels of retrotransposition in germ cells.
From the perspective of the element, only L1 insertions that get transmitted to the next generation are favorable outcomes (Figure 1-3). New L1 insertions in somatic tissues can never be of any benefit, because they are an evolutionary dead-end. In contrast, all insertions in the germline are favorable, although those that reduce the fitness of the host organism are marginally less so.

Therefore, mechanisms by which host cells restrict retrotransposition in somatic cells are likely to confer a selective advantage, while ensuring that L1s are under no selective pressure to escape this somatic restriction. In the germline, host cells would be under pressure to permit retrotransposition, albeit at low levels. In contrast, L1 elements would be under pressure to insert as frequently as possible in the germline. This is supported by evidence that the human genome has experienced waves of L1 expansion over evolutionary time [96]. Phylogenetic analysis of recent L1 insertions within the past 25 million years demonstrates limited branching. This suggests that L1s exist in equilibrium with their host, accumulating slowly for millions of years. Occasionally, a single L1 element obtains a mutation or set of mutations that enable it to escape from regulation. That L1 expands clonally until the host evolves new mechanisms of regulation that bring this element under control. Mutational analysis of the coiled-coil domain of ORF1 provides further evidence that L1s have been under selective pressure during these expansions [97].

While somatic cells are clearly not permissive for the expression and retrotransposition of endogenous L1 elements, it is not clear whether new L1 insertions
are regulated in the same way. Studying the regulation of new L1 insertions in the genome might yield clues as to how repression of endogenous L1 elements is initiated and maintained. The expression of new L1 insertions will be discussed in chapter 3.

1.9 Methylation can affect L1 mobility

Regulation of L1 elements in primary somatic tissues is probably governed by multiple redundant mechanisms. In somatic cells, transcriptional control of L1 sequences appears to be an important mechanism of preventing retrotransposition. L1 RNA is abundant in germ cells, but has been detected only in a few primary somatic cell types outside of the testis and ovary, such as placental tissue [98]. As discussed above, transcription factor availability may play a role in determining in which tissues L1s can be expressed. Epigenetic regulation has also been shown to play a role in negatively regulating the transcription of L1 elements [99].

One mechanism of epigenetic regulation, DNA methylation, is important for the regulation of L1 elements. Methylation of cytosines has been negatively correlated with transcriptional activity [100, 101]. Although L1s are AT-rich, the 5' UTR of L1 contains a CpG island [102]. In humans, methylation sensitive restriction mapping suggests that nearly all of the CpGs within this island are indeed methylated [103]. Methylation of the CpG island in L1 is unusual in that CpG islands within genes are typically not methylated [104]. If methylation of L1 elements is necessary and sufficient to transcriptionally repress them, then demethylation should permit L1 re-expression. Many cancer cell lines
are notable for decreased levels of genomic methylation [105]. Demethylation of L1 elements in cancer cells is associated with the expression of endogenous L1 elements, while L1s in surrounding normal tissue remain methylated and transcriptionally silenced [106, 107]. A subset of cancer cell lines also support high levels of L1 retrotransposition in transient transfection assays [11]. In addition, demethylation has been shown to increase L1 RNA levels in mice. Mice deficient in Dnmt3L, a DNA methyltransferase associated enzyme that is active in the testis, have increased levels of L1 RNA in the testis [108].

Methylation of CpGs in L1 may not repress the retrotransposon directly, but may function by recruiting transcriptional repressors to repetitive DNA. Methylation of L1 is associated with recruitment of the Methyl-CpG-binding protein 2 (MeCP2) or the related protein MBD1 to the L1 internal promoter [109]. These factors then act to repress transcription and retrotransposition. In differentiating cells, MeCP2 is associated with the organization of DNA into heterochromatin and the recruitment of histone deacetylases [110, 111]. Both of these functions can repress transcription.

DNA methylation might be a mechanism by which DNA damage affects retrotransposition. Oxidative DNA damage has been shown to decrease the affinity of MeCP2 for damaged methylated DNA [112]. DNA damage near L1 elements might release them from repression. Gamma radiation has also been shown to induce global hypomethylation in cell lines [113] and in mouse livers and spleens [114]. One potential mechanism for hypomethylation in the setting of irradiation is an alteration in the folate
pool. Gamma irradiation has been shown to reduce the activity of the enzyme methylenetetrahydrofolate reductase in the livers of mice [115]. A polymorphism associated with reduced activity of this enzyme has been linked to hypomethylation and cancer susceptibility in humans [116].

Although methylation appears to be important for the regulation of L1s, repression of L1 mobility may be established through multiple pathways. In plants, inhibition of transposons by methylation is leaky, and other mechanisms such as histone methylation and RNAi contribute to their repression [117]. Similarly, primary mammalian cells are not competent for retrotransposition, even when transiently transfected with an L1 element driven by a strong promoter [11]. This suggests that other mechanisms besides transcriptional control govern the repression of L1 retrotransposition. DNA methylation may be necessary but not sufficient to regulate new L1 elements.

The importance of CpG methylation in the regulation of L1 elements is well established. However, it has not yet been determined whether transcription of new L1 insertions is also regulated by DNA methylation. Studying the pattern of methylation in new L1 insertions could yield valuable information as to how endogenous L1s are recognized and targeted for methylation dependent repression. For example, is the CpG island required to initiate methylation? New L1 insertions are often 5’ truncated and lack the CpG island contained within the 5’ UTR. Is DNA methylation triggered by L1 transcription, or by recognition of L1 sequence? What is the minimal length of L1
sequence that initiates methylation? The regulation of new L1 insertions will be discussed in chapter 4.
Chapter 2: Gamma Radiation Increases L1 Retrotransposition

Introduction

L1 elements are potential agents of genomic instability in cells. The generation of a new genomic L1 copy requires the disruption and repair of DNA. L1 proliferation throughout the genome is associated with insertional mutagenesis, chromosomal translocations, and large deletions [80, 82, 118]. I became interested in determining if L1 elements could be activated by genomic instability. DNA lesions created by DNA damaging agents may serve as substrates for L1 insertion via an alternate mechanism of retrotransposition.

It is likely that most L1s enter the genome via a pathway that uses the element-encoded endonuclease to create a nick in the target DNA that produces a free DNA end that can be used as a cDNA primer. This process of target primed reverse transcription is based mainly on elegant studies with the non-LTR retrotransposon R2Bm [73, 119]. The R2Bm element-encoded endonuclease creates sequential staggered DNA breaks at the genomic target site [120]. A similar mechanism likely operates for L1 insertion, albeit at a less specific target site (5'-TTTT/A-3') [74, 80-82, 121, 122]. It is less clear how, after first-strand L1 cDNA synthesis, the second cDNA strand is generated and the insertion site is repaired. Non-homologous recombination machinery may use short stretches of sequence homology between the L1 cDNA and the genomic insertion flank to patch the 5' end of a nascent L1 insertion and/or prime second strand cDNA synthesis [9, 81, 123-125]. Using a bioinformatic analysis, it has recently been reported that the 5' ends of
truncated L1 insertions contain short stretches of microhomology to the target site, whereas longer L1 insertions and Alu elements do not show this pattern [13]. These and other findings (see below) suggest that there is more than one pathway by which L1s can breach genomic DNA.

L1s can also enter the genome in a manner that does not require L1 endonuclease. An endonuclease-deficient tagged L1 element can retrotranspose, albeit with greatly diminished activity compared to the wild type tagged L1 element [79]. Presumably, this endonuclease-deficient L1 inserts into pre-existing DNA breaks. This hypothesis is supported by studies that document that other mobile elements, including retrotransposons, can insert into DNA breaks [76-78]. Endonuclease-deficient L1s can retrotranspose at high frequencies in cells deficient in non-homologous end joining (NHEJ) DNA repair, lending further credence to the idea that L1 insertions can integrate and possibly repair double strand DNA breaks [33, 126]. The use of a transient plasmid-based assay offers unique advantages for determining the mechanism of genotoxic induced L1 mobility. A plasmid-based assay measures retrotransposition directly rather than as a byproduct of L1 activity such as L1 RNA or protein levels. By studying the activity of a single active element, the retrotransposition assay is not confounded by the genomic context or varying activities of different elements.
Results

2.1 L1 elements induce γH2AX foci in human osteosarcoma cells

While the current model of L1 integration posits that L1 endonuclease creates staggered strand cleavages during insertion, it is not known if cells recognize L1 insertions as damaged DNA. To test the ability of the L1 endonuclease to create DNA lesions that are recognized as DSBs in vivo, L1 elements were transfected into human 143B osteosarcoma cells and cells were stained for phosphorylated Histone H2AX foci (γH2AX). H2AX is a histone variant that is rapidly phosphorylated by ATM at or near the site of a double strand DNA break [127]. γH2AX foci appear rapidly, persist for hours, and are proportional to the number of DSBs [127, 128]. Within 24 hours after transfection with an active L1 element, tens of γH2AX foci were observed per cell (Figure 2-1). These findings are similar to observations by Prescott Deininger (P. Deininger, personal communication). The numbers of γH2AX foci were proportional to the activity of the transfected element under the conditions of a retrotransposition assay performed by Shane Horman, a postdoc in the Luning Prak laboratory. The number of breaks is remarkable because it is far in excess of the number of insertions per cell as measured by conventional assays. No increase in γH2AX foci over controls was seen at 48 hours (Figure 2-1), suggesting that negative regulation of L1s occurs shortly after transfection. The return of γH2AX foci to near baseline levels by 48 hours further suggests that L1 activity is negatively regulated by cells beginning shortly after transfection.
Figure 2.1. *Active L1 elements induce formation of γH2AX foci.* 143B cells were transfected with active or inactive L1 elements. At 3, 24, or 48 hours after transfection, cells were permeabilized, fixed, and stained for γH2AX. **A.** Fluorescent images showing γH2AX foci at 24 hours post transfection. EF06R: wild type L1, EF13E: endonuclease deficient L1, EF03N: no L1. All images are 400x. **B.** Average numbers of γH2AX foci per cell at 3 (grey) and 24 hours (black) post transfection. Stuffer/delP: wild type L1 with promoterless reporter cassettes. Percentages below elements represent activity in retrotransposition assays. * p<0.05, ** p<0.01, *** p<0.001 as compared to untransfected by 2 tailed Student’s t-test. n>300 cells per group. **C.** Average numbers of γH2AX foci per cell at 48 hours post transfection. EF05J: retrotransposition-incompetent L1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
To test whether L1 induced DNA damage was specific to the activity of the L1 endonuclease, an L1 containing a point mutation in the endonuclease active site (Figure 2-6A) was also assayed. This element is incapable of nicking DNA [79]. The endonuclease deficient L1 generated a much smaller increase in γH2AX foci at 24 hours, comparable to transfection with a plasmid lacking L1 (Figure 2-1). Thus, the L1 endonuclease generates DNA lesions that are perceived by the cell as DSBs.

2.2 Gamma radiation increases L1 retrotransposition in a cultured cell assay

To determine if DNA damage and the induction of DSBs affects L1 mobility, L1 retrotransposition frequency was measured in a cultured cell assay in the presence and absence of gamma radiation (Figure 2-3). This assay made use of an active human element, L1RP, tagged with an enhanced green fluorescent protein (EGFP) reporter as described previously [129] (Figure 2-2). An L1 insertion of sufficient length into a transcriptionally permissive location in the genome will express EGFP. In human 143B osteosarcoma cells selected for the presence of the L1-EGFP plasmid with puromycin, 2 and 4 Gy of gamma radiation increased the percentage of EGFP+ cells (Figure 2-3). At these doses, DSBs generated were also proportional to radiation exposure, suggesting a correlation between DNA damage and retrotransposition (Figure 2-5). A functionally inactive L1 with two missense mutations in ORF1 [130] showed no detectable retrotransposition with or without irradiation (Figure 2-3), indicating that the radiation-induced increase in EGFP+ cells was specific to the L1 retrotransposition mechanism. Because inhibition of protein synthesis by puromycin may act independently
Figure 2-2. L1-EGFP is a human L1 element that expresses an EGFP reporter following retrotransposition. The L1-EGFP construct used in tissue culture assays is based on the human L1RP element. It's transcription and RNA stability are enhanced by an upstream CMV-IE promoter and a downstream SV40 poly-A signal. Inserted in the antisense orientation in the element's 3'UTR is an EGFP reporter driven by a CMV-IE promoter and containing an HSVtk poly-A signal. This reporter is interrupted by an intron in the sense orientation relative to L1. Consequently, EGFP cannot be expressed from the plasmid because it cannot splice out the intron. Because L1 retrotransposes through an RNA intermediate, the intron will be spliced out before integration. Consequently, EGFP will be able to be expressed from a genomic L1-EGFP insertion of sufficient length in a transcriptionally permissive region of chromatin.
Figure 2-3. Gamma radiation increases the L1 retrotransposition frequency. A. 143B human osteosarcoma cells transfected with L1-EGFP were exposed to gamma radiation and subjected to puromycin selection for the presence of plasmid. The percentage of cells expressing EGFP was measured by flow cytometry. B. Percentage of EGFP+ cells on day 12 from three independent L1-EGFP transient transfections (triangles, circles, squares) following 0, 2, or 4 Gys of gamma irradiation and puromycin selection. * p<0.0005, ** p<0.0001, as compared to 0 Gy by 2 tailed Student's t-test. A functionally inactive L1 was also tested and had undetectable activity at each dose of irradiation (not shown). C. 143B cells were transfected with L1-EGFP and irradiated or mock treated two days later. EGFP was measured on day 8. Cotransfection with DsRed was used to normalize transfection efficiency. Percentage of EGFP+ cells on day 8 from two independent L1-EGFP transient transfections (squares, circles) following 0, 1, 2, or 4 Gys of gamma irradiation. Adjusted for transfection efficiency by DsRed cotransfection. * p<0.05, ** p<0.01, *** p<0.001, as compared to 0 Gy by 2-tailed Student's t-test. D. A dead L1 (EF06R) shows no EGFP expression above background autofluorescence after exposure to gamma radiation.
or in concert with radiation to increase retrotransposition [131], the experiment was also performed without antibiotic selection. Without antibiotic selection, gamma radiation again increased the percentage of EGFP+ cells up to two-and-a-half-fold over mock treated cells (Figure 2-3). The difference in absolute retrotransposition in unselected cells is discussed below.

2.3 Gamma radiation does not increase L1-EGFP RNA levels

A further consideration is that gamma radiation may affect the dynamics of the L1-EGFP expression plasmids used in this study. To test this, I measured the transcriptional activity of the L1 plasmid by quantifying L1-EGFP RNA via quantitative strand specific real-time PCR. L1-EGFP RNA levels were indistinguishable between irradiated and unirradiated cells at 3 and 24 hours post radiation (Figure 2-4). This suggests that plasmid copy number remains constant in the face of gamma radiation. Moreover, it also rules out a direct effect of gamma radiation on L1-EGFP transcription as the cause of the increase in the retrotransposition frequency.

2.4 Other genotoxic agents do not increase L1 retrotransposition

If damaged DNA serves as a preferred substrate for L1 insertion (Figure 1-2), then other DNA damaging agents besides gamma radiation should increase the L1 retrotransposition frequency. Therefore, L1 retrotransposition was monitored in cells subjected to a variety of DNA damage agents. These agents had various mechanisms of action, including oxidative damage, topoisomerase inhibition, and adduct formation.
Figure 2-4. *L1 RNA is not increased in cells exposed to gamma radiation.* A. 143B cells were transfected with L1-EGFP and exposed to 4 Gy of gamma radiation or mock treated 2 days later. Strand specific quantitative RT-PCR was performed on RNA harvested at 3 and 24 hours post irradiation. See materials and methods for primers and conditions. Cycle counts were normalized to GAPDH expression. Data are expressed in log\textsubscript{10} units relative to average L1 expression in unirradiated cells at 3 hours. Black, white, and striped bars represent independent L1-EGFP transfections. RT-PCRs were performed in quadruplicate for each sample. Standard deviations are expressed as error bars.
Only gamma radiation resulted in increased L1 retrotransposition (Table 2-1). DSBs induced by gamma radiation therefore may not cause increased retrotransposition directly, but rather may correlate with other radiation responsive cellular processes that promote L1 mobility. Alternatively, these other forms of genotoxic stress induce fewer DNA breaks at the doses tested.

2.5 Gamma radiation and calicheamicin γ1 generate similar numbers of DSBs

To rule out the possibility that toxicity was reached before a comparable number of DNA breaks were created, the capacity of gamma irradiation and calicheamicin γ1 treatment to induce DNA breaks was analyzed in detail. Calicheamicin γ1 was chosen for this more detailed analysis because of its higher activity in producing DSBs than the other chemotherapeutic agents [132]. To monitor DNA damage, cells were subjected to γ irradiation or calicheamicin γ1 treatment and stained for γH2AX. When γH2AX foci were measured 6 hours after treatment, 5 to 10 pM calicheamicin γ1 elicited a comparable number of foci to 4 Gy of gamma radiation (Figure 2-5). Thus, calicheamicin γ1 produces comparable numbers of DSBs to gamma radiation at the doses used in the retrotransposition assay, yet the retrotransposition frequency in calicheamicin γ1 treated cells decreased. These data suggest that the presence of large numbers of DSBs is not the primary factor increasing the retrotransposition frequency in irradiated cells. Either gamma radiation induced DNA lesions are uniquely preferred substrates for L1 insertion, or gamma radiation increases retrotransposition via an alternative mechanism.
Most genotoxic agents do not increase L1 retrotransposition in a cultured cell assay. 143B cells were transiently transfected with L1-EGFP and exposed to DNA damaging agents at various doses up to and including a toxic dose. The agent, dose range tested, spectrum of single stranded (ss) vs. double strand (ds) break generated, mechanism of damage, and effect on L1-EGFP mobility are summarized. + increase in L1 mobility, - decrease in L1 mobility, N/S no significant change. All retrotransposition assays were performed in 143B osteosarcoma cells.

<table>
<thead>
<tr>
<th>Agent</th>
<th>Dose</th>
<th>Break Type</th>
<th>Mechanism</th>
<th>Retrotransposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma Irradiation</td>
<td>1-10 Gy</td>
<td>ds</td>
<td>Radical Attack</td>
<td>+</td>
</tr>
<tr>
<td>Calicheamicin γ1</td>
<td>5-20 pM</td>
<td>ds</td>
<td>Radical Attack</td>
<td>-</td>
</tr>
<tr>
<td>Etoposide</td>
<td>0.25-100 ug/mL</td>
<td>ds/ss</td>
<td>Topoisomerase II inhibition</td>
<td>-</td>
</tr>
<tr>
<td>Cisplatin</td>
<td>0.25-10 ug/mL</td>
<td>ds/ss</td>
<td>Strand Crosslinking</td>
<td>N/S</td>
</tr>
<tr>
<td>Camptothecin</td>
<td>1-5 nM</td>
<td>ds/ss</td>
<td>Topoisomerase I inhibition</td>
<td>N/S</td>
</tr>
<tr>
<td>Arsenite</td>
<td>100-500 uM</td>
<td>ds/ss</td>
<td>Oxidative</td>
<td>N/S</td>
</tr>
<tr>
<td>UV</td>
<td>25-200 J/m²</td>
<td>ss</td>
<td>TT dimers</td>
<td>N/S</td>
</tr>
</tbody>
</table>

Table 2-1. Most genotoxic agents do not increase L1 retrotransposition in a cultured cell assay. 143B cells were transiently transfected with L1-EGFP and exposed to DNA damaging agents at various doses up to and including a toxic dose. The agent, dose range tested, spectrum of single stranded (ss) vs. double strand (ds) break generated, mechanism of damage, and effect on L1-EGFP mobility are summarized. + increase in L1 mobility, - decrease in L1 mobility, N/S no significant change. All retrotransposition assays were performed in 143B osteosarcoma cells.
Figure 2-5. *Gamma radiation and Calicheamicin γT generate comparable numbers of γH2AX foci.*
143B cells were permeabilized and fixed 6 and 24 hours after exposure to gamma irradiation (54 and 72 hours post transfection). Double strand breaks were detected by staining for γH2AX. Nuclei were visualized by DAPI staining. A. Fluorescent images showing γH2AX foci at 6 and 24 hours post irradiation. All images are 400x. B. Average numbers of repair foci per cell 6 hours after exposure to mock, 4 Gy gamma radiation, or 5 and 10 pM of Calicheamicin γT. The bar graphs show the mean number of γH2AX foci per cell determined from three separate assessments of at least 100 cells in different regions of the coverslip. Error bars indicate standard deviation.
2.6 Endonuclease deficient L1 retrotransposition does not increase with gamma irradiation

If gamma radiation induced DNA breaks were preferred substrates for insertion, then an endonuclease-deficient L1 element should exhibit increased retrotransposition in the presence of radiation-induced DNA damage. A hybrid L1<sub>Rp</sub>/L1.3 element was generated that uses the EGFP reporter and harbors a mutation abrogating endonuclease function (Figure 2-6) [33]. L1<sub>Rp</sub> has greater activity than L1.3 in a retrotransposition assay [129, 133]. There are 10 nucleotide differences causing 3 changed amino acids in the swapped region between L1.3 and L1<sub>Rp</sub> [134, 135]. Because these sequence differences have uncharacterized effects on retrotransposition rates, the parental L1.3 was also swapped for use as an endonuclease-competent control (Figure 2-6).

Retrotransposition activity of these elements was tested in CHO-K1 cells, previously shown to support high levels of endonuclease-independent retrotransposition [33]. Following irradiation, an increased fraction of CHO-K1 cells transfected with the endonuclease-competent L1-EGFP expressed EGFP (Figure 2-6). Gamma radiation is therefore able to increase retrotransposition in an additional mammalian cell line. In contrast, cells transfected with an endonuclease-deficient L1 did not exhibit a statistically significant increase in retrotransposition following irradiation (Figure 2-6). Therefore, nearly all of the increased retrotransposition in the setting of gamma radiation occurs via an endonuclease-dependent pathway.
Figure 2-6. Gamma radiation does not greatly increase endonuclease-independent L1 retrotransposition in CHO-K1 cells. A. An endonuclease-deficient L1 was generated by swapping in a portion of ORF2 from an L1.3 containing a point mutation at the endonuclease active site (see materials and methods). A control endonuclease competent chimeric L1 element was also generated. B. CHO-K1 cells were transfected with endonuclease competent L1-EGFP and retrotransposition measured as shown in fig 1. n = 9 (3 independent transfections measured in triplicate) for each bar. * p<0.001 by 2 tailed Student’s t-test. C. CHO-K1 cells were transfected with endonuclease deficient L1-EGFP and retrotransposition was measured. Total EGFP+ events per gated (live) events: 0Gy: 43/2.3*10^6, 4Gy: 154/2.2*10^6. The difference in retrotransposition frequency between the irradiated and unirradiated endonuclease deficient L1s is not significant. A retrotransposition-incompetent L1 did not exhibit detectable retrotransposition with or without irradiation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
2.7 Genomic flanks of novel insertions have endonuclease-dependent features

To gain further insight into the mechanism of L1 insertion in the setting of genotoxic stress, genomic sequences flanking L1 insertions were cloned. Irradiated cells that expressed EGFP were isolated by flow cytometry and expanded in tissue culture. Suppression PCR and inverse PCR were used to characterize the genomic flanks of new L1 insertions (see materials and methods). Typical endonuclease-dependent insertions have 7-20 base pair target site duplications, an AT rich target site, and a poly-A tail, as evidenced by an insertion cloned from an unirradiated 143B cell (Figure 2-7) [11]. In contrast, endonuclease independent insertions lack these features and may instead contain large deletions, cDNA transduction, or 3' element truncation [33, 82, 122]. Insertion flanks sequenced from irradiated cells had features consistent with endonuclease-dependent retrotransposition (Figure 2-7). All had target site duplications, AT rich target sites, and 3' resided within older mobile elements.

To ensure that the method of recovering L1 insertion flanks could recover endonuclease-independent insertions, an insertion from an endonuclease-deficient L1 was cloned. This insertion lacked target site duplications and had a large deletion at the site of insertion when compared to the corresponding genomic empty site (Figure 2-7). In addition, SINE cDNA capture occurred at the 5' end of this L1 insertion. The co-integration of a SINE sequence along with the L1 sequence has been described previously for endonuclease-independent L1 insertions [78]. These data suggest that, despite large numbers of gamma radiation induced DNA breaks, L1 insertion usually
Figure 2-7. L1 insertions from irradiated clones have endonuclease-dependent features. Single L1-EGFP transfected 143B cells expressing EGFP were isolated by flow cytometry and expanded. DNA was extracted and the 3' genomic flank amplified and sequenced by suppression PCR or inverse PCR (see materials and methods). The 5' flank was then amplified and sequenced using primers designed from the human genome database. Hallmarks of endonuclease dependent L1 insertion include 7-20 bp target site duplications (TSD), AT rich consensus target sites, and poly-A tails. Grey boxes denote TSDs. TSD sequences are displayed beneath each grey box. Numbers represent map positions in L1-EGFP; a full length insertion (including the spliced EGFP cassette) is 7814 bp long. Poly-A tail length is given as the subscripted number next to A/T. Chromosome insertion location is given in the 3' flank. A. L1 insertion flanks from an unirradiated cell. B. Insertion flanks recovered following 4Gy irradiation resemble most endonuclease dependent genomic L1 insertions. C. An endonuclease deficient L1 insertion in a CHO-K1 cell has a deletion at the site of insertion, lacks target site duplications, and has 5' transduced sequence.
occurs via an endonuclease-dependent pathway. Exposure to gamma radiation therefore probably changes the cellular environment in a manner that is more favorable for endonuclease-dependent retrotransposition.

**Discussion**

*Gamma radiation increases L1 retrotransposition*

L1 retrotransposition increases when cultured cells are subjected to gamma radiation. Using different assay conditions, gamma radiation increased L1 retrotransposition up to four-fold. The increase is most evident in the presence of antibiotic selection for the L1-EGFP plasmid. L1 retrotransposition also increased following irradiation without antibiotic selection. The lower retrotransposition frequencies in assays without antibiotic selection are due in part to measurement of EGFP at an earlier time point. The frequency of EGFP+ cells is known to increase for at least the first week of the assay [129]. Antibiotic exposure may also help drive retrotransposition by increasing the copy number of the L1 plasmid or by acting as a stressor in concert with irradiation. The increase in retrotransposition is proportional to the dose of gamma radiation in the 0-4 Gy range. These doses induce many DNA breaks per cell, as measured by staining for γH2AX foci. Increased retrotransposition following irradiation was observed in two different immortalized cell lines, 143B human osteosarcoma cells and CHO-K1 Chinese hamster ovary cells. These cell lines were chosen for this analysis because they were known to support high levels of L1 retrotransposition [6, 33]. Increased retrotransposition following irradiation in both cell
types suggests that the increase in L1 activity is not a unique peculiarity of a single species or cell line.

**Endonuclease-dependent retrotransposition predominates following gamma radiation**

I considered whether gamma radiation promoted integration of L1 into DSBs. To address this possibility, I compared the retrotransposition frequency of an endonuclease deficient L1 in irradiated and unirradiated cells. Despite inducing large numbers of DSBs, as ascertained by staining for γH2AX foci, no significant increase in endonuclease deficient retrotransposition was observed. In addition, Calicheamicin γ1, while creating comparable numbers of H2AX foci to gamma irradiation, did not increase L1 mobility. Therefore, L1 insertion into broken DNA ends appears to occur infrequently in cells.

The analysis of genomic sequences flanking insertions also supports the predominant use of an endonuclease-dependent pathway: All 5 genomic flanks had target site duplications, A/T rich target sites and lacked deletions in the flanking DNA. In addition, some of these insertions had interesting features. A 5' inversion, a feature common to genomic L1 insertions, was identified [124]. This inversion was unusually long, as were many of the insertions sequenced. Long insertions have been associated with highly active elements and may occur naturally in 143B cells. One insertion contained inversions of both the genomic flanking sequence and the L1 sequence. Specifically, 10 bases of the 5' genomic flank were inverted, as well as 7 bases of the L1 proximal to the point of 5' truncation. These are similar to recently described L1
insertions containing palindromic sequences [82]. In addition, two of the insertions were in introns. Insertion of L1 sequences near or in genes has been noted to alter transcription, disrupt coding sequences and interfere with transcriptional elongation [19, 20, 118].

Taken together, these results suggest that L1 usually retrotransposes via an endonuclease-dependent pathway in irradiated cells. At first glance, these results seem inconsistent with earlier studies that document the insertion of L1 elements into pre-formed DNA breaks. However, in these studies, L1 integration events into DNA breaks either conferred a selective advantage or were assayed using methods designed to specifically recover insertions into breaks [33, 76-78]. Interestingly, XRCC4 deficient cells support increased absolute numbers of endonuclease-independent insertions, while DNA-PKcs deficient cells have fewer insertions relative to parental lines [33]. One possible explanation for these findings is that endonuclease-independent insertion is favored at the sites of persistently unrepaired breaks, but requires the assistance of early steps of the DNA repair cascade for targeting to breaks or resolution of integration. In contrast, in the present study, the cells had a largely intact DNA repair apparatus, so breaks were presumably repaired swiftly.

Endonuclease-dependent integration also predominates in published surveys of L1 insertions in the human genome and in tissue culture assays. An in silico survey of 399 human-specific L1 insertions identified only 2 with possible endonuclease-independent features [3]. Because large deletions often accompany endonuclease-independent L1
insertions, it is possible that such insertions were selected against during evolution. However, the analysis of new L1 insertions in cultured cells also reveals a predominance of what appear to be endonuclease-dependent integration events [80-82]. Thus, when faced with multiple DNA breaks, a cell appears to use conserved mechanism of DNA repair, if intact. L1 mediated DNA break repair appears to be infrequent compared to conventional, endonuclease-dependent retrotransposition.

**DNA damage and L1 mobility**

Alternatively, L1 insertion into pre-formed DSBs occurs frequently but is sufficiently toxic that most such cells undergo apoptosis. This is supported by studies showing upregulation of the pro-apoptotic and apoptotic proteins Bax and Caspase-3 via a p53 mediated pathway in breast cancer cells and transformed cells transfected with an active L1 element [72, 136]. A caveat to these findings is that the cell types studied likely had altered DNA damage and apoptotic signalling cascades. However, by this reasoning, it is likely that these studies underestimate the degree of toxicity of L1 activity under physiologic conditions. Indeed, cell death due to L1 activity would not only cause highly active L1 elements to be counter-selected over evolutionary time, but would also bias retrotransposition assays. The large number of γH2AX foci induced following transfection with highly active L1s underscores the threat that mobile elements pose to genomic integrity. Presumably, most of these foci represent sites of DNA cleavage by the L1 endonuclease. The number of foci far exceeds the number of L1 insertions that have been detected, either in single cells or in whole populations.
There are a number of potential explanations for this disparity. First, these foci might represent unsuccessful insertions. L1 endonuclease nicks DNA, but the reverse transcriptase fails to generate L1 cDNA. Second, extremely short insertions might occur frequently. Most retrotransposition assays recover insertions by using a spliceable reporter system. Only insertions at least as long as the reporter can be recovered. If short insertions were more frequent than expected, then total L1 insertions might be underestimated. Third, these foci could represent integration of other RNA templates into the genome. Both SINEs and processed pseudogenes hijack L1 proteins to retrotranspose. SINE and pseudogenes may be mobilized by L1s more frequently than L1s are themselves mobilized. Fourth, cells with multiple L1 insertions might not be recovered because they undergo apoptosis. Transfection with L1 elements is associated with increased cell death. Cells subject to the most L1 endonuclease induced DNA damage might be more likely to undergo apoptosis. These explanations are not mutually exclusive.

Because of the genomic instability that mobile elements induce, organisms have a vested interest in curtailing retrotransposition levels and have likely acquired several strategies for regulating parasitic DNA. I have found that, even in a transient plasmid-based assay, L1-EGFP is subject to considerable negative regulation. No new γH2AX foci above background levels were detected at 48 hours post transfection. Similarly, the amount of L1-EGFP declined significantly from 51 to 72 hours following transfection in both irradiated and unirradiated cells. These results suggest that human cells can interfere with L1 transcription, decrease L1 RNA stability, or have mechanisms for controlling plasmid copy numbers. Cells may also have strategies for preventing L1
translation or inhibiting L1 endonucleolytic activity. Strategies for regulating L1 activity will be explored in further depth in chapters 3 and 4.

*LI mobility in individuals after exposure to DNA damaging agents*

That gamma radiation increases L1 retrotransposition has potential implications for individuals who have been exposed to radiation, particularly in the setting of treatment of neoplasia [137, 138]. First, only a subset of genotoxic agents causes a measurable increase in L1 mobility in the cultured cell assay. In addition to gamma radiation, an increase in L1 retrotransposition has recently been observed in cells exposed to the carcinogens mercury, cadmium, and nickel, but not other heavy metals [139]. In mouse cells exposed to etoposide, a human Alu SINE, thought to utilize L1 retrotransposition machinery, had increased endonuclease-dependent retrotransposition [49]. Although structural differences in the DNA lesions cannot be ruled out as a cause for altered L1 mobility, the cellular response to these different genotoxic agents may play a critical role in modulating L1 retrotransposition. Second, L1 mobility, especially in the setting of genotoxic stress is likely to be damaging to cells. Although studies have documented L1’s ability to insert into pre-existing DNA breaks, the present study suggests that most L1 insertions in cells subjected to genotoxic stress occur via an endonuclease dependent pathway. Like gamma radiation, the L1 endonuclease is potentially highly damaging to genomic integrity and provokes a robust DNA damage response. Third, given the potential toxicity of L1 mobilization, it seems reasonable to speculate that cells with a greater resistance to cell cycle check point arrest and apoptosis
would be more likely to survive genotoxic stress and L1 mobilization. Unfortunately, the
cells most likely to achieve this pernicious state of L1 and DNA damage induced
genomic instability are probably the tumor cells. Radiation therapy may therefore cause
further mutation and L1-mediated genetic instability in a subset of neoplastic cells.
Chapter 3: Expression of new L1 insertions is restricted to the testis

Introduction

One intriguing aspect of L1 biology is the restriction of expression and mobility of L1 elements by tissue type. L1 RNA and protein have been detected primarily in germ tissues and some transformed/cancer cell lines, in particular cancers of the germ tissues [98, 107, 140, 141]. A subset of transformed and cancer cell lines also support retrotransposition when transiently transfected with plasmid-based L1s [5, 130]. Novel L1 insertions from genomic L1s have also been observed occurring in the germline [142, 143]. In contrast, most primary somatic tissues, with a few notable exceptions, do not express L1 RNA or protein [144, 145]. Primary cells also do not support retrotransposition in a transient assay [5] (and Table 4-2). Only a single disease-causing insertion has been documented as occurring in somatic cells [146].

In humans and other multicellular eukaryotic organisms, unrestrained L1 retrotransposition is likely to be deleterious to the survival of the host organism [147-149]. Therefore, organisms that evolved a mechanism to limit L1 activity to a subset of cells gained a survival advantage and were probably selected for. To explore the nature of L1 regulation in vivo, Nina Luning Prak generated a transgenic mouse containing an L1 tagged with an EGFP reporter. This model system offers two distinct advantages over previous studies:
First, transformed and cancer cell lines permit high levels of L1 retrotransposition but have dysregulated gene expression compared to the primary cell types from which they were derived [150-152]. Most tissue culture assays also measure the retrotransposition of a plasmid-based L1 element, which may be subject to different regulation than a chromosomally integrated L1. The L1-EGFP mouse permits the study of a highly active genomically integrated L1 element in primary cells in situ, which is more physiologic. Second, RNA and protein analysis detects not only active L1 elements but also inactive and fossilized L1 elements that have integrated near strong promoters. In contrast, the L1-EGFP mouse model allows study of a single L1 element in relative isolation.

During the characterization of the L1-EGFP mouse, I became interested in the regulation of L1-EGFP insertions. Because L1-EGFP contains a reporter sequence unique within the mouse (and human) genomes, the integration site of insertions can be identified and characterized. In addition, expression of EGFP RNA and protein can be measured more easily than L1 RNA and protein. Analysis of the regulation of new genomic L1 insertions in the mouse may provide insights into the maintenance of regulation of new L1 elements in the human genome.

3.1 A transgenic mouse model of L1 retrotransposition expresses L1-EGFP in the testis

Using the human L1RP element tagged with an EGFP reporter in the 3'UTR, Nina Luning Prak generated two transgenic mouse lines, 57 and 59, (Figure 3-1). This
Figure 3-1. A human L1 is used to generate a transgenic model of L1 retrotransposition. The L1-EGFP transgenic construct is similar to the L1-EGFP used in the tissue culture assay. A. The transgene is derived from the human L1RP element and terminates at a SV40 poly-A signal. The L1-EGFP transgene is driven by a mouse polymerase II promoter in addition to its own 5'UTR. The 3' UTR contains an EGFP reporter in the antisense orientation. The EGFP reporter is interrupted by an intron in the sense orientation. EGFP is driven by a CMV-IE promoter and has an HSVtk poly-A signal. B. A plasmid containing the L1-EGFP transgene was linearized and stably transfected into embryonic stem cells. The embryonic stem cells were injected into a blastomere and used to impregnate a pseudopregnant mouse. Two transgenic lines, 57 and 59, were generated in this fashion. Offspring of mice with the L1-EGFP transgene may contain germline L1-EGFP insertions.
transgenic construct differs from the L1-EGFP used in the cell culture assay system in the use of an upstream promoter to drive L1 expression. In the transgenic mice, L1-EGFP is driven by an upstream murine Polymerase II promoter in addition to its own 5’UTR internal promoter. (The cell culture assay uses a CMV-5’UTR-L1-EGFP.) As in the cell culture assay, EGFP in the transgene is interrupted by an intron, while new insertions derived from the transgene lack the intron. New insertions can therefore be scored genetically for loss of the intron, or phenotypically for expression of EGFP.

To determine what tissue types are permissive for expression of the L1-EGFP transgene, I analyzed organs from transgenic mice for the presence of L1-EGFP RNA. Total RNA was extracted from harvested organs. Using a primer situated in the SV40poly-A signal, sense strand specific L1-EGFP cDNA was reverse transcribed. This cDNA was then amplified in a semi-quantitative PCR using primers flanking the intron (Figure 3-2A). L1-EGFP transcripts were detected at high levels in the testis, weakly in the brain and liver, and not at all in other somatic tissues (Figure 3-2B). Expression of L1-EGFP transcripts in ovaries was inconclusive (data not shown). Contaminating genomic DNA was ruled out as an amplicon source by an RT- control. GAPDH expression by RT-PCR confirmed that the RNA samples could be amplified successfully (Figure 3-2C). Therefore, L1-EGFP, like endogenous L1 elements, is primarily expressed in germ tissues.
Figure 3-2. Strand specific amplification detects L1-EGFP transcripts in the testis. Organs were harvested from an L1-EGFP transgenic mouse and total RNA was extracted. A. L1-EGFP transcripts were reverse transcribed using an antisense primer in the SV40 poly-A signal. cDNA was amplified in a PCR reaction using primers flanking the intron. B. L1-EGFP transcripts are present as a 550 bp product. Control reactions lacking reverse transcriptase failed to detect contaminating genomic DNA. Reverse transcription with random hexamer and amplification of the GAPDH housekeeping gene confirmed that the RNA sample has not degraded.
3.2 *L1 retrotransposition is limited to the testis in a mouse model*

Since the L1-EGFP transgene is only expressed in a subset of tissues in the mouse, L1-EGFP retrotransposition might also be limited to these tissues. I first screened tissues for EGFP expression by fluorescence microscopy. Organs from an L1-EGFP transgenic mouse were observed in the presence of 488 nm light using a Leica M2FL3 stereofluorescent dissecting microscope outfitted with a 100 watt mercury bulb. At 20x-40x magnification, not one of the mouse organs was grossly fluorescent. We considered 3 possibilities: EGFP expression from insertions was too faint to be observed, too infrequent to be observed, or L1-EGFP might not retrotranspose in the mouse model.

To detect faint or infrequent L1-EGFP expressing cells, I generated single cell suspensions from mouse organs and analyzed 500,000 live events by flow cytometry. EGFP expressing cells were found at a rate of approximately 1 in 5000 in the testis (Figure 3-3). No EGFP expressing cells were observed in liver, spleen, or bone marrow. This suggests that L1-EGFP is capable of retrotransposition in the mouse testis. In the other organs tested, L1-EGFP is not active, EGFP is incapable of being expressed from insertions, or the rate of retrotransposition is below 1 in 100,000 cells.

Although L1-EGFP insertions in somatic tissues appear to be incapable of being detected phenotypically, they may be able to be detected genotypically. DNA from L1-EGFP transgenic mouse organs was analyzed by a genotyping PCR with a short
Figure 3-3. Detection of L1-EGFP insertions in the mouse testis by flow cytometry. Testes were harvested from mice with the L1-EGFP transgene, an L1-EGFP insertion, or wild type littermates. The capsule was disrupted and tubules were digested by incubation with collagenase, trypsin, and DNAse. A single cell suspension was generated by syringe dissociation and mesh filtration. Cells were analyzed for EGFP expression. A. Mice with the L1-EGFP transgene have infrequent L1-EGFP positive cells. B. Mice with L1-EGFP insertions express EGFP in a majority of the cells in their testis. C. No EGFP expressing cells are detected in wild type mice. No EGFP+ cells were identified in spleen (0/675,673 events), bone marrow (0/736,523 events) or liver (0/186,552 events).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
extension time to preferentially detect the low molecular weight spliced insertion product. Serial dilution of genomic DNA containing an L1 insertion into a genomic DNA containing the L1 transgene suggests that this method can successfully detect a single insertion in the background of 5000 transgene containing genomes. All tissues tested besides the testis were negative for the presence of insertions (data from NLP). Therefore, L1-EGFP retrotransposition occurs rarely, if at all, in somatic tissues in this mouse model.

3.3 Insertion of L1-EGFP in the germline and during early embryogenesis

The presence of L1-EGFP expressing cells in the testis suggests that germline L1-EGFP insertions may be transmitted to the offspring of L1-EGFP transgenic mice. The L1-EGFP transgene was backcrossed onto 129/SvEmsJ mice, which have a high rate of developing germ cell tumors [153]. As L1 elements are thought to be highly active in germ cell tumors such as teratocarcinomas and choriocarcinomas, L1-EGFP might be most active on this genetic background [141, 154]. To identify offspring with L1-EGFP insertions, genotyping PCRs on tail DNA were performed by Julia Sarino, Sarah Fox, Janet Sallit, and Peffin Lee. Offspring with new insertions were identified by the presence of a low molecular weight product in a genotyping PCR of genomic DNA from tail (Figure 3-4). After breeding these mice for 4 years, 5 insertions have been identified out of 533 total offspring, as summarized in Table 3-1. These insertions are independent and distinct. When bred to a wild type mouse, the insertion segregates independently from the transgene. This suggests that the insertion has integrated on a different
Figure 3-4. **L1-EGFP insertions segregate independently from the L1-EGFP transgene.** Tail DNA was extracted from the offspring of an L1-EGFP tg/LMW+ x wt mating. A. Genomic DNA was amplified in a genotyping PCR using primers that flank the intron. The unspliced transgene is detected as a 1400 bp product, while spliced insertions are detected as 500 bp products. B. Both 500 bp and 1400 bp products can be detected in the offspring from an L1-EGFP transgenic/insertion+ mouse.
chromosome from the transgene, or on the same chromosome at a distant location. Different insertions are distinct from each other in their extent of 5' truncation as detected by "ruler PCR" (Figure 3-5). Some insertions have only a few hundred bases of L1 sequence 5' of the reporter, while others have several thousand bases. Taken together, this suggests that the average rate of unique retrotransposition events in the L1-EGFP mice is approximately 1 per 100 offspring.

L1-EGFP insertions have been identified in offspring of both line 57 and line 59 transgenic mice (Table 3-1). Insertions were also identified after crossing the L1-EGFP transgene onto two other genetic backgrounds. Insertion WB.1 was identified while crossing the line 57 mouse to a Werner/Bloom double knock out background (from Brad Johnson, UPenn). Werner/Bloom deficient mice have increased genomic instability due to progressively shortening telomeres [155]. In yeast, RNAi induced deficiency in mut7, the Werner/Bloom homologue, increased mobilization of transposons [156]. Insertion Bclxl.1 was identified while crossing a line 57 mouse onto a BclXt transgenic background. BclXt mice have enhanced cell survival due to decreased apoptosis [157]. Since L1 retrotransposition has been associated with apoptosis, enhancing cell survival may increase L1-EGFP by allowing more cells with successful insertions to survive [72]. A caveat to this insertion event is that it is not clear if this event occurred on the Bcl-Xt background or was not identified in a previous generation. Given the single event per background, it is difficult to say whether these genetic backgrounds affected the overall rate of L1-EGFP retrotransposition.
Table 3-1. Analysis of L1-EGFP insertions in transgenic mice. Germline L1-EGFP insertions have occurred in both lines 57 and 59 L1-EGFP transgenes. These insertions occurred in multiple genetic backgrounds. The insertions lengths of some of the insertions have been determined by ruler PCR. The insertion sites of some of the insertions have been determined by inverse PCR and suppression PCR. ND: Not done.

<table>
<thead>
<tr>
<th>Insertion</th>
<th>Parental genotypes (MxF)</th>
<th>L1-EGFP line</th>
<th>Insertion length</th>
<th>Insertion site</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.1</td>
<td>embryonic (founder)</td>
<td>57</td>
<td>2010</td>
<td>Ch 4 (63-endonexin/musL1)</td>
</tr>
<tr>
<td>57.2</td>
<td>tg+ x tg+</td>
<td>57</td>
<td>5945-6713</td>
<td>Ch 12</td>
</tr>
<tr>
<td>59.1</td>
<td>tg+ x tg+</td>
<td>59</td>
<td>1886-1968</td>
<td>ND</td>
</tr>
<tr>
<td>Bcl-XI</td>
<td>derivation unclear</td>
<td>-</td>
<td>ND</td>
<td>Ch 17 (musL1)</td>
</tr>
<tr>
<td>WB.1</td>
<td>Wm+/+, Blm +/-, tg+ x Wm-/-, Blm +/-</td>
<td>57</td>
<td>1957-2326</td>
<td>ND</td>
</tr>
</tbody>
</table>

5' primers

| 1700 | 3435 | 4622 | 5485 | 5854 |

3' primer

| 6502 |

Figure 3-5. L1-EGFP insertions are polymorphic in 5' truncation by ruler PCR. A. Tail DNA from L1-EGFP insertion mice was PCR amplified using a 3' primer in EGFP and individual 5' primers at different locations in L1. B. Short L1 insertions will only be amplified by the most proximal primers, whereas long L1 insertion will be amplified by most or all 5' primers.
During generation of these transgenic mouse lines, an L1-EGFP insertion occurred early during embryogenesis of the founder of line 57 (57F₀). Genotyping PCRs of 57F₀ organs were positive for insertion in all tissues tested (Figure 3-6). However, ear fibroblast cultures derived from single cells were heterogeneous for the presence of the insertion [158]. Ear fibroblast subclones were cloned by limiting dilution and found to harbor one of three genotypes: tg-, tg+, or tg+/LMW+. No line was LMW+ only. These data suggest that L1 retrotransposition occurred in a totipotent or pluripotent cell, but after the single cell stage. It is unknown if the insertion occurred at the same time or after the L1-EGFP transgene integrated into the genome.

3.4 L1 insertions express EGFP primarily in testis

Transcription and mobility of the L1-EGFP transgene, despite being driven by a ubiquitous promoter, was primarily confined to the testis. I was interested in determining if expression of EGFP from new insertions was similarly restricted. Organs were harvested from mice with insertions that had been bred away from the L1-EGFP transgene and were screened for EGFP expression by fluorescence microscopy. Fluorescence was strongly present in the testis but absent in all other tissues (Figure 3-7). All five independent L1-EGFP insertions shared a phenotype of testis-only fluorescence. Restriction of EGFP fluorescence to the testis was also evident in the founder mouse in which the insertion occurred during early embryonic development. In the founder mouse, some of the tubules in the testis were fluorescent while others were not, as would be expected from a mouse that is chimeric for an insertion (Figure 3-8).
Figure 3-6. The L1-EGFP insertion is present in multiple tissues in a mouse with an early embryonic L1-EGFP insertion. DNA was harvested from the liver, intestine, brain, and lung of the L1-EGFP line 57 founder mouse and amplified in a genotyping PCR. Spliced product was amplified from all tissues tested.
Figure 3-7. Mice with L1-EGFP insertions express EGFP in cells of the testis. A mouse with a suspected L1-EGFP insertion by genotyping PCR of tail DNA was harvested and organs were screened for fluorescence using a Leica M2FL3 dissecting scope. Left: EGFP fluorescence in the testis of a mouse with an L1-EGFP insertion. Right: The testis of a wild type littermate. Magnification: 20x
Figure 3-8. Tubules are chimeric for EGFP expression in a mouse with an early embryonic L1-EGFP insertion. The line 57 founder mouse had an L1-EGFP insertion during early embryogenesis. Organs from this mouse were screened for fluorescence using a Leica M2FL3 dissecting scope. Some of the cells of this mouse were derived from the embryonic stem cell injected with the transgene. Other cells of this mouse were derived from the blastomere that the stem cell was injected into. The pattern of fluorescence in the 57 founder testis is chimeric as compared to its offspring that have inherited the L1-EGFP insertion.
The lack of fluorescence in somatic tissues does not preclude the transcription of low levels of EGFP RNA. To test for the presence of EGFP RNA, eight organs (thymus, liver, kidneys, spleen, heart, brain, ovary, testis) from an insertion+ mouse were harvested. Total RNA was extracted and reverse transcribed into cDNA using an EGFP specific primer. The cDNA was then amplified in a genotyping PCR. EGFP transcripts were detected at high levels in the testis, but not in any of the other seven organs tested (Figure 3-9). A control reaction lacking reverse transcriptase had no bands, suggesting that successful amplification in the testis sample was not due to the presence of contaminating DNA. A constitutively expressed transcript, Beta-actin, was successfully amplified from all organs, indicating that all samples had an adequate quantity of non-degraded RNA. Taken together, these findings suggest that very few cells are permissive for high levels of expression of L1 insertions outside of the testis.

3.5 L1-EGFP insertions are expressed in a subset of cells in the testis

One interesting observation is that the pattern of EGFP expression within insertion+ mouse testes appears non-uniform. The lumens of some tubules contain fluorescent cells, while others do not. Indeed, fluorescence is heterogeneous along the length of a single tubule (Figure 3-10). One possible interpretation is that EGFP expression within the testis is confined to a subset of cell types or stages. Tubules in the testis contain germ cells at many stages of spermatogenesis as well as supporting somatic cells. To detect EGFP expression within individual cells, testes from insertion+ mice were paraformaldehyde fixed, paraffin sectioned, and stained with an anti-EGFP
EGFP

Figure 3-9. *EGFP* transcripts are detected only in the testis in L1-EGFP insertion mice. A. RNA was extracted from insertion line 59.1 mouse tissues and double-digested with RNase-free DNase. cDNA was reverse transcribed using a primer in EGFP. A 40 cycle PCR reaction amplified the EGFP reporter (~700bp). EGFP transcripts are detected in the testis but not in somatic tissues. No PCR products were generated in the absence of reverse transcriptase. B. Beta-actin RNA was amplified by random hexamer and PCR primers spanning an intron (504bp spliced, 1080bp genomic). All samples have adequate RNA and are free of containing DNA. Performed by Sara Smith.
Figure 3-10. Heterogeneous EGFP expression within the tubules of mice with a germline L1-EGFP insertion. Tubules in the testes of mice with L1-EGFP insertions are not uniformly fluorescent. A single tubule was dissected out from the testis of a mouse with an L1-EGFP insertion. Expression of EGFP is non-uniform along the length of the lumen of this tubule.
antibody. Sectioning and staining were performed by Dan Martinez at the Stokes Pathology Core Facility. George Gerton, Norman Hecht, and Tim Bestor were consulted on the analysis of the stained sections. EGFP immuno-histochemistry is in agreement with the observed pattern of fluorescence. Cells within tubule cross-sections are all positive for EGFP expression, all negative for EGFP expression, or contain a mix of positive and negative cells (Figure 3-11A). Both EGFP expressing and non-expressing cells are present at all stages of germinal cell development, from spermatogonia to spermatids (Figure 3-11B). In addition, Sertoli cells surrounded by EGFP expressing germ cells also appear to express EGFP. It is difficult to determine whether myoid cells express EGFP due to their high levels of background staining. Overall, it is unlikely that EGFP expression is uniquely associated with one or more stages in germ cell development.

Interestingly, when EGFP expressing cells are present in the same tubule as non-expressing cells, the EGFP expressing cells are grouped tightly and do not commingle with non-expressing cells. There is a clear demarcation between EGFP expressing and non-expressing cells in both transverse and radial tubule sections. This pattern is strikingly similar to the progression of spermatogenesis from a single germ cell precursor along a tubule [159]. Daughter cells from different germ cell precursors may occupy the same tubule, but are organized into distinct groupings according to lineage, with more advanced stages of spermatogenesis found closer to the lumen [160]. Daughter cells from a germ cell precursor have also been reported to progress in a helical pattern along a tubule, although this is controversial [161-163]. In one tangential tubular section, EGFP
Figure 3-11. L1-EGFP is expressed at multiple stages of germ cell development. Testes were harvested from a mouse with an L1-EGFP insertion and fixed in 10% formalin. The testis was then embedded in paraffin and sectioned. Sections were stained either with mouse anti-EGFP antibodies and a secondary anti-mouse linked to horseradish peroxidase. The section was developed to stain EGFP expressing cells brown. Sections were counterstained with eosin (blue). Multiple stages of germ, Sertoli cells, and myoid cells stain positive for EGFP expression. Within individual tubules, clusters of cells suspected of being derived from single germ cell precursors all stain positive or negative for EGFP expression. A. 200x. B. 400x.
positive cells appear to spiral inward toward the lumen (Figure 3-12). We hypothesize that during embryogenesis, some germ cell precursors make a cell fate decision to express EGFP, while others do not. This fate decision may be maintained in all of the daughter cells of the germ cell progenitor.

3.6 Expression of L1 insertions is restricted during embryonic development

To determine when during embryonic development L1 insertions begin to be expressed, I looked for EGFP expression in D14.5 mouse embryos. A male mouse with an L1-EGFP insertion was crossed to a wild type female mouse. The female was sacrificed 14 days after a vaginal plug was observed. With the help of Raluca Verona from the Bartolomei laboratory, genital ridges and mesonephrons were dissected out of the embryos. We observed no significant fluorescence in genital ridges at 20-40x using a dissecting scope (Figure 3-13) or at 200x-400x using an inverted scope (data not shown). Genital ridges/mesonephrons were fixed in paraformaldehyde, sectioned in paraffin, and stained with anti-EGFP antibodies by Dan Martinez at the Stokes Pathology Core. Genotyping of fibroblast cultures derived from the embryos confirmed that some were positive for the L1-EGFP insertion (data from JS). None of the male or female genital ridges with L1-EGFP insertions had significant staining above the level of secondary antibody alone (Figure 3-14). There were no differences in morphology or staining between genital ridges that had L1-EGFP insertions and those that did not. Despite the presence of large number of postmigratory primordial germ cells in the genital ridges by day 14.5 [164], it appears unlikely that any express EGFP at that stage.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Helical EGFP expression along a transverse tubule section. Waves of germ cells derived from single germ cell precursors are suspected of progressing along the length of a tubule in a helical pattern. One cluster of cells that uniformly expresses EGFP is oriented in a helix in a tangential section of a single tubule from an L1-EGFP insertion mouse. Magnification: 400x.
Figure 3-13. *Genital ridges from day 14.5 embryos are not fluorescent.* A wild type female was mated to a line 57.1 insertion mouse and sacrificed at day 14.5. Genital ridges (inferior structure) and mesonephrons (superior structure) were harvested and screened for fluorescence under a dissecting scope. No fluorescence above background autofluorescence was observed. Parallel structures in the genital ridge indicate that this is a developing testis. Magnification: 80x.
Figure 3-14. Genital ridges from D14.5 embryos do not express EGFP. A male mouse with an L1-EGFP insertion was mated to a wild type female mouse. The female was sacrificed at day 14.5 post fertilization, and genital ridges/mesonephrons were removed from the embryos. Genital ridges were fixed in 10% formalin and sectioned as above. A. Detection of EGFP expression by staining with mouse anti-EGFP and horseradish peroxidase anti-mouse secondary with eosin counterstain. B. Background staining by secondary antibody alone. Magnification: 400x
of development. The fate decision to re-express L1 insertions must therefore be made during a later stage of development.

Discussion

L1 retrotransposition in the testis and early embryo in a mouse model

The L1-EGFP transgenic mouse model was developed to study the activity of L1 elements in vivo in both somatic and germ cells. This transgenic L1 element actively retrotranspososes in the mouse germline, generating new L1-EGFP insertions at a rate of approximately 1 per 100 offspring. This is similar to the retrotransposition frequency of 1 in 70 obtained by Ostertag et al, using a L1-EGFP with an upstream acrosome promoter [165]. However, it differs from the observed frequency of approximately 1 EGFP expressing cell per 5000 live cells in the testis of L1-EGFP transgenic mice. One interpretation is that L1-EGFP insertions are positively selected during spermatogenesis or fertilization. However, a more likely interpretations is that quantification of EGFP expressing cells by flow cytometry underrepresents the true numbers of L1-EGFP retrotransposition events. In tissue culture experiments, EGFP only begins to be expressed 2 to 3 days after transfection with L1-EGFP, suggesting that there is a lag between retrotranspososition and accumulation of detectible quantities of EGFP protein [129]. Flow cytometry of the testis also inadvertently included cells from many somatic tissues, such as blood vessels and the capsule of the testis, both of which do not express EGFP. Finally, EGFP is a cytoplasmic protein, and may not be present in detectible quantities in sperm, which are numerous but have scant cytoplasm.

69

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In an attempt to induce high levels of retrotransposition in multiple tissues, a murine polymerase II major subunit promoter was added (in addition to the internal promoter in the L1 5'UTR). The polII promoter had been previously reported to induce high levels of expression in a broad distribution of somatic tissues, including liver, lung, intestine, and kidney [165]. Surprisingly, I have determined that L1-EGFP RNA is primarily expressed in the testis in these transgenic mice. This is interesting in that it is similar to published reports of the expression and mobility of endogenous L1 elements which are driven by the 5'UTR only. In adult mice, endogenous L1 RNA transcripts have been identified primarily in germ cells: leptotene and zygotene spermatocytes in the testes, and oocytes in the ovaries [140, 145]. The L1 ORF1 protein is also found in these germ cells in mice [140, 145]. In humans, ORF1 and ORF2 proteins have been detected in spermatids and spermatocytes [98]. There are two likely explanations for the lack of broad tissue distribution of L1-EGFP expression.

The first possibility is that the polymerase II promoter fusion lost function due to accumulated mutations during cloning. This is unlikely in that pPol II/5'UTR transgenic mouse lines have a much higher rate of retrotransposition than two L1-EGFP transgenic mouse lines that lack an additional upstream promoter. All five L1-EGFP insertions arose in the pPol II/5'UTR mouse lines, while the 5'UTR-only lines have no detectable rate of retrotransposition in the germline, or any other tissue. This is similar to findings in other labs and my own observations that an L1-EGFP with a strong, ubiquitous
upstream promoter upstream of the 5′UTR is more active in tissue culture experiments than an L1-EGFP lacking the promoter [165].

The second possibility is that the factors that control the tissue specificity of endogenous L1 elements are predominant in determining the permissiveness of L1-EGFP expression by cell type. These factors could be intrinsic to the L1 element, such as the internal promoter contained within the 5′UTR. Other factors could be extrinsic to the L1 element, such as host cell inhibitors of retrotransposition that limit L1 RNA levels in somatic cells. This is supported by experiments in tissue culture by me and other members of the Luning Prak lab in which the presence of a ubiquitous promoter is not sufficient to permit L1 retrotransposition in cells traditionally non-permissive for L1 retrotransposition, such as ear fibroblasts or embryonic stem cells (data not shown).

In addition to germ tissues, there were two somatic tissues in which low levels of L1-EGFP expression were observed: brain and liver. In the brain, the presence of L1-EGFP might be due to ectopic or illegitimate expression. Regions such as the hippocampus might simply be transcriptionally promiscuous. Alternately, a subset of neurons might actually be permissive for L1 transcription. Recently, an L1-EGFP element was described to be transcribed and capable of retrotransposition in neuronal precursor cells of the rat hippocampus [67]. The report of genomic L1 insertions in rat neuronal precursors remains puzzling in that retrotransposition of L1-EGFP has not been replicated in mouse brains in the Luning Prak lab.
Detection of L1-EGFP transcripts in the liver might have a simpler explanation. Sequencing of the genomic flanks of one of the L1-EGFP transgenes revealed that it had integrated into the cytochrome P\textsubscript{450} gene 7b1 on Chromosome 3. This gene is expressed in the liver [166]. The genomic context of the L1-EGFP transgene may influence its expression. Transcription from the L1-EGFP transgene promoter could be permissive in the liver due to high levels of transcription of flanking DNA. Alternately, the cytochrome P\textsubscript{450} transcript could include L1-EGFP transgene sequence. RT-PCR would detect L1-EGFP RNA, even if it was not present as an independent transcript.

Retrotransposition of L1-EGFP was also primarily observed to occur in the germ tissues. No L1 insertions were detected in any other tissues by flow cytometry or genotyping PCR. This is similar to surveys of L1 activity in humans, in which new insertions have been observed to be transmitted through both the male and female germlines [142, 143]. In addition, the majority of human disease-causing L1 insertions have occurred as the result of germline insertions [118]. There is a single report of an L1 insertion into the adenomatous polyposis coli gene in a somatic tissue causing colon cancer [146]. However, it is not clear whether this L1 insertion occurred after cellular transformation or was the initial mutagenic insult in a primary cell. Identification of an L1-EGFP insertion during early embryogenesis is the first experimental report of an L1 insertion in an embryonic cell \textit{in vivo}. However, this also does not definitively represent a physiologic somatic L1 retrotransposition event, as the insertion may have occurred before or after the L1-EGFP transgene had integrated into the genome.
These observations lend further credence to the notion that regulation of L1s is a complex process and that organisms robustly repress L1 expression in most tissues. Tissue specific expression is established by the host cell or the L1 element itself, and is not easily altered by the addition of an upstream promoter. However, in permissive cells, the presence of a strong upstream promoter can increase L1 retrotransposition, presumably by increasing L1 RNA levels.

*Expression of new L1 insertions is also restricted by tissue type*

Similar to the L1-EGFP transgene and endogenous L1 elements, genomic L1 insertions appear to also be primarily expressed in the testis. Measurement of expression of L1 insertions utilized an EGFP reporter driven by the strong, ubiquitous cytomegalovirus immediate early promoter. This promoter would presumably allow high levels of EGFP transcription in multiple tissues. Surprisingly, fluorescence was only detected in the testis, much like the L1-EGFP transgene and endogenous L1 elements. This suggests that some of the same mechanisms that regulate L1 expression are also acting on the EGFP reporter. These mechanisms must be unusually dynamic to repress such a strong promoter.

One factor that may play a role in regulating L1 elements *in vivo* is post transcriptional gene silencing. The L1-EGFP transgene is an attractive substrate for PTGS because the EGFP reporter is antisense relative to the L1. In the presence of transcripts from the L1 promoter, antisense transcripts may form double stranded RNA.
DsRNAs can be subject to post-translational gene silencing mechanisms such as RNAi. However, L1-EGFP insertions are much less susceptible to RNAi than the L1-EGFP transgene. Most L1-EGFP insertions are extensively 5' truncated and do not contain the L1 promoter [125]. Without the L1 promoter, the sense strand is not transcribed and double stranded RNA does not form. If the L1 transgene is present, it can possibly form double stranded RNA in trans with L1-EGFP insertions. However, all insertions have been bred away from the L1-EGFP transgene for several generations. Finally, the EGFP transcript contains no L1 sequence and is therefore unlikely to form dsRNA in trans with endogenous L1 sequences. Repression of EGFP expression from L1-EGFP insertions in somatic tissues is unlikely to be achieved by PTGS.

RNAi is not the only cellular mechanism that could potentially regulate EGFP expression from L1-EGFP insertions. The L1-EGFP insertion mouse is a potentially powerful model system to study the regulation of L1s by their host organisms. Currently, the major tool for studying L1 elements is the retrotransposition assay, which uses as its read-out the end point of the L1 lifecycle. Interpreting changes in retrotransposition frequency is difficult because the L1 lifecycle is complex and may be regulated by unknown host factors at multiple stages. It is difficult to study retrotransposition intermediates such as RNA levels, because mammalian genomes are cluttered with hundreds of thousands of endogenous L1 elements with a high degree of sequence homology. In contrast, the endpoint of EGFP expression is cellular fluorescence. Detection of EGFP RNA is easier because it is a unique sequence within mammalian
genomes. In chapter 4, EGFP expression from L1-EGFP insertions will be used as a tool to understand the mechanisms that restrict L1 expression.

*L1 insertions are re-expressed after germ cell progenitor migration*

The difference in somatic and germ tissue regulation of L1-EGFP insertions suggests a model for the silencing of new L1 insertions:

Selective germline re-expression: An L1 insertion expressed in the germline is negatively regulated in the early embryo. This negative regulation is maintained in cells destined to form somatic tissues. In germ cell precursors, the L1 is selectively reexpressed. L1 is expressed in the adult testis because germ cells actively reverse negative regulation or lack some of the regulatory mechanisms found in somatic cells (Figure 3-15).

This model is supportive of a retrotransposition event during early embryogenesis. The insertion occurred early in embryonic development, before L1s were shut down in all cells. The cell of insertion was presumably totipotent and was not committed to a somatic lineage. Heterogeneous expression of EGFP in the tubules of mice with L1 insertions also supports and qualifies this model. A subset of germ cells precursors either fails to de-repress the L1-EGFP insertion or actively negatively regulates it. The lack of EGFP expression in D14.5 embryo genital ridges further defines the selective germline reexpression model. L1 insertions remain negatively regulated in
Figure 3- 15. L1 insertions are regulated during development. L1 insertions are expressed in germ tissues but not somatic tissues. To account for this pattern of regulation, L1s must undergo 2 cycles of de novo regulation during development. An L1 insertion is expressed in germ cells but is negatively regulated during embryogenesis. Repression is maintained in somatic cells. During germ cell development, L1 insertions are reexpressed.
premigratory germ cell progenitors, and must therefore be reexpressed some days after colonization of the genital ridges.

The selective germline re-expression model is supported by evidence in the literature. L1 ORF1 protein is first seen to be reexpressed in primitive spermatagonia at day 15.5 post fertilization [145]. Expression of ORF1 protein in ovaries follows a similar timetable [145]. However, no ORF1 protein can be detected in any cells at D14.5 or earlier. In humans, ORF1 and ORF2 proteins can be detected in primordial germ cells at both 18 and 28 weeks post fertilization [98]. The only primary somatic cell types observed to express L1 RNA or proteins are Sertoli, Leydig and myoid cells in the testis, thecal cells in the ovary, and syncytiotrophoblasts in the placenta [98, 140, 145].

The selective germline re-expression model implies that massive changes in the regulation of the human transcriptome occur regularly. This model predicts that the pattern of L1 regulation changes completely within a single generation of the parent organism. L1 expression is permissive in germ tissues but is shut down during embryonic development. Negative regulation is maintained in somatic cells, but is reversed during development of germ tissues. The mechanism of repression of L1 insertions in embryo and somatic cells is highly stringent, as no EGFP expressing cells were detected in somatic tissues. In contrast, the mechanism of derepression in germ cell precursors appears to be error prone, as not all germ cells express EGFP. L1 sequences comprise 17% of the human genome. Even if only a fraction of genomic L1s are subject to this regulation, this suggests that the mechanisms that repress and derepress L1
expression during germ cell development affect expression from hundreds of megabases of DNA.
Chapter 4: Methylation of CpG DNA in Novel L1 Insertions

Introduction

During characterization of a mouse model of retrotransposition, it was observed that expression of EGFP from L1-EGFP insertions was restricted to the testis. I considered three alternatives for the restriction in EGFP insertion. First, the structure or sequence of the EGFP reporter could restrict expression to the testis. Second, L1-EGFP could preferentially insert into chromatin that was only permissive for expression in the testis. Third, L1-EGFP insertions might be regulated by some of the same mechanisms that restrict expression from the L1-EGFP transgene and endogenous L1 elements to germ cells. If the first possibility is ruled out, then L1-EGFP insertions can be studied to gain insights into how organisms regulate the expression of newly integrated L1 copies.

Results

4.1 EGFP is driven by a strong promoter that is active in multiple mouse tissues

A trivial reason for the lack of fluorescence in tissues outside of the testis is that the reporter driving EGFP transcription, the cytomegalovirus-immediate early promoter, might only be active in the testis. However, this conflicts with published evidence that the CMV-IE promoter induces strong, ubiquitous expression in transgenic mouse models [167]. Moreover, the CMV-IE promoter fused to EGFP reporters has successfully expressed EGFP in a variety of mouse tissues in a number of transgenic models (Table 4-
1) [168-179]. Based on a literature survey, CMV-EGFP should be expressed ubiquitously at high levels throughout the mouse.

One possibility is that the CMV promoter accumulated disabling mutations during cloning such that it no longer functions as an effective promoter in somatic tissues. To test this, I amplified the CMV promoter using flanking primers and sequenced it. By CLUSTAL-W comparison, the amplified sequence had 100% identity with the CMV sequence in the Clontech pEGFP-N1 vector from which it was derived. Therefore, it is unlikely that the CMV promoter was inappropriately restricting expression to the testis.

To rule out the remote possibility that the CMV-EGFP fusion used in the L1 transgene cannot be expressed in primary somatic cells, mouse embryonic fibroblasts were transiently transfected with pEGFP-N1 (Clontech), which constitutively expresses EGFP and from which the L1 transgene was derived. Within 24 hours of transfection, the fibroblasts were visibly fluorescent (Figure 4-1). This suggests that somatic cells from the L1-EGFP transgenic mouse line should be capable of expressing CMV promoted EGFP.

4.2 L1 inserts into transcriptionally permissive chromatin

Although EGFP can be expressed from a plasmid in somatic cells, it cannot be expressed from L1-EGFP insertions. With the exception of extremely rare cells,
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Expression system</th>
<th>Citation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>embryonic stem cells</td>
<td>transgenic</td>
<td>Ward and Stem, 2002</td>
<td></td>
</tr>
<tr>
<td>glial cells</td>
<td>lentiviral</td>
<td>Mclver et al, 2005</td>
<td></td>
</tr>
<tr>
<td>heart</td>
<td>transgenic</td>
<td>Fleury et al, 2003</td>
<td></td>
</tr>
<tr>
<td>kidney</td>
<td>transgenic</td>
<td>Provenzano et al, 2005</td>
<td>bicistronic fusion</td>
</tr>
<tr>
<td>kidney</td>
<td>transgenic</td>
<td>Villuendas et al, 2001</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>liver</td>
<td>transgenic</td>
<td>Provenzano et al, 2005</td>
<td>bicistronic fusion</td>
</tr>
<tr>
<td>liver</td>
<td>adenoviral</td>
<td>Yang et al, 2002</td>
<td></td>
</tr>
<tr>
<td>lung</td>
<td>transgenic</td>
<td>Provenzano et al, 2005</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>lung</td>
<td>transgenic</td>
<td>Villuendas et al, 2001</td>
<td>bicistronic fusion</td>
</tr>
<tr>
<td>pancreas</td>
<td>transgenic</td>
<td>Villuendas et al, 2001</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>pups (all tissues)</td>
<td>transgenic</td>
<td>Klysik and Singer, 2005</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>retina</td>
<td>adenoviral</td>
<td>Reich et al, 2003</td>
<td></td>
</tr>
<tr>
<td>skeletal muscle</td>
<td>plasmid uptake</td>
<td>Bartlett et al, 1996</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>skeletal muscle</td>
<td>lentiviral</td>
<td>Ouyang et al, 2004</td>
<td>bicistronic fusion with IRES</td>
</tr>
<tr>
<td>skin</td>
<td>transgenic</td>
<td>Klysik and Singer, 2005</td>
<td></td>
</tr>
<tr>
<td>spleen</td>
<td>transgenic</td>
<td>Tupin et al, 2003</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>spleen</td>
<td>transgenic</td>
<td>Villuendas et al, 2001</td>
<td>bicistronic fusion</td>
</tr>
<tr>
<td>stomach</td>
<td>transgenic</td>
<td>Provenzano et al, 2005</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>tail</td>
<td>transgenic</td>
<td>Klysik and Singer, 2005</td>
<td></td>
</tr>
<tr>
<td>testis</td>
<td>transgenic</td>
<td>Villuendas et al, 2001</td>
<td>similar to L1-EGFP fusion</td>
</tr>
<tr>
<td>thymus</td>
<td>plasmid uptake</td>
<td>Li et al, 1997</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-1. **CMV-IE promoter is expressed in multiple murine tissues.** Many transgenic mouse models utilize an EGFP reporter expressed from a CMV-IE promoter. The promoter-reporter fusion has been reported to be active in multiple somatic tissues in the mouse.
Figure 4-1. Mouse fibroblasts are fluorescent after transfection with an EGFP expression plasmid. Ears harvested from wild type mice were digested with collagenase and cultured. Fibroblasts that grew out of the ear were transfected with EF03N using the FuGENE6 transfection system. This plasmid contains the same EGFP expression cassette that L1-EGFP is derived from. Cells were screened for fluorescence under a Leica inverted microscope and EGFP filter set. A high percentage of fibroblasts express EGFP under these conditions.
fibroblasts derived from embryos with L1-EGFP insertions are not fluorescent (data not shown). It is not clear whether these rare cells express EGFP or exhibit autofluorescence due to a granular deposition of keratin or some other protein. The difference in EGFP expression suggests that genomic L1-EGFP insertions are recognized by somatic cells to different from plasmid-based EGFP in some way that prevents their expression. One possibility could be that L1 insertions are regulated by their genomic context [180].

Based on genomic surveys, L1 insertions have been thought to occur randomly throughout the genome [6, 181-186]. However, tissue culture models of L1 insertion conflict over whether insertion into genes is rare or frequent [28, 82]. Recent experiments with transgenic L1 models suggest that they may actually insert preferentially near genes transcribed at the time of insertion [67]. If L1-EGFP is retrotransposing in the germline, then it may be inserting into genes transcribed in the testis. If these genes are not transcribed outside of the testis, then perhaps L1-EGFP insertions in these genes might also be negatively regulated.

The genomic flanks of L1-EGFP insertions were cloned from genomic DNA by inverse PCR or suppression PCR. Insertion sites were identified by BLASTing flank sequence against the mouse genome. Of these, one particular insertion proved to be of particular interest. Insertion 57.1 had integrated into the beta3-endonexin gene in the anti-sense orientation (Figure 4-2) [158]. According to the Novartis Gene Expression Atlas, transcripts from this gene have been identified not only in the testis, but also in many somatic tissues. This suggests that expression from this L1-EGFP insertion is not limited by its genomic context.
Figure 4-2. *L1-EGFP inserted into a gene transcribed in multiple somatic tissues*. Genomic DNA was extracted from a mouse with an L1-EGFP insertion. Inverse PCR was performed on digested and re-ligated genomic DNA. Amplified bands from the inverse PCR were sequenced and identified by NCBI BLAST against the mouse genome database. L1-EGFP had inserted in the antisense orientation into the Beta3-endonexin gene, initially identified as RNA transcript 4930471016 by the Riken Genome Exploration Project (Accession number AK015557). In the Novartis Mouse Gene Atlas GNF1M, this gene is expressed not only in testis, but many somatic organs as well.
4.3 *EGFP and its CMV-IE promoter are CpG rich*

If the L1-EGFP insertions sites are not primarily responsible for establishing the tissue-specific pattern of EGFP expression, then the L1-EGFP insertion must be specifically targeted for negative regulation in somatic cells. Since EGFP can be expressed from a plasmid, but not from a genomic insertion, this suggests that chromatin at the site of the L1-EGFP insertion is no longer transcriptionally permissive. Therefore, L1-EGFP insertions are epigenetically regulated in somatic tissues.

One epigenetic strategy that cells utilize to downregulate transcription of genes is DNA methylation. DNA methyltransferases add methyl groups to cytidine nucleotides at CpG sequences [187]. Transcription of a gene has been shown to be inversely correlated with the extent of CpG methylation of its promoter [100]. CpG methylation within coding regions is also likely to be important for regulation of expression, but is not as clearly correlated with transcription as promoter methylation [188]. If expression of EGFP from the L1-EGFP insertions is susceptible to DNA methylation, then it should have a high density of CpG sites. Within L1-EGFP, the density of CpG sequences was plotted by Petr Svoboda (Figure 4-3). The L1 5’UTR, the CMV-IE promoter, and the EGFP coding regions all have a high density of CpGs. Thus, both the L1-EGFP transgene and the EGFP reporter are potential substrates for DNA methylation.
Figure 4-3. *L1* 5'UTR, CMV-IE promoter, and EGFP coding regions are all CpG rich. CpG sites were identified in the L1-EGFP sequence. The number of CpG sites per 100 bases is plotted as a function of length. The red bar signifies low CpG densities, while the blue region signifies high CpG densities. The numbers in the boxes indicate the average number of CpGs per 100 bases in each region of the L1. Arrows pointing right indicate L1 sequence that is oriented in the sense direction, while arrows pointing left indicate EGFP sequence that is oriented in the anti-sense direction.
4.4 The L1-EGFP transgene and L1-EGFP insertions are hypomethylated in the testis

If DNA methylation contributes to the silencing of L1-EGFP insertions in somatic tissues, then CpGs in EGFP should be hypermethylated in somatic tissues and hypomethylated in germ tissues. To test the level of CpG methylation, organs from a mouse harboring an L1-EGFP insertion were harvested and genomic DNA was extracted, purified, and bisulfite treated. The entire reporter cassette and some flanking L1 sequence were amplified in four strand specific PCRs and sequenced. Cytosines were converted to uracil during the bisulfite treatment and read as thymine during the sequencing reaction, while methylated cytosines were protected from conversion [189]. Bisulfite sequencing revealed that the CMV promoter is almost completely unmethylated in the testis, but heavily methylated in somatic tissues (Figure 4-4). These results support a role for methylation in regulating the activity of the promoter driving EGFP transcription. Methylation of CpGs in the CMV promoter likely contributes to transcriptional repression of L1 insertions in somatic tissues, and demethylation or lack of methylation of CpGs in the promoter likely contributes to permissive expression of L1s in the germ tissues. In addition, the EGFP coding sequences are also unmethylated in the testis and methylated in somatic tissues (Figure 4-4). While the lack of methylation of the coding regions suggests that the entire insertion is being targeted for epigenetic regulation, Peter Jones has found that methylation of coding regions does correlate with expression as well as methylation of CpG islands in promoters [188].
**Figure 4-4.** *L1-EGFP insertions are relatively unmethylated in testis.* Genomic DNA was extracted from the organs of line 57.1 insertion+ mouse. This DNA was sodium bisulfite treated and amplified in a strand specific PCR reaction in the sense orientation relative to the EGFP cassette. Amplicons were sequenced and methylated cytosines were detected by their conversion to thymines. Each column represents a potentially methylated CpG site in the L1-EGFP transgene sequence. Replicate sequences from each organ are shown in rows. Methylated cytosines are shown as filled circles, while unmethylated cytosines are shown as empty circles.
Organisms may use the same strategy to negatively regulate both new L1 insertions and full length L1 elements in the genome. The L1-EGFP transgene has a similar pattern of expression to L1-EGFP insertions, and might also be subject to DNA methylation. To test this, Petr Svoboda performed bisulfite sequencing on organs from L1-EGFP transgenic mice. The L1 5'UTR, the CMV promoter, and the EGFP coding regions were all unmethylated in the testis and methylated in somatic tissues (Figure 4-5). This suggests methylation is at least partly responsible for establishing and maintaining repression of both the L1-EGFP transgene and L1-EGFP insertions.

4.5 Demethylation increases expression from L1-EGFP insertions in cultured cells

If DNA methylation inhibits L1s, then demethylation might reverse this inhibition. One strategy for reversing DNA methylation is administration of 5'Aza-2'deoxyctydine. 5'Aza-2'deoxyctydine is incorporated into replicating DNA in the place of cytidine, but inhibits methyltransferases [190]. Over several generations, the genomes of cells grown in media supplemented with 5'Aza-2'deoxyctydine will become progressively demethylated.

143B osteosarcoma cells were transfected with either L1-EGFP or the constitutively expressing pEGFP-N1 and selected for the presence of plasmid with puromycin. Cells were then cultured in 0, 10, 100, or 1000 ng/mL of 5'Aza-2'deoxyctydine. Treatment with 1000 ng/mL of 5'Aza-2'deoxyctydine was associated with significant cell death and/or growth delay. After 7 days, cells were harvested and
Figure 4-5. The L1-EGFP transgene is relatively unmethylated in testis. Genomic DNA from the organs of an L1-EGFP transgenic mouse was analyzed by sodium bisulfite sequencing. Each column represents a potentially methylated CpG site in the L1-EGFP transgene sequence. Replicate sequences from each organ are shown in rows. Methylated cytosines are shown as filled circles, while unmethylated cytosines are shown as empty circles.
EGFP was measured by flow cytometry. Treatment with 5’Aza-2’deoxycytidine had little to no effect on the percentage of EGFP expressing cells (Figure 4-6A). Since EGFP is a marker for retrotransposition, this suggests that 5’Aza-2’deoxycytidine did not affect the retrotransposition frequency. However, EGFP+ cells treated with 1000 ng/mL of 5’Aza-2’deoxycytidine exhibited quadruple the mean fluorescence index of untreated EGFP+ cells (Figure 4-6B). By Western blotting, fluorescence intensity correlates well with the amount of EGFP protein [191]. Therefore, the highest dose of 5’Aza-2’deoxycytidine may have increased EGFP expression from L1-EGFP insertions. Constitutive EGFP expression from pEGFP-N1 was unaffected by 5’Aza-2’deoxycytidine treatment (Figure 4-6). This suggests that L1-EGFP insertions, rather than EGFP sequences per se, are targeted for methylation.

To determine whether 5’Aza-2’deoxycytidine affected RNA levels of EGFP, cells from the above experiment were pooled and total RNA was extracted. EGFP RNA was reverse transcribed with a strand specific primer and quantified using real time PCR. In 143B cells transfected with L1-EGFP and treated with 1000 ng/mL of 5’Aza-2’deoxycytidine, EGFP transcripts increased 40-fold over untreated cells (Figure 4-7). Demethylation presumably increased transcription of EGFP from L1-EGFP insertions, increasing fluorescence. 5’Aza-2’deoxycytidine exposure had only minimal effect on RNA levels of L1 transcripts from the L1-EGFP plasmid and EGFP transcripts from the pEGFP-N1 plasmid. Taken together with the fluorescence data, expression from plasmids is relatively insensitive to methylation in this assay.
Figure 4-6. 5'Aza-2'deoxycytidine increases the fluorescence from L1-EGFP insertions, but not their frequency. L1-EGFP or pEGFP-N1 was transfected into 143B cells using FuGENE6. Successful transfectants were selected with puromycin and cultured in 0, 10, 100, or 1000 ng/mL of 5'Aza-2'deoxycytidine. After 7 days of treatment, cells were harvested and EGFP expression in live cells was measured by flow cytometry. A. The percentage of EGFP expressing cells among live cells. n=3 for each group. Error bars represent standard deviations. B. Mean fluorescence intensity of EGFP positive cells. n=3 for each group. Error bars represent standard deviations. * p < 0.005 by student's t-test.
Figure 4-7. 5′Aza-2′deoxycytidine increases EGFP RNA in cells transfected with L1-EGFP. A. 143B cells transfected with L1-EGFP were treated with 0-1000 ng/mL of 5′ Aza-2′deoxycytidine for 7 days. Cells were then pooled and total RNA was extracted (RNeasy kit – QIAGEN). RNA was reverse transcribed using a primer specific for the EGFP reporter transcript (antisense). cDNA was amplified using qRT-PCR. Relative RNA levels were normalized to untreated cells. Each bar represents a single RT-PCR. B. EGFP RNA was quantified from N1-EGFP transfected cells and normalized to untreated cells. C. RNA from L1-EFGFP transfected cells was reverse transcribed using a primer specific for the L1-EGFP transgene transcript (sense). cDNA was amplified using qRT-PCR and normalized to untreated cells.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
4.6 Demethylation does not induce EGFP expression in primary somatic cells under tested conditions

Since demethylation increased EGFP transcripts from recently integrated L1 insertions in a tissue culture assay, demethylation might also be able to permit expression of EGFP from L1 insertions in the mouse model. Embryonic fibroblasts derived from 10 embryos with L1-EGFP insertions harvested at day 14.5 post fertilization. These fibroblasts were cultured with 0, 100, or 1000 ng/mL of 5’Aza-2’deoxycytidine for 14 days. Once again, 1000 ng/mL of 5’Aza-2’deoxycytidine was highly toxic, and very few cells remained alive after 2 weeks of treatment. EGFP expressing cells were present infrequently in all wells after 7 days of exposure, and by FACS analysis, there was no significant increase in the percentage of cells expressing EGFP as a function of 5’Aza-2’deoxycytidine dose. After 14 days, no EGFP expressing cells were detected visually or by flow cytometry (Table 4-2). Dosages of 5’Aza-2’deoxycytidine sufficient to increase expression of EGFP from recent L1-EGFP insertions were not sufficient to induce EGFP expression in primary cells from an L1-EGFP insertion that has been integrated for several generations.

5’Aza-2’deoxycytidine was also tested to determine if it could induce EGFP expression from L1-EGFP insertions in somatic tissues of an adult mouse. 0.15 mg/kg of 5’Aza-2’deoxycytidine was administered by intraperitoneal injection 3 times per week to a 9 week old L1-EGFP insertion mouse. This dose has been shown to induce significant DNA demethylation and histological and functional changes in the murine testis [192]. This dose induced weight loss relative to a littermate injected with PBS (Figure 4-8A).
Figure 4-8. 5'AzA-2'deoxycytidine does not induce EGFP expression in somatic tissues. A 9 week old male mouse with an L1-EGFP insertion and a wild type mouse were administered 0.15 mg/kg 5'AzA-2'deoxycytidine i.p. 3 times per week. A littermate was similarly administered PBS. A. Weight of mice receiving 5'AzA-2'deoxycytidine or PBS. At day 14, the Aza-treated mice were within 1 standard deviation below normal weight, and the PBS-treated mice were within 1 standard deviation above normal weight for the 129/SVJ strain. B. Mice were harvested and single cell suspensions were generated from their testis, spleen, bone marrow, and liver. EGFP expression was measured by flow cytometry. EGFP expression is compared to non-specific autofluorescence.
After 2 weeks of 5’Aza-2’deoxycytidine injections, mice were sacrificed. Only the testis contained fluorescent cells when analyzed at 20x-40x under stereofluorescence. Single cell suspensions were generated from bone marrow, liver, spleen and testis and EGFP was measured by flow cytometry. While the testis contained many EGFP expressing cells, no EGFP expressing cells were found in the bone marrow, liver, or spleen (Figure 4-8B). The course of 5’Aza-2’deoxycytidine treatment may have been too brief to induce demethylation in somatic tissues. Alternately, CpG methylation may be sufficient to shut down expression from new L1 insertions, but demethylation via 5’Aza-2’deoxycytidine exposure may not be sufficient to reactivate L1-EGFP insertions in primary somatic cells.

4.7 Histone acetylation does not induce L1 retrotransposition or EGFP expression from L1 insertions under tested conditions

Another mechanism of epigenetic regulation is histone acetylation. Histone acetylation is directly correlated with transcriptional permissiveness [193]. To test whether histone acetylation could allow EGFP transcription, embryonic fibroblasts containing an L1-EGFP insertion were cultured in Sodium butyrate and Trichostatin A (TSA). These agents increase histone acetylation by inhibiting histone deacetylases [194-196]. To determine if histone acetylation and demethylation had a synergistic effect, embryonic fibroblasts were also treated with these agents in combination with 5’Aza-2’deoxycytidine. Not one of these treatment conditions caused cells to become fluorescent (Table 4-2). In addition, no new insertions were detected in mouse embryonic fibroblasts with the L1-EGFP transgene cultured under these conditions. It
Table 4-2. Combination therapy of DNA demethylating and histone acetylating agents does not permit EGFP expression or retrotransposition in primary cells. Embryonic fibroblasts contained L1-EGFP insertions were cultured in the presence of the DNA demethylating agent 5’Aza-2’deoxycytidine and the histone acetylating agent Sodium Butyrate, either singly or in combination. After 7-14 days of treatment, no EGFP expression was detected visually or by flow cytometry. Embryonic fibroblasts contained L1-EGFP insertions were also cultured with these agents, as well as the acetylating agent Trichostatin A. After 7-14 days of treatment, no retrotransposition events were detected by flow cytometry or by genotyping PCR. High doses or combinations of these agents were associated with significant toxicity (>95% cell death). ND: Not Done
may be possible that histone acetylation coupled with demethylation is sufficient to re-express EGFP or permit retrotransposition in somatic cells. However, the toxicity of these agents, especially when administered in combination, may limit detection of EGFP re-expression. Primary cells appear to be more sensitive to epigenetic manipulation than the 143B osteosarcoma cell line, and a toxic dose may have been achieved before EGFP expression or retrotransposition was detected.

Discussion

L1 inserts into a gene transcribed in testis

The insertion site preference of L1 retrotransposons is an active topic of current mobile element research. Some retrotransposons only insert into highly conserved sequences [197-199]. For example, the R2Bm element, a non-LTR retrotransposon like L1, only inserts into the 28s ribosomal RNA gene [200]. In contrast, the L1 endonuclease recognizes and nicks DNA at the sequence TTTT/A, which occurs frequently throughout genomes [201]. In cultured cell assays, L1s can also been observed to insert into degenerate sequences, albeit at a lower frequency [82].

For many years, chromosome banding analysis experiments in mice and men have generally suggested that L1s insertions are found randomly in the genome, with a slight preference for AT rich regions of the genome [181-185]. Surveys of L1 sequences in the human genome database confirm this distribution [1]. There are two ways in which this pattern could have been established [4]. L1s could have inserted randomly, and the slight
AT predilection is due to the TTTT/A target site. Alternately, L1 inserts nonrandomly but has been subjected to selective forces so as to appear random. Since genes tend to be found in GC rich regions, this model suggests that L1 insertions into or near genes are deleterious and are selected against. Indeed, some deleterious L1s seem to be under selective pressure [94]. If L1s insert non-randomly, newer L1 insertions should have a different pattern of distribution than older elements. Experiments are in conflict over the insertion pattern of recent L1 elements. One study found that recent L1 insertions are found equally in AT and GC rich regions [4]. Another study found that recent L1 insertions have an AT preference, but integrate in non-random clusters [186].

These studies are difficult to interpret because they are all retrospective. Even the most recent L1 insertions have presumably been subjected to generations of selective pressure. Using a cell culture assay of retrotransposition, a prospective study of 100 insertions in HeLa cells observed frequent insertion into genes [82]. Transgenic L1-EGFP models such as the L1-EGFP transgenic mouse are more physiologic systems for the detection of new L1 insertions than cancer cells because they are integrated on chromosomes and are active in the same cell types as endogenous L1 elements. Recently, L1-EGFP retrotransposition was reported to occur in rat neuronal precursor cells in a transgenic model [67]. Surprisingly, many of the 21 insertions were in or near genes expressed in neurons. This suggests that L1 insertions might arise preferentially in regions of chromatin that are actively being transcribed. Chromatin accessibility may contribute to L1 target site selection. Although it is difficult to determine the state of

99
chromatin in a single cell at the moment of L1 integration in vivo, transcriptionally active chromatin is assumed to be "open."

A single insertion into a gene that is actively transcribed in testis has been demonstrated. This is consistent with L1 preferentially inserting into "open" chromatin. A second insertion site which has been cloned was not informative due to its distance from characterized genes. Identification of more genomic flanks of germline L1-EGFP insertions could help determine if insertion in or near testis expressed genes occurs frequently or is atypical. Another way to test this would be determine if the pre-insertion target sites of L1 insertions are expressed in germ cells by Northern blot analysis. Gene expression has been classically associated with open chromatin [202]. However, a caveat to this approach is that the relationship between chromatin structure and transcriptional permissiveness is not absolute [203].

New L1 insertions are methylated in somatic tissues.

Using bisulfite sequencing, we have determined that almost all CpGs in the 5'UTR of the L1-EGFP transgene are methylated in somatic tissues. This is similar to findings that endogenous L1 elements are heavily methylated in primary cells. The 5'UTR of most endogenous L1, like L1-EGFP, contains a CpG island. Restriction digestion of genomic DNA with enzymes sensitive to CpG methylation suggests that most CpGs within L1s are methylated [103]. Indeed, the bulk of methylated cytosines in the genome of differentiated cells is contained within L1 and other mobile elements [99].

100

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
New L1-EGFP insertions in mice are also heavily methylated outside of the testis. Because the tissue specificity of methylation is similar between the L1-EGFP transgene and L1-EGFP insertions, this suggests that the same mechanism targets them for methylation. The presence of mobile element sequence may be the trigger for methylation of the EGFP reporter. DNA methylation of L1 elements was proposed by Tim Bestor to be a component of the nuclear host-defense strategy by which mobile elements, including L1s, are kept in check [99]. This "methylation defense" theory implies that cells have a mechanism for recognizing L1 sequences within the genome [204]. Recognition of L1 sequence would be followed by recruitment of a methyltransferase and methylation of CpGs within L1.

Our findings have several implications for this theory. First, at least one of the L1-EGFP insertions that have been identified is 5' truncated 200 - 500 bases from the EGFP reporter. Only a few hundred bases of L1 sequence are sufficient to recognize and methylate an L1 insertion. Second, in this short L1-EGFP insertion, the L1 sequence is located 5000 bases from the CpG island in the 5' UTR. The L1 sequence recognition step must therefore be able to recognize small fragments of L1 sequence, not just full length insertions that include the CpG island. Third, EGFP has no resemblance to L1 sequence. Therefore, once recruited, the methyltransferase must also methylate adjacent, non-L1 sequences. Methylation has been proposed as one mechanism by which L1s modulate the transcription levels of nearby genes [20]. If the methylation of mobile element sequences induces heterochromatin formation, this heterochromatin may spread to nearby
genes, downregulating them. For example, an Intracisternal A Particle (IAP), a long terminal repeat retrotransposon currently active in mice, is integrated 5' of the agouti gene and has been shown to epigenetically regulate coat color in mice [205].

The EGFP reporter may itself contribute to this process. Published reports of broad tissue expression of EGFP as well as experimental expression of EGFP from plasmids in embryonic fibroblasts suggest that EGFP does not trigger de novo methylation by itself. However, the unusual CpG richness of EGFP may make an L1-EGFP insertion a more attractive site for epigenetic regulation. During human codon optimization of GFP, the number of CpGs was increased from 12 to 60 [206]. Retroviral gene therapy vectors that contain an EGFP in which all CpG sites have been mutated have been generated. These vectors express EGFP less efficiently, but are less prone to methylation dependent inactivation [207]. In a mouse model of transposition, a transgenic Sleeping Beauty transposon with an EGFP reporter gene was much more susceptible to methylation than one with an agouti reporter [208]. While retrotransposons may be epigenetically regulated differently than retroviruses and transposons, it is also possible that the EGFP contributes to methylation dependent inactivation of L1-EGFP insertions.

*Demethylation increases expression of EGFP from L1-EGFP insertions under some conditions*

In 143B cells, transcription of EGFP from L1-EGFP insertions is approximately 40-fold higher when the cells are grown under demethylating conditions. Demethylation
has previously been associated with an increase in transcription and retrotransposition of mobile elements. Many cancers are associated with altered epigenetic regulation [209]. Cellular transformation is often associated with L1 hypomethylation [107, 210, 211]. Demethylation of L1s in cancer cells is associated with increased L1 transcription and is a marker for poor prognosis [107, 212]. Transformation is a complex process that may involve mutation or dysregulation of many genes. One gene that has been identified to be involved in regulating L1 elements through methylation is the methyltransferase associated enzyme Dnmt3L. Mice that lack Dnmt3L have increased levels of L1 and IAP RNA in prospermatogonia, dividing spermatogonia and spermatocytes [108]. In contrast, other methyltransferases, such as Dnmt1, have not been associated with alterations in L1 retrotransposition or retrotransposition frequency. This suggests that regulation of mobile elements is established by a subset of methyltransferases.

Interestingly, demethylation has only been shown to be associated with increased L1 expression in germ cells and cancer cells, cell types that are already known to be permissive for retrotransposition. Accordingly, in mice with L1-EGFP insertions, demethylation is not sufficient to permit EGFP expression from L1-EGFP insertions in mouse embryonic fibroblasts. In addition, injection of a mouse with demethylating agents did not induce EGFP expression in somatic tissues.

There are several possible explanations why demethylation increased EGFP expression in a cultured cells assay but not in primary cells. First, regulation of L1 elements in somatic cells could be a complex process involving multiple layers of
regulation. Regulation of L1 elements in germ cells and some cancer cells may be less stringent, relying on just methylation. Second, methylation might be an early step in the negative regulation of L1 insertions [111]. Prevention of methylation prevents further regulation, such as the incorporation of L1-EGFP insertions into heterochromatin. Third, demethylation might be necessary and sufficient to permit reactivation of L1 elements in all cells. However, L1s are so heavily methylated that a toxic dose of demethylating agents is achieved in primary cells before L1s are reactivated. Transformed and germ cells are more hardy or can tolerate more permissive transcription. These possibilities are not mutually exclusive.

The timing of de novo methylation of L1 sequences in the mouse is informative. In mouse embryos, L1 sequences undergo a cycle of de novo methylation shortly after fertilization [103, 213]. Methylation of L1 sequences is maintained in somatic tissues. However, primordial germ cells begin to undergo a cycle of epigenetic reprogramming beginning prior to their migration into genital ridges [214, 215]. During this time, L1 sequences may become partially demethylated in primordial germ cells [215, 216]. By day 17.5, many L1 sequences are remethylated in male genital ridges [215, 216]. Correspondingly, day 17.5 is also when the Dnmt3L methyltransferase is expressed at high levels in male germ cell progenitors [217].

I speculate that EGFP expression from L1-EGFP insertions is regulated primarily by methylation (Figure 4-9). L1-EGFP is methylated early in embryogenesis, abrogating EGFP expression. This methylation recruits L1 sequences into
Figure 4-9. A model of L1-EGFP insertion regulation by methylation during embryogenesis. After fertilization, new L1-EGFP insertions are methylated and repressed by day 2.5. In somatic cells, this leads to further epigenetic modifications, such as histone deacetylation and heterochromatin formation that reinforces the negative regulation. In migrating male germ cell precursors, L1-EGFP insertions are partially demethylated during genetic reprogramming between days 9.5 and 12.5. By day 17.5 the methyltransferase Dnmt3L is reexpressed in germ cells and methylates L1-EGFP insertions in a subset of germ cell precursors. In adult mice, EGFP is expressed in all germ cells except those arising from the methylated germ cell precursors.
heterochromatin in somatic tissues, a process which may not be easily reversed. During migration of primordial germ cells, L1-EGFP insertions become partially demethylated. After migration, L1-EGFP insertions either remain unmethylated or are remethylated by Dnmt3L in individual germ cell precursors. Adult mice with L1-EGFP insertions reflect this patterning. EGFP expression is limited to daughter cells of germ cell precursors that did not remethylate their L1-EGFP insertions.
Chapter 5: Implications and Future Directions

5.1 Identifying cellular factors that mediate gamma radiation induced retrotransposition

I have observed that exposure of cells to gamma radiation increases L1 retrotransposition. However, I have been unable to establish a biochemical basis for this observation. Identification of factors that can increase the rate of L1 retrotransposition would yield insights into the mechanism of L1 mobility and its regulation. In addition, it would be important in the design of highly active L1 elements for use in gene therapy or random mutagenesis studies. There are a number of clues that may help narrow down the possible mechanisms by which gamma radiation increases retrotransposition. First, the lack of increase of endonuclease-independent retrotransposition suggests that gamma radiation enhances endonuclease-dependent retrotransposition rather than induces an alternate pathway. Second, the fact that L1-EGFP RNA levels were the same in irradiated and unirradiated cells suggests that gamma radiation does not affect L1 transcription and RNA stability. However, this does not rule out a more complex effect, such as increasing L1 transcription while coordinately increasing L1 degradation. Gamma radiation may increase retrotransposition by potentiating a later step in the L1 lifecycle.

Gamma radiation could increase L1 retrotransposition by acting either directly on the L1 (element intrinsic) or by changing the host cell environment (element extrinsic). These alternatives are not mutually exclusive. If gamma radiation is affecting retrotransposition through an element intrinsic pathway, then exposure to gamma
radiation may be able to complement mutations in L1 that cause defects in retrotransposition. I have tested one such mutation that inactivates the L1 endonuclease and found that gamma radiation does not increase retrotransposition under these conditions. The effects of gamma radiation on an L1 deficient in reverse transcriptase could also be tested. Another approach would be to create a library of randomly mutated L1 elements and screen them in a retrotransposition assay to determine if sensitivity to gamma radiation was associated with mutations in particular nucleotides, amino acids, or functional domains of L1. A caveat to screening mutant L1s is that mutations which fully abrogate L1 activity are likely to be uninformative.

If gamma radiation increases retrotransposition through an element extrinsic pathway, the most likely mechanism is by increasing the cellular availability of an unknown cofactor or decreasing the availability of an inhibitor involved in L1 translation, nuclear transport, or integration. To identify this cofactor or inhibitor, one approach will be to use microarrays to look for changes in transcription specific to gamma radiation in a retrotransposition competent cell line. The control condition will be a genotoxic agent that induces double strand breaks similar to gamma radiation but does not increase L1 retrotransposition. One such agent is Calicheamicin γ1, which I have demonstrated generates similar numbers of double strand breaks as gamma radiation, yet does so without increasing retrotransposition. 143B cells will be exposed to gamma radiation, Calicheamicin γ1, or mock treatment. RNA could be extracted, reverse transcribed to cDNA, and analyzed on an Affymetrix microarray. Candidate cofactors would be expressed at higher levels in irradiated cells, but at lower levels or not increased in
Calicheamicin γ1 treated cells. Alternately, candidate inhibitors would be expressed at lower levels in irradiated cells, but at higher levels or not decreased in Calicheamicin γ1 treated cells.

Because the Affymetrix microarrays contain some 22,000 transcripts, it is likely that we will need to use some other criteria to narrow the field of potential genes affecting retrotransposition. Candidate genes would be subject to further analysis based on the following considerations:

1. Magnitude of increase/decrease
2. Known biochemical functions related to the mechanism of retrotransposition, such as translation, nucleic acid trafficking, cell cycle arrest, or double strand break repair
3. Dose dependent increase/decrease between 1 Gy and 4 Gy
4. Homology to \textit{S. cerivisiae} genes already identified as regulators of Tyl retrotransposition [90]

Candidate genes would be tested in a retrotransposition assay for their ability to affect L1 retrotransposition. siRNA against the candidate genes would be compared to non-specific siRNAs to determine if deficiency in the candidate genes increases or decreases L1-EGFP retrotransposition. The effects on retrotransposition of high levels of candidate genes expressed from a plasmid will also be tested.

A caveat to the microarray approach is that gamma radiation affects cells in ways other than transcription. Gamma radiation might affect translation or subcellular compartmentalization of proteins that are important for retrotransposition. For example,
gamma radiation has been shown to release proteins sequestered in the nucleolus [218].
In one study, L1 ORF2 was detected at high concentrations in the nucleolus [219]. A proteomics approach might be merited, as it is an additional strategy to identify factors that interact with or influence L1s.

A general strategy to identify element extrinsic factors would be to develop a random mutagenesis screen of genes that affect L1 retrotransposition in yeast. This approach has been used successfully to identify regulators of Ty1 retrotransposons and Tn7 transposons [90, 220]. First, a yeast system would have to be found and/or engineered such that L1s would be capable of mobilizing. This seems likely given that a hybrid Ty1/L1 element is capable of retrotransposition in yeast [76]. Next, His- Ura- yeast would be mutagenized using a Ty1 retrotransposon with a His reporter [90]. Successful Ty1 integrants would be selected for using media deficient in histidine. Next, an L1 construct will be generated using a Ura reporter. L1-Ura will be transfected into the mutagenized yeast and grown on plates deficient in histidine and uracil. Yeast clones that support high or low rates of L1 retrotransposition as measured by colony formation will be isolated, and the genes interrupted by the Ty1-His element will be cloned and identified. Mammalian homologues of these genes will be identified. Genes that are known to be upregulated or downregulated by gamma radiation or other forms of DNA damage will be of particular interest. These genes will be tested in a retrotransposition assay in 143B cells to determine if they are mediators of gamma radiation induced retrotransposition.
5.2 Artemis deficiency increases L1 retrotransposition by an unknown mechanism

Once a gene or protein is identified, it will be studied to determine its contribution to L1 retrotransposition. One candidate class of genes known to be affected by gamma radiation that may play a role in retrotransposition is DNA double strand break repair factors [52]. These factors have been shown to be important for transposition of the Sleeping Beauty DNA transposon in human cells and retrotransposition of Ty1 in yeast [41, 90]. Deficiency in these factors generally decreases L1 endonuclease-dependent retrotransposition but increases endonuclease-independent retrotransposition [33]. Mechanistically, these proteins could serve as either cofactors or inhibitors of retrotransposition. Resolution of an L1 insertion requires ligation of the insertion with the genomic flanks. This is thought to use elements of the non-homologous end joining pathway, but more recent evidence suggests that homologous recombination may also play a role [13, 82].

To test whether DNA double strand break repair factors affect retrotransposition, I screened DT40 chicken bursal lymphocytes deficient in components of non homologous end joining and homologous recombination (Figure 5-1). Notably, Artemis deficient cells supported three-fold to four-fold higher levels of retrotransposition as compared to parental cells. Increased retrotransposition in Artemis deficient DT40 cells has been confirmed with the assistance of Zach Faber, a rotation student in the laboratory (Figure 5-2). This suggests that Artemis functions as an inhibitor of L1 retrotransposition, and
Figure 5-1. Retrotransposition in chicken lymphocytes deficient in DNA repair enzymes. DT40 chicken bursal lymphocytes deficient in DNA repair enzymes and the wild type parental cell line were co-nucleofected with L1-EGFP and pDsRed-N1 on day 0. Cells were analyzed by flow cytometry for EGFP and DsRed expression on days 1 and 6 post transfection. To control for transfection efficiency, the percentage of cells expressing EGFP on day 6 was normalized to DsRed expression on day 1. n=1 for each DT40 cell line.
Figure 5-2. Artemis deficiency increases EGFP+ cells in chicken lymphocytes. DT40 chicken bursal lymphocytes deficient in Artemis and the wild type parental cell line were co-nucleofected with L1-EGFP and pDsRed-N1 on day 0. Cells were analyzed by flow cytometry for EGFP and DsRed expression on days 1 and 6 post transfection. To control for transfection efficiency, the percentage of cells expressing EGFP on day 6 was normalized to DsRed expression on day 1. n=3 independent transfections for each DT40 cell line. Error bars represent standard deviations. *: p<0.01

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
its absence permits increased L1 mobility. To further demonstrate the importance of Artemis, Zach Faber and other members of the Luning Prak lab are currently reconstituting the Artemis deficient DT40 line with a human Artemis gene from an expression plasmid. If the reconstituted cells have a retrotransposition frequency at or below the parental retrotransposition frequency, this will suggest that the increased retrotransposition was due to the deficiency in Artemis, rather than another mutation that arose during the generation of the knockout cell line. One caveat is that it is difficult to extrapolate observations in DT40 cells to human cells, as they are non-mammalian. DT40 cells are also unusual in that they support high rates of homologous recombination. As discussed above, this may influence the pathway by which L1 insertions are resolved. Experiments are in progress to determine if other cell lines deficient in Artemis also support increased rates of retrotransposition compared to Artemis-sufficient cells.

Artemis is non-homologous end joining repair factor that has multiple functions. Because Artemis opens DNA hairpins, an important intermediate in VDJ recombination, deficiency in Artemis causes a severe combined immunodeficiency [221]. Artemis has also been shown to have 5’ to 3’ exonuclease activity and, after phosphorylation by DNA-PKcs, can act as an endonuclease on both 5’ and 3’ overhangs [222]. RNA is also susceptible to endonucleolytic cleavage by Artemis [222]. Any of these functions of Artemis has the potential to inhibit L1 retrotransposition.

Artemis may inhibit L1 by processing intermediates of retrotransposition. In tissue culture, L1 insertions occasionally have small inversions of either L1 or genomic
sequence in the 5' genomic flank [82]. I have identified one L1 insertion in which both
the L1 and the genomic flank are inverted. One explanation for this feature is that the 5'
end of L1 insertions may sometimes form a hairpin intermediate, which is then opened by
the L1 endonuclease during resolution of the insertion. Premature opening of the hairpin
by Artemis may result in an inversion or an unsuccessful insertion (Figure 5-3). If this
is the case, then an Artemis deficient cell line should have fewer or no L1 insertions with
small inversions in the 5' genomic flank.

Alternately, Artemis could prevent the initial steps of reverse transcription. L1
endonuclease creates an overhang that the reverse transcriptase uses as a primer to initiate
reverse transcription. If the Artemis endonuclease digests this overhang, then L1
insertion would not be able to proceed. Alternately, the presence of Artemis at genomic
DNA nicked by L1 endonuclease may sterically hinder the retrotransposition machinery.
Both of these situations would be difficult to detect, as there would be no L1 sequence at
the aborted insertion site. These sites could be repaired either with no loss of genetic
information, or a small deletion or insertion (Figure 5-3). One strategy to detect this
function is to attempt to coimmunoprecipitate L1 endonuclease and Artemis. Successful
coimmunoprecipitation would suggest that the two proteins are in close proximity.
Another strategy is to immunoprecipitate Artemis and nucleic acids in a ChIP assay.
Subsequent detection of L1 by RT-PCR would suggest that Artemis is in close proximity
to L1 RNA. Alternately, Jef Boeke has developed an in vitro assay for L1 insertion [74].
Endonuclease of overhangs would be detectible in this assay by a gel shift of the target
DNA.

115
Figure 5-3. Artemis inhibits L1 retrotransposition by digesting insertion intermediates. A. At the site of an L1 insertion, L1 cDNA is reverse transcribed from the L1 RNA template. If the 5' genomic flank forms a hairpin intermediate, then Artemis (yellow pacman) can open the hairpin. L1 cDNA recombines with the 5' flank overhang in a region of microhomology, resulting in a 5' truncated L1 insertion with a small inversion of 5' flanking genomic sequence. B. During target primed reverse transcription, nicked genomic DNA forms a 3' overhang. Artemis either cleaves the genomic DNA overhang or sterically hinders reverse transcription at the target site of insertion. Without an effective DNA primer, L1 cDNA cannot be reverse transcribed, resulting in a DNA lesion with no integrated L1 sequence. C. During target primed reverse transcription, the L1 RNA forms a 5' overhang. Artemis cleaves this L1 RNA overhang. Destruction of the L1 RNA template results in an extremely 5' truncated insertion.
Finally, Artemis might digest the L1 RNA intermediate. During reverse transcription, the L1 RNA template would essentially be a large 5' overhang. Endonucleolysis or exonucleolysis of the L1 RNA before reverse transcription is complete would cause 5' truncation of the new insertion (Figure 5-3). The prevalence of 5' truncated L1s has never been adequately explained. Poor processivity of the L1 reverse transcriptase has been ruled out, and a recent theory is that homologous recombination occurs at regions of microhomology in the genomic flank [13, 125]. Artemis activity provides an alternate mechanism for the large numbers of 5' truncated L1 insertions in the genome. If Artemis causes 5' truncation, then Artemis deficient cells should support more long and full length insertions. In this case, Artemis may not necessarily increase the L1 retrotransposition frequency. Rather, more L1 insertions would meet the minimum length required for expression of the reporter (2000 bp).

5.3 Identifying when and where L1 insertions are expressed

The activating effect of gamma radiation is likely to impact retrotransposition frequency only in cells that are already permissive for L1 retrotransposition. One strategy that organisms use to inhibit L1 mobility is to epigenetically restrict L1 expression to germ cells. I have observed that new L1 insertions in the genome are also subject to this regulation. EGFP is highly expressed from new L1-EGFP insertions in a subset of cells in the adult testis but not in adult somatic cells. Understanding the mechanisms by which new L1 insertions are regulated may yield information about the biology of L1 silencing and how the genome defends against parasitic nucleic acids. It
could also lead to the design of L1 gene therapy vectors that permit expression in both somatic and germ cells.

While adult testes with L1-EGFP insertions express EGFP, embryos at day 14.5 post fertilization do not express EGFP, even in germ cell precursors in genital ridges. Therefore, L1 insertions undergo a cycle of de novo repression and re-expression in every generation. The re-expression in germ cells of EGFP from L1-EGFP insertions occurs at a developmental stage between D14.5 and adolescence. Future studies of the expression of new L1 insertions will focus on isolating the stage of germ cell development at which L1 insertions are re-expressed. Embryos will be harvested from matings between males with L1-EGFP insertions and wild type females at days D15.5, D16.5, and D17.5 post fertilization. In addition, neonatal and prepubescent mice will be harvested. Genital ridges and developing testes will be extracted and screened for EGFP expression using fluorescence microscopy and histology. Because EGFP may be transcribed before visually detectible quantities of EGFP protein accumulate, low levels of L1 RNA will be measured by quantitative RT-PCR.

One form of epigenetic regulation that appears to play a role in regulating L1 insertions in somatic tissues is DNA methylation. Both the L1 promoter and the EGFP reporter are CpG rich, and therefore susceptible to repression by methylation. Methylation status is correlated with the expression of EGFP in somatic and germ tissues. It is likely that demethylation or lack of methylation plays a role in the re-expression of L1 insertions in the early embryo. Genital ridges from embryonic mice with L1-EGFP
insertions will be analyzed by bisulfite sequencing at D14.5 and beyond. If methylation plays an important role in stage-specific expression of L1 insertions in developing germ cells, then the developmental stage at which EGFP is expressed should be immediately preceded by demethylation of the CMV promoter and/or the EGFP coding sequence.

5.4 The contribution of L1 and EGFP sequences to expression from L1-EGFP insertions

The most likely explanation for the restriction of expression of new L1-EGFP insertions to the testis is that organisms have a mechanism for identifying and repressing newly integrated L1 sequences. Both very long and very short insertions share this phenotype, suggesting that restriction of L1 sequences to the testis requires no more than a few hundred bases of L1 sequence. However, the amount of L1 sequence present may determine the level of EGFP expression within the testis or the penetrance of EGFP expression within cells of the testis. Longer insertions, which are potentially active and therefore more harmful, might be more tightly regulated than shorter insertions, which are inert. To test this, the testes of age matched mice with long and short L1-EGFP insertions will be harvested. Using histology and flow cytometry, the number of EGFP expressing cells vs. non-expressing cells will be compared. EGFP expression will be compared in these mice by measuring the mean fluorescence index of EGFP expressing cells, as well as by measuring EGFP RNA levels by strand-specific quantitative RT-PCT. A caveat to this approach is that L1 insertion sites differ, and these sites may affect the expression of individual insertions.
There are two strategies to reduce the uncertainty introduced by insertion site variability. First, large numbers of insertions could be screened. Second, L1 insertions of varying lengths could be introduced into the same genomic site in tissue culture. The contribution that L1 sequence length makes to EGFP expression will be measured in 143B osteosarcoma cells. Long and short insertions will be generated by PCR amplification and cloned into a unique site in the genome. The mean fluorescence and frequency of EGFP expression in clones with insertions of different lengths will be measured by flow cytometry. If the L1 sequence contributes to reporter repression in a length dependent fashion, then clones with longer insertions might express less EGFP, or even shut down EGFP expression altogether. One possibility is that, over time, all cells will shut down EGFP expression. If this effect is mediated by methylation, then cells with longer insertions might also be more recalcitrant to re-expression of EGFP by demethylation. Clones will be treated with 5’Aza-2’deoxyctydine and analyzed by flow cytometry. If the length of L1 sequence determines the strength of repression of L1 insertions, then clones with shorter insertions will express EGFP at lower doses of 5’Aza-2’deoxyctydine.

Independent of the effects of insertion length, the EGFP reporter might contribute to its own downregulation. In a transposon model of gene therapy, the CpG rich EGFP reporter was more sensitive to methylation induced repression than a CpG poor agouti gene [208]. Because it contains CpG islands, the EGFP reporter may contribute to the lack of expression from new L1 insertions. To test this, a new L1 construct will be generated using dmEGFP, which has no CpGs. Expression from this construct will be
compared to EGFP in a cell culture assay by introducing full length L1-EGFP and L1-dmEGFP insertions into the same genomic location. Expression from full length L1-dmEGFP insertions will be measured by flow cytometry strand specific quantitative RT-PCR and compared to expression from full length L1-EGFP insertions. If the CpGs in EGFP contribute to the regulation of the expression from new L1-EGFP insertions, then dmEGFP will be expressed at higher levels and in more cells than EGFP. This would suggest that, to obtain maximal expression, the payloads of L1 gene therapy vectors should be optimized to contain fewer CpGs. Alternately, L1 vectors could be designed with extra CpGs so as to be highly effective at silencing genes [223].

5.5 DNA methylation and L1 insertions: Necessary but not sufficient?

Repression of L1 insertions is likely to involve multiple pathways of regulation. Methylation may be an important initial step in the establishment of epigenetic regulation of L1 elements. This is supported by increased EGFP expression from L1-EGFP insertions in cells cultured with 5’Aza-2’deoxycytidine. However, maintaining repression of L1 insertions is likely to involve other, overlapping mechanisms. Concordantly, 5’Aza-2’deoxycytidine treatment failed to induce EGFP expression in vivo and in fibroblast cultures, in which repression of L1-EGFP insertions was already established. This would suggest that methylation is necessary to repress L1 insertions, but demethylation is not sufficient to permit their re-expression.
One possibility is that the dose of 5′Aza-2′deoxycytidine was not sufficient to induce EGFP expression. Indeed, demethylation might be sufficient to re-express repressed L1 insertions, albeit at very high levels. The treatment of mice with 5′Aza-2′deoxycytidine may have been too brief or at too low a dose to demethylate L1 insertions in somatic tissues. This experiment will be repeated using a longer (6 week) exposure, as well as a higher dose of 5′Aza-2′deoxycytidine. In fibroblast culture, a toxic dose of 5′Aza-2′deoxycytidine may have been achieved before the L1-EGFP insertion was sufficiently demethylated to permit EGFP expression. The survival of cells can be enhanced by overexpressing anti-apoptotic proteins, such as Bc1-X\textsubscript{L} [157]. New fibroblasts with L1-EGFP insertions will be derived after crossing to a mouse transgenic for Bc1-X\textsubscript{L} overexpression or existing fibroblasts will be transfected with a Bc1-X\textsubscript{L} expression plasmid. These fibroblasts will be treated with high doses of 5′Aza-2′deoxycytidine and analyzed by flow cytometry. If demethylation alone can regulate L1 insertions, then L1-EGFP insertion mice and fibroblasts expressing Bc1-X\textsubscript{L} should express EGFP at high doses of 5′Aza-2′deoxycytidine.

A more likely possibility is that methylation triggers other mechanisms which are the true effectors of transcriptional repression of L1 insertions. These mechanisms are not easily reversed once established. One of these downstream mechanisms is histone deacetylation. Acetylation of histones H3 and H4 is associated with transcriptional permissiveness, and deacetylation is associated with transcriptional repression [224]. To test whether histone deacetylation plays a role in repressing L1 insertions, mouse embryonic fibroblasts were cultured with both methyltransferase inhibitors and histone
deacetylase inhibitors. While no EGFP expression was observed, this does not rule out regulation of L1 insertions by histone deacetylation. These agents, singly or in combination, were highly toxic. As discussed, fibroblasts expressing Bcl-X\textsubscript{L} might tolerate a higher dose of these agents. These fibroblasts will be cultured in the presence of 5'Aza-2'deoxycytidine, Trichostatin A, and/or Sodium Butyrate, and analyzed by flow cytometry. If both methylation and histone acetylation are redundant mechanisms of L1 insertion regulation, EGFP should be re-expressed under these conditions.

Another strategy to explore the regulation of L1 insertions would be to uncouple DNA methylation and histone deacetylation. One of the mediators of histone deacetlylation is Methyl-CpG-binding protein 2 (MeCP2). MeCP2 binds to methylated CpGs and is thought to recruit histone deacetylases [111]. MeCP2 has been shown to be recruited to methylated DNA in the L1 5'UTR [109]. To test whether MeCP2 is required for repression of L1 insertions, cells lines will be obtained that have been derived from patients deficient in MeCP2, a condition known as Rett Syndrome, as well as reconstituted cells with transgenic MeCP2 [225]. Cells will be transiently transfected with L1-EGFP, and expression of EGFP measured in a retrotransposition assay. If cells with L1-EGFP insertions have a higher mean fluorescence index in a MeCP2 background, this would suggest that MeCP2 is required to negatively regulate methylated L1 insertions. Another approach would be to use cells derived from a mouse in which the MeCP2 is flanked by loxP sites [226]. If MeCP2 catalyzes the repression of L1 insertions, then the introduction of Cre from an exogenous source might induce the re-expression of EGFP. MeCP2 deficiency may be more likely to cause EGFP re-
expression than inhibition of histone deacetylation because MeCP2 is also thought to recruit proteins that induce the formation of heterochromatin [110]. Heterochromatin formation and histone deacetylation may play synergistic roles in repressing L1 insertions. Alternately, repression of L1 insertions may still occur in the absence of MeCP2 because homologues such as MBD1 could provide overlapping functions or MeCP2 might be unable to convert established heterochromatin to euchromatin [109].

5.6 L1s and genomic stress

I have observed increased L1 retrotransposition following exposure to gamma radiation. Conversely there appears to be stringent regulation of the expression of L1 insertions in somatic tissues. These findings are consistent with the idea that high levels of L1 retrotransposition cause genomic instability and are detrimental to cell survival. Organisms normally keep L1 proliferation in check by redundant mechanisms. It is a testament to the stringency of L1 regulation that retrotransposition only seems to occur under extremely toxic conditions. Gamma radiation induces elevated levels of retrotransposition in cancer cells only at near lethal levels. Whole genome demethylation and histone acetylation do not permit re-expression of epigenetically regulated L1 insertions, even at doses that induce cell death. Unrestrained proliferation of mobile elements appears to be a fate worse than death.
Chapter 6: Materials and Methods

Recombinant DNA plasmids

All plasmids consist of the L1RP element tagged with an EGFP cassette (pL1RP-EGFP) in a pCEP4 backbone (Invitrogen) as described previously [129]. The L1 element is driven by its 5' UTR and an upstream CMV MIE promoter. L1-EGFP (EF06R) was derived from L1RP-EGFP by blunt ligation of nt 1-1392 of pPur (BD/Clontech) into the Nru I site of pCEP4. A negative control “dead” L1-EGFP (EF05J) was similarly derived from pL1RP(JM111)-EGFP, which contains disabling mutations in ORF1 [130]. A positive control with constitutive EGFP expression (EF03N) was derived from pEGFP-N1 (BD/Clontech) by cloning pPur into the EcoO191 site of pEGFP-N1. An EN-plasmid (EF13E) was created by swapping nt 1927-3708 (Age I-Bcl I) from JM102D205A [33], which contains a point mutation in the EN domain, into EF06R. A control plasmid (EF12J) was similarly derived from JM102 [130]. JM102 and JM102D205A are derived from LI, which differs from L1RP by 10 base changes in the region swapped.

Cell Lines

143B human osteosarcoma cells were a gift from H. Kazazian (University of Pennsylvania). Both 143B and CHO-K1 cells were cultured at 37°C, 5% CO2, and 100% humidity in DMEM (Gibco BRL) supplemented with 10% Fetal Calf Serum (US...
Biotechnologies) and 100 U/mL of penicillin and streptomycin. DT40 cells deficient in Ku70, RAD54, Ku70/RAD54, DNA-PKcs, and ATM/Ku70 were a gift from S. Takeda (Kyoto University). Parental DT40 cells and DT40 cells deficient in Artemis were a gift from M. Takata (Kawasaki Medical School). DT40 cells were cultured at 37°C, 5% CO\(_2\), and 100% humidity in RPMI1640 (Gibco BRL) supplemented with 10% Fetal Calf Serum (US Biotechnologies), 1% chicken serum (Sigma), 2mM glutamine (Gibco BRL), 50uM B-mercaptoethanol (Sigma) and 100 U/mL of penicillin and streptomycin.

**Gamma Irradiation**

On day 0, 1x10\(^5\) adherent 143B cells were transfected with 1 μg of L1 plasmid and 0.2 μg of DsRed-Express-N1 (Clontech) in 6 μL FuGENE6 (Roche) and 100μL OptiMEM (Gibco BRL). On day 2, cells were harvested with Versene (2% EDTA in PBS) and irradiated with 0-4 Gy using a \(^{137}\)Cesium source. Cells were returned to culture, harvested on day 8, and the percentage of EGFP+ cells analyzed by flow cytometry (BD FACSCaliber), gating on live cells by forward/side scatter and ToPro3 exclusion. Transfection efficiency was normalized using pDsRedExpress-N1 (Clontech). Alternately, cells were selected for transfection with 0.6 μg/mL of puromycin on days 4-6 and analyzed by flow cytometry on day 12.

**Other chemotherapeutic agents**

143B cells were transfected as above. Cells were exposed to genotoxic agents for 1 hour on day 2 in standard culture conditions, washed twice with PBS, and returned to
culture. Agents tested were Arsenite (Sigma, in PBS), Calicheamicin γ1 (gift from Wyeth, in DMSO), Camptothecin (Sigma, in DMSO), and Cisplatin (Sigma, in PBS). Etoposide (Sigma, in DMSO) exposure was for 10 minutes. UV irradiation was performed in a Stratalinker (Stratagene) with cells in 3 mm depth of PBS.

**DNA damage quantification**

Cells grown on coverslips were fixed and permeabilized with 3.5% paraformaldehyde + 1% TritonX-100, washed with PBS, and blocked with KB (0.01 M Tris pH 7.5, 0.15 M NaCl, 0.1% BSA, 0.1% Na Azide). Coverslips were stained with mouse anti-phosphorylated H2AX (Upstate, Charlottesville) and washed first 1x with 0.1% Triton in PBS and then 2x with KB. Secondary staining was with anti-mouse AlexaFluor 594 (Molecular Probes, Eugene), and nuclei were visualized by DAPI staining (Sigma). Stained cells were examined with a 100X PlanNeofluor objective mounted on a Nikon TE-200 microscope. For assessments of H2AX foci, the number of foci detectable in each nucleus was counted by focusing through the entire thickness of the nucleus and recording the number of foci from all levels. At least 300 cells were counted per time point/treatment. Images were captured with a Hamamatsu CCD camera.

**Quantitative RT-PCR**

143B cells stored in RNAlater (Ambion) were homogenized on QIAshredder spin column in 600 μL Buffer RLT (QIAGEN) and RNA was extracted using the RNeasy
protocol for Isolation of Total RNA from Animal Cells with on Column DNase digestion (QIAGEN). 100 ng of total RNA was reverse transcribed using Taqman Multiscribe (ABI). Primers used were in L1 3’UTR (SV40rev-TCCAAACTCATCAATGTATCTTATCAT), in EGFP (geno5- TTTATTGCGGATCCCTCAGAAGAA) and a random hexamer for GAPDH (ABI). Quantitative real-time PCR was performed using High Capacity cDNA archive kit (ABI) and detected on a PRISMR 7900HT (ABI) using a custom designed TaqmanL1 probe/primer (ABI). GAPDH was similarly quantified using random hexamer reverse transcription and human GAPDH Taqman probe/primers (ABI). Data were analyzed using SDS 2.2.1 (ABI) and Excel.

Cloning and sequencing of genomic flanks

EGFP+ clones were handpicked or isolated by flow cytometry (FACSDiva). Clones were expanded in culture for 2 weeks, and DNA purified using a DNeasy kit (Qiagen). Amplification was either by inverse PCR as previously described [158] or by modified suppression PCR [142]. For suppression PCR, genomic DNA was digested by Ase1, EcoR1, or Hind III (New England Biolabs). A linker/dummy combination was annealed and ligated (New England Biolabs) to the digested genomic DNA: ATlink-GTGGCGGCGCCAGTGATTCCTGGAGGGCGCGGCGATAGATAGAACG. Excess linkers were removed with Centricon-40 columns (Princeton Separations). Genomic flanks were amplified using a primer in the linker ATX-GTGGCGGCGCCAGTGATTC and seminested primers in L1-EGFP: SV40for1-ATGATAAGATACATTGATGAGTGTGGA and
SV40for2-GGACAAACCACAACCTAGAATGC. The primary PCR was done in a reaction volume of 25\u00b5L containing 1 \mu M SV40for1, 50nM ATX, 1x FailSafe buffer D, 4% DMSO, 1.25U FailSafe Taq (Epicentre) and 8 ul linked DNA. Amplification used the following cycling conditions: 68° for 5 s (extension of dummy), 94° for 4 min (primary denaturation), followed by 40 cycles at 94° for 30 s, 85° for 10 s, 75° for 10 s, 63° for 30 s, and 68° for 120 s, with a final extension at 68° for 10 min. The secondary PCR was done in a reaction volume of 20\mu L containing 1 \mu M SV40for2, 50nM ATX, 50\mu M dNTPs, 1x PCR buffer D, 4% DMSO, 1.25U FailSafe Taq and 1\mu l primary PCR product. Amplification used the following cycling conditions: 94° for 4 min, followed by 40 cycles at 94° for 30 s, 63° for 30 s, and 68° for 120 s, with a final extension at 68° for 10 min. PCR products were run on a 0.8% agarose gel and extracted by QIAQuick (Qiagen), TA cloned into pCR2.1 (Invitrogen), and transformed into One-Shot bacteria (Invitrogen). Plasmid DNA was extracted from expanded clones by QiaPrep (Qiagen) and sequenced. 3' flank DNA was localized in the human genome using MegaBLAST against the NCBI database and 5' primers were generated from upstream sequence (IDT DNA). The 5' flank was amplified using insertion specific primers as well as primers in L1 (L1ORFIAS: GGTTGTTCCCTTCCATGTTCAGC, L1RP3435AS: GCTGTGAATCCATCTGGTCC, and HSVtk: CCGATTGCAGCGCATCGCCTT). The PCR was performed in a reaction volume of 20\mu L containing 300nM of each primer, 50\mu M dNTPs, 1x PCR buffer I (Roche), 1.5U TaqGold (Roche) and 200ng genomic DNA and used the following cycling conditions: 95° for 10 min, followed by 40 cycles at 95° for 25 s, 58° for 30 s, and 72° for 210 s, with a final extension at 72° for 10 min. PCR products were isolated and sequenced as above.
To detect L1-EGFP transgene transcripts, RNA was extracted from mouse tissues using an RNAeasy midi kit (Qiagen, Chatsworth, CA) and digested with RNase-free DNAse1 (Roche). cDNA synthesis was performed with AMV reverse transcriptase following the manufacturer's instructions (Roche). The SV40rev oligonucleotide was used to produce L1 sense-strand cDNA (from the transgene). Each 50-μl PCR contained 5 μl of 10× PCR buffer (Roche buffer 1 with 15 mM MgCl2), 1 μl of 10 mM dNTPs, 200 ng of each oligonucleotide, 0.3 μl of AmpliTaq Gold, and 5 μl of reverse transcriptase extension product. Amplifications were performed on a Peltier thermal cycler (Hybaid, Asford Middlesex, U.K.) by using the following program: 95°C for 15 min, followed by 40 cycles of 95°C for 25 sec, 62°C for 30 sec, and 72°C for 90 sec. This was followed by a final extension at 72°C for 10 min and a 4°C hold. The following tissues were surveyed for L1 sense-strand transcripts in L1-EGFP line 57 transgenic mice: testis, pooled ovaries, spleen, liver, and brain. Testis, pooled ovary, liver, and lung were surveyed in L1-EGFP line 59 transgenics. Testis and lung were surveyed in line 63 L1-EGFP transgenics.

To detect EGFP transcripts from insertions, RNA was extracted from mouse tissues (thymus, liver, kidney, spleen, brain, heart, ovary, testis) using an RNAeasy mini kit (Qiagen) and digested with RNase-free DNAse1 (Qiagen) followed by an additional DNAse treatment (DNA-free Kit, Ambion). cDNA synthesis was performed following the manufacturer’s instructions (Taqman, ABI). The geno5 oligonucleotide was used to
produce EGFP cDNA and a random hexamer primer (Roche) was used to produce beta-actin cDNA. Each 25-μl PCR contained 12.5 μl of 2× PCR buffer (Epicentre Premix D), 100 ng of each primer (see below), 0.25 μl of Failsafe PCR Enzyme Mix (Epicentre), and 5 μl of reverse transcriptase extension product. Amplifications were performed using the following program: 95°C for 15 min, followed by 40 cycles of 95°C for 25 sec, 62°C for 30 sec, and 72°C for 90 sec. This was followed by a final extension at 72°C for 10 min and a 4°C hold. The expected product size for the EGFP reaction is ~700 bp. The expected product size for beta actin is 504 bp while a PCR product from genomic DNA would be 1080 bp.

Primers:

EGFP cDNA detection: geno5 – TTTATTGCCGATCCCTCAGAAGAA; geno3 – TTCAAGGACGACGGCAACTACAAGA

beta-actin cDNA detection: Bactinfor1 – GTGGGCCGCTCTAGGCACCAA; Bactinrev1 – CTCTTTGATGTCACGCACGATTTC

Genotyping PCR

Genotyping PCR was carried out with geno5 and geno3 primers, 250 ng of tail DNA, and a mix like the one described for RT-PCR. Amplification conditions were 94°C for 15 min, followed by 35-40 cycles of 94°C for 30 sec, 58°C for 30 sec, 72°C for 90 sec. This was followed by a final extension at 72°C for 5 min and a 4°C hold. Tail DNA from Fo57 offspring with the insertion band by genotyping PCR were separately
amplified by using intron primers (and the same amplification conditions) to confirm the presence (or absence) of the transgene.

**Fluorescence Microscopy**

Stereofluorescence microscopy was performed with a Leica MZFLIII stereofluorescent dissecting microscope, using a 100-W mercury bulb and the following filters: exciter HQ470/40 and emitter 515 nm LP (Chroma Technology, Brattleboro, VT). Images were captured by using a MagnaFire charge-coupled device (CCD) camera (Optronics International, Goleta, CA) and formatted in PHOTOSHOP (Adobe Systems, Mountain View, CA). Genital ridges were imaged for fluorescence at ×200-400 with the MagnaFire CCD camera mounted on an inverted fluorescence microscope (DM-IL, Leica, Deerfield, IL). Long exposures revealed a consistent pattern of autofluorescence among all of the blastocysts at a given developmental stage. The following tissues were surveyed by stereofluorescence in all three lines of L1-EGFP transgenic mice: liver, lung, heart, kidney, spleen, testis, ovary, uterus, skeletal muscle, and brain. In addition, thymus, bone marrow, eye, skin, and stomach were evaluated in line 57 L1-EGFP transgenics. The following organs were evaluated by fluorescence from mice with the L1-EGFP insertion: liver, lung, heart, kidney, thymus, spleen, testis, ovary, uterus, brain, eye, embryo and skin.
Flow cytometry of organs

Testis & liver: Capsule was cut open using scissors (testis). The organ was put in 4 mL 1X trypsin-EDTA (Gibco BRL) and incubated at 37°C for 20 min with shaking. The trypsin and organ was dissociated by passing through an 18 gauge needle using a 5mL syringe into 4mL fresh trypsin and incubated at 37°C for 20 min with shaking. The solution was filtered through 70 μm Nylon cell strainer (BD Falcon). This was centrifuged for 10 minutes at room temperature at 1300 rpm to form pellet. The supernatant was aspirated and the cells were resuspended in 1 mL FACS buffer and 10 μL DNase and put on ice.

Spleen: Tissue was completely dissociated by rubbing between 2 frosted slides in 2 mL PBS. Cell suspension was filtered through 70 μm Nylon cell strainer (BD Falcon). This was centrifuged for 10 minutes at room temperature at 1300 rpm to form pellet. The supernatant was aspirated and the cells were resuspended in 1 mL FACS buffer and put on ice.

Bone Marrow: The bone marrow was put into 1mL PBS and centrifuged for 10 minutes at room temperature at 1300 rpm to form a pellet. The supernatant was aspirated and the cells were resuspended in 2 mL FACS buffer and put on ice.

ToPro3 was added to all cell suspensions for live dead gating and all organs were analyzed by flow cytometry on a FACSCaliber as above.
Histology

L1-EGFP insertion+ mice were mated to wild type females. 14 days after a vaginal plug was observed, genital ridges were harvested from embryos with the generous assistance of Raluca Verona from the Bartolomei laboratory. Genital ridges were fixed in 10% formalin. Replicate genital ridges were also stored in RNAlater for future RNA quantification studies. In addition, organs from L1-EGFP insertion and transgene mouse harvests were fixed in 10% formalin. Fixed organs were submitted to Dan Martinez in the Stokes Pathology Core Facility for embedding, sectioning, and staining. To increase specific staining and minimize background staining, three unique primary anti-EGFP antibodies have been tested: Mouse Biclonal IgG1K Anti-GFP (Roche), Mouse monoclonal IgG2a Anti-GFP (Molecular Probes), and Rabbit Polyclonal IgG Anti-GFP (Novus). The Molecular Probes antibody gives the strongest signal with the least background. Sections were stained with HRP-linked secondary antibodies specific for allotype and isotype. Sections were developed and counterstained with eosin.

Demethylation and histone acetylation (tissue culture)

143B cells were transfected with L1-EGFP or pEGFP-N1 using FuGENE6 (Roche) as above. 2 days after transfection, cells were selected for the presence of plasmid for 2 days using 0.7ug/mL puromycin. 6 days after transfection, cells were cultured in varying doses of 5’Aza-2’deoxycytidine (Sigma), Trichostatin A (Sigma), and/or Sodium Butyrate (Sigma). 14 days after transfection, wells were harvested with Versene and analyzed by flow cytometry on a FACSCaliber (BD Biosciences) as above.
Replicate wells were harvested, pooled, spun in a tabletop centrifuge at 1350 rpm for 5 minutes. Supernatant was removed, the pellet resuspended in 1 mL PBS, transferred to a 1.5mL eppie, spun at 3000 rpm in a benchtop centrifuge for 10 minutes. Supernatant was removed, and cells were resuspended in 70ul PBS. Cells were further resuspended in 450ul RNAlater (Ambion) and stored at 4° for quantification of RNA.

Embryonic fibroblasts were generated by trypsin (Gibco/BRL) digestion and dissociation of L1-EGFP insertion, L1-EGFP transgenic, and w.t. D14.5 embryos. Insertion and transgenic status was verified by genotyping and cells were cultured under demethylating/histone acetylating conditions and analyzed by flow cytometry as above. W.t. embryos were transfected with L1-EGFP by FuGENE6 (Roche) as above, cultured under demethylating/histone acetylating conditions and analyzed by flow cytometry as above.

Demethylation (in vivo)

A L1-EGFP insertion mouse was injected with 0.15 mg/kg 5’Aza-2’deoxycytidine (Sigma) in PBS i.p. three times per week for 2 weeks. A w.t. littermate was mock injected with PBS alone, and an unrelated w.t. mouse was also injected with 0.15 mg/kg 5’Aza-2’deoxycytidine. Mice were weighed before every injection. After 2 weeks, mice were sacrificed, and testis, bone marrow, spleen and liver analyzed by flow cytometry as above. Testis, spleen, liver, heart, lungs, thymus, brain, intestine, and kidney frozen in RNAlater (Ambion) for future RNA quantification studies.

135
**Bisulfite sequencing**

Spleen, brain, and testis were harvested from a line 57.1 insertion+ mouse. DNA was extracted from these tissues by sequential phenol:chloroform and chloroform extractions, followed by spooling out and washing in EtOH. Resuspended DNA was further purified by a DNeasy kit (QIAGEN) according to kit protocol. 2 µg of the purified DNA (OD260/280 1.83-1.86) was bisulfite modified using the MethylEasy kit (Human Genetic Signatures) according to kit protocol. Modified DNA was amplified in four strand-specific touchdown PCRs using the Failsafe system, buffer D. The annealing temperature dropped from 54° to 50° over 8 cycles, followed by 24 cycles at 50°. The first reaction was semi-nested, the second reaction was nested, the third reaction was single-round, and the fourth reaction was nested. Products were band purified from an agarose gel using a QIAquick kit (QIAGEN), cloned using the TOPO kit (Invitrogen), and transformed into Top10 cells. Colonies were picked, DNA extracted using the FastPlasmid kit (Eppendorf), and sequenced using the M13rev standard primer.

**Nucleofection**

1 million DT40 cells from various lines were nucleofected according to manufacturer’s instructions. Cells were washed in PBS to remove serum and resuspended in 100ul Reagent T (amaxa). 5ug of EF06R, EF05J or EF03N was added in combination with 1ug pDsRedExpress-N1 (Clontech) in a total volume of 5uL. Cells/DNA were transferred to a cuvette and nucleofected using program B-09. Cells
were returned to culture and analyzed by flow cytometry on days 1 and 6 post nucleofection on a FACSCaliber with ToPro3 staining as above.
References


89. Min, B., B.T. Weinert, and D.C. Rio, *Interplay between Drosophila Bloom's syndrome helicase and Ku autoantigen during nonhomologous end joining repair*


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


