CELLULAR CONTROL OF HUMAN MOBILE ELEMENTS

AN ABSTRACT

SUBMITTED ON THE TWENTY THIRD DAY OF MARCH 2012

TO THE BIOMEDICAL SCIENCES PROGRAM

IN PARTIAL FULFILLMENT OF THE REQIUREMENTS

OF THE SCHOOL OF MEDICINE

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

Zachary J. Faber

Approved:

Dr. Prescott Deininger, PhD
Director

Dr. Victoria Belancio, PhD

Dr. Erik Flemington, PhD

Dr. Steven Hill, PhD

Dr. Zachary Pursell, PhD

**Abstract**

The human genome contains over 500,000 copies of the retrotransposon LINE-1. Although only a limited number remain active they still represent a potent source of genetic instability in the form of insertions and the formation of DNA double strand breaks. As a result of the deleterious effects of LINE-1 activity the cell has evolved multiple mechanisms to control LINE-1 including methylation, RNA processing, and DNA repair. Here we describe two new DNA repair mechanisms involved in the control of LINE-1.

By creating cell lines that express LINE-1 constitutively we show that cells adapt to the chronic exposure to the LINE-1 endonuclease created DSBs by increasing the rate of DSB repair. The increase in DSB repair also leads to a reduction in retrotransposition of both LINE-1 and Alu. This response represents a protective mechanism utilized by the cell to limit the deleterious effects of LINE-1.

The insertion mechanism of LINE-1 and Alu involves a priming of the retroelement A-tail to a thymidine-rich endonuclease target site in the genome. The LINE-1 endonuclease cleaves sites that contain non-T bases, allowing for the creation of mismatches between the A-tail and the genomic target site, creating a substrate for the mismatch repair (MMR) pathway. Using isogenic cell lines that differ in their expression of MLH1, we show that MMR deficiency leads to an increase in LINE-1 retrotransposition. By sequencing *de novo* insertions from MMR negative cells we find an increased frequency of guanine in the target sites, making the target sites more closely

resemble the spectrum of sequences capable of being cleaved by the LINE-1 endonuclease. Using an oligo-based capture assay we determined the G/A mismatch created by an A-tail priming on a target site containing a G can be recognized by MMR, indicating the mechanism of repression by MMR involves direct recognition of mismatches.

CELLULAR CONTROL OF HUMAN MOBILE ELEMENTS

A DISSERTATION

SUBMITTED ON THE TWENTY THIRD DAY OF MARCH, 2012

TO THE BIOMEDICAL SCIENCES PROGRAM

IN PARTIAL FULFILLMENT OF THE REQIUREMENTS

OF THE SCHOOL OF MEDICINE

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

Zachary J. Faber

Approved:

Dr. Prescott Deininger, PhD
Director

Dr. Victoria Belancio, PhD

Dr. Erik Flemington, PhD

Dr. Steven Hill, PhD

Dr. Zachary Pursell, PhD

**Acknowledgements**

I would like to thank Dr. Prescott Deininger for all of the help and support throughout my PhD. You taught me how to conduct real, thoughtful science. I would also like to thank my current committee members: Dr. Belancio, Dr. Flemington, Dr. Hill, and Dr. Pursell, as well as the former members, Dr. Hemenway and Dr. Brindly. Your guidance was appreciated.

Thank you to the past and current Deininger lab members, especially Dr. Nick Wallace, for allowing me to bounce crazy ideas off you. Also thank you to Dr. Engel and the other members of the Engel and Belancio labs.

Thank you to my parents, who remained positive, even when I was not.

Lastly, I would also like to thank Suzanne. Without you I probably would have never made it. Thanks for everything.

TABLE OF CONTENTS

**List of Tables**

**List of Figures**

**Chapter 1**



**Chapter 2**

**Chapter 1: An introduction to mobile elements**

A majority of the human genome is made up of repetitive DNA (Lander *et al*, 2001). Repetitive DNA can be grouped into four distinct categories: 1) transposable elements, 2) low complexity simple repeats, 3) large duplications and 4) processed pseudogenes. Of these four classes, transposable elements are the largest group, contributing over 1200 mega bases (Mb) to the human genome. Although mobile DNA already makes up a significant portion of the genome, its contribution is most likely underestimated because old, divergent mobile elements harboring enough mutations may not resemble any current consensus element (De Koning *et al*, 2011).

Mammalian genomes contain four classes of transposable elements. First, sequences belonging to the Long INterspersed Element (LINE) family account for just over 20% of the human genome. Short INterspersed Elements (SINEs), the second largest group of transposable elements, comprising approximately 13% of the genome. The last two groups are Long Terminal Repeat (LTR) elements and DNA transposons, which contribute approximately 8% and 3% respectively to the human genome.

LINEs, SINEs, and LTR elements are mobilized throughout the genome utilizing a "copy-and-paste" mechanism that occurs through an RNA intermediate. Each successful mobilization event leads to an increase in copy number of the respective class of transposable element. DNA transposons, however, do not progress through an RNA stage in their mobilization, and instead employs a "cut-and-paste" mechanism where the DNA transposon sequence is excised from the DNA and inserted in another location.

This mechanism allows the copy number of DNA transposons to remain steady, unlike RNA-based transposable elements.. Currently, there is no evidence that demonstrate DNA transposons are actively mobilizing in the human genome. To date, the only documented, active mobile elements are LINE-1, Alu, and SINE-VNTR-Alu elements (SVA) (Belancio *et al*, 2009).

The four classes of mobile elements can be further subdivided into autonomous and nonautonomous classifications. Autonomous elements encode for key factors required for their mobilization. Nonautonomous elements do not contribute to their mobilization machinery, instead relying on another autonomous element to provide the required proteins. LINE-1 is a non-LTR autonomous retroelement that also mobilizes Alu, SVA, and processed pseudogenes.

LINE-1 sequence makes up approximately 17% of the human genome. However, when the contribution of other sequences that use the LINE-1 retrotransposition machinery for mobilization are added, the proportion of the genome that can be attributed to LINE-1 activity jumps to over one third (Lander *et al*, 2001). LINE-1 activity continues to shape the human genome, with an estimated 1 in every 1000 diseases caused by an insertion being LINE-1-mediated (Kazazian & Moran, 1998).

**LINE-1 replication cycle overview**

LINE-1 retrotransposition (Figure 1) begins with the transcription of a full length, genomic element. The resulting messenger RNA (mRNA) then travels to the cytoplasm where its two open reading frames are translated into their respective proteins. Because of the strong *cis* preference of LINE-1 proteins, the mRNA that encodes these proteins is

Figure 1

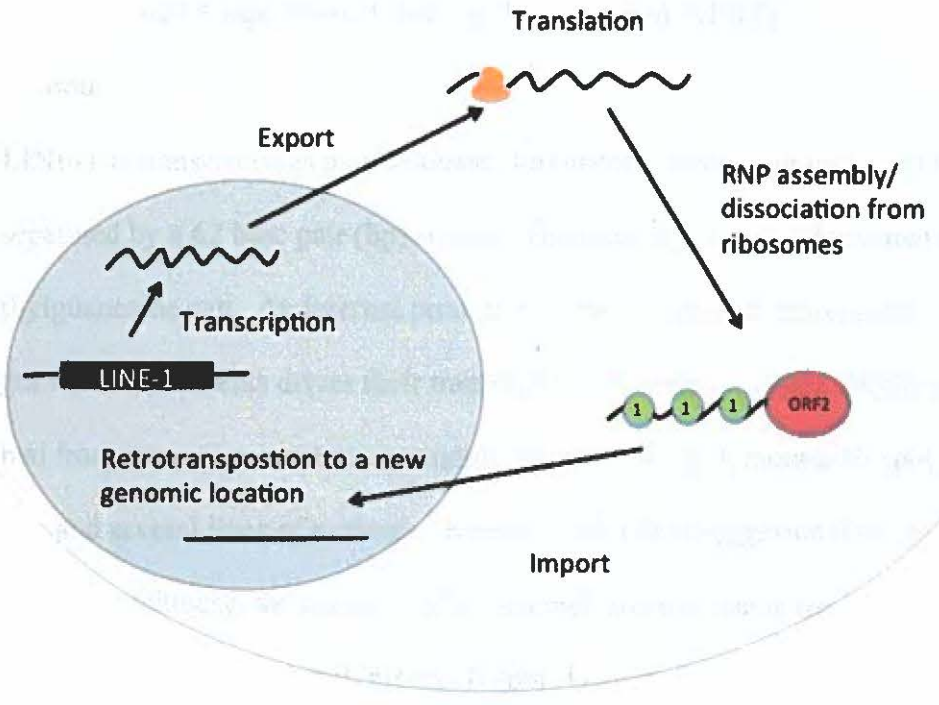



**Figure 1: LINE-1 replication cycle**

The LINE-1 replication cycle begins with transcription of a genomic LINE-1 element. The resulting mRNA then moves from the nucleus into the cytoplasm to be translated. An RNP is then formed and moves back to the nucleus where a new copy of LINE-1 can be inserted into a new genomic locus.

Kasahara (Kubo *et al*, 2006), which demonstrates that retrotransposition of a transfected LINE-1 element does not require a cell division. Without the breakdown of the nuclear membrane during cell division, the LINE-1 mRNA would be unable to escape the nucleus and therefore could not translate its protein machinery. Therefore, another mechanism must be utilized during the LINE-1 life cycle to allow the LINE-1 mRNA to translocate out of the nucleus and return with its associated proteins after translation. A transport mechanism would explain the ability of LINE-1 to propagate in non-dividing cells, as the RNP formed would be too large to cross the nuclear membrane. However, a specific transport system has not yet been identified.

**Translation**

The LINE-1 mRNA is unique because it contains two open reading frames (ORFs). Bicistronic mRNAs commonly possess an internal ribosomal entry site (IRES) that functions to facilitate ribosome initiation onto the mRNA before the second open reading frame. This ensures the proper translation of the downstream open reading frame. Studies of the mouse LINE-1 element have suggested that an IRES element is responsible for the translation of both ORFs (Li, 2006). Initially, translation of ORF2 protein was though to use a similar IRES-dependent mechanism (McMillan & Singer, 1993). However, it appears that ORF2 in humans is not translated by an IRES-dependent mechanism. Several mutations to the 62 bp inter-ORF spacer have been shown to reduce LINE-1 retrotransposition, as measured by a cell culture-based colony-forming assay. The first AUG codon in ORF2 has also been shown to be dispensable for translation. In the context of a full-length element, ORF2 translation requires an upstream ORF. The lack of an IRES and the requirement of an upstream ORF support a ribosome retention

and reinitiation model as the most likely mechanism of ORF2 protein translation. Evidence in support of this method include: the introduction of a hairpin in the inter-ORF spacer leading to a significant decrease in ORF2 protein translation and the introduction of a premature stop codon within ORF1 that abolishes ORF2 protein translation (Alisch *et al*, 2006). Additionally, human ORF1 translation is not driven by the presence of an IRES sequence, but is instead dependent on a 5' cap in human cells. Furthermore, it is possible to delete large portions of the 5' UTR and retain high levels of translation initiation (Dmitriev *et al*, 2007). High levels of cap-dependent translation, coupled with suboptimal translation of ORF2, helps to explain the large difference observed in the ratio of ORF1 protein to ORF2 protein.

**LINE-1 RNP**

Human LINE-1 proteins and RNA were first shown to form a large complex by Hohjoh and Singer (Hohjoh & Singer, 1996). Differential centrifugation experiments allowed these researchers to conclude that both ORF1 and LINE-1 mRNA were found in a complex together in the cytoplasmic fraction containing microsomes, polysomes and large macromolecules. They also described the role RNA plays in stabilizing the LINE-1 RNP and speculated that ORF1 dimers, or perhaps even higher order multimers, are formed in the RNP complex. The formation of this complex is required for retrotransposition (Kulpa & Moran, 2005). Although the other components of the RNP have yet to be elucidated, studies using a GFP-tagged ORF1 demonstrated a colocalization between complexes containing ORF1 protein with polyadenylated RNA and cellular stress granules (Goodier *et al*, 2007). Putative RNPs were also detected in the nucleus as well as the nucleolus (Goodier, 2004). These previous reports relied on

reverse transcriptase (RT) activity as a surrogate for the presence of ORF2 protein. Using an immunofluorescence-based approach, tagged versions of ORF1 protein, ORF2 protein, and LINE-1 mRNA were found to colocalize into what the Goodier group termed the "basal RNP." These RNPs also colocalized with markers of cellular stress granules, agreeing with past findings (Doucet *et al*, 2010). However, it is not known if the RNPs seen colocalizing at stress granules and in the nucleolus are actually retrocompetent or if they represent a molecular dead end.

LINE-1 proteins exhibit a *cis* preference (Wei *et al*, 2001); proteins that are encoded by an mRNA molecule tend to associate with that mRNA. Evidence supporting the *cis* preference of LINE-1 proteins consists of several lines of reasoning. First, disease causing insertions can usually be traced back to a near identical source element. If LINE-1 proteins were highly functional in *trans*, the older, mutated copies of LINE-1 would be mobilized and able to cause disease. Second, studies have shown that the LINE-1 reverse transcriptase preferentially works on LINE-1 substrates, despite the high abundance of several cellular RNAs in the RNP preparations (Kulpa & Moran, 2006). In these assays the LINE-1 RNP was enriched and collected from cells transfected with LINE-1 by sucrose gradient centrifugation. The RNPs were then incubated with nucleotides in a modified reverse transcription reaction using a primer containing an anchor at the 5' end and a stretch of thymidine bases at the 3' end. After the RT step the resulting cDNA was then be used as a template for PCR with primers specific for the anchor from the RT primer to an internal LINE-1 primer. Sequencing of the resulting products revealed that it was indeed LINE-1 sequence that was being reverse transcribed. This group found that when they targeted abundant cellular RNAs with the second internal primer, very small

amounts of products were formed from only three of the non-LINE-1 RNAs that were tested, even though those RNAs were thought to be present at a higher level than the LINE-1 mRNA. Western blotting for ORF1 protein confirmed that the RT activity found in the RNP could be attributed to LINE-1. Later, direct evidence of the interaction of ORF1 and ORF2 proteins associating with LINE-1 mRNA in cells was demonstrated using immunofluorescence. In that study, the RNA binding activity of ORF1 protein, but not its ability to form protein multimers, was important in the formation of cytoplasmic foci (Goodier *et al*, 2010).

Given the unconventional translation method of ORF2 protein, it is unsurprising that only one or two molecules of ORF2 may be present in each LINE-1 RNP and, therefore, it is advantageous for the element of those proteins to remain closely associated with the LINE-1 RNA. Although LINE-1 exhibits a strong *cis* preference, other RNAs can also be mobilized. For these RNAs to be mobilized they must gain access to the LINE-1 proteins. The mechanism by which they disrupt the LINE-1 RNP and "highjack" the retrotransposition machinery is yet unknown.

**L1 5' UTR/Promoter**

The 5' UTR contains binding sites for pol II as well as for a variety of factors. Minakami *et al.* identified a region at the immediate 5' end of the LINE-1 5' UTR that was required for transcription. He went on to describe a cellular factor that bound to that segment through a series of footprinting, band shift, and mutation assays. Those data led to the finding that the first 155 bp were responsible for the promoter activity of the element, and that the first 40 bp were the most critical (Minakami *et al*, 1992). This agreed with the previous studies by Swergold *et al.* that mapped the core LINE-1

promoter element to the first 100 bp, with the subsequent ~700 bp providing sequences that assisted in LINE-1 transcription.

The protein factor that Minakami described binding to the first 40 bp of the LINE-1 5' UTR was the transcription factor yin yang 1 (YY1) (Becker *et al*, 1993). In their analysis, deletion of the YY1 binding site decreased the expression of LINE-1 by 5-fold. They also presented data depicting the direct binding of YY1 utilizing electromobility shift assays coupled with Western blot analysis. YY1 also directs proper transcription initiation to an unconventional string of guanine nucleotides (Athanikar *et al*, 2004). Interestingly, that study also found that mutation of the YY1 site in the context of the full length 5' UTR did not decrease overall transcription or the ability of the retroelement to jump. However, when those mutations were introduced in the context of the 150 bp minimal promoter both transcription and retrotransposition potential were significantly reduced. From these results the authors concluded that downstream factors contributed to transcription and that YY1 served to maintain the full-length internal promoter in the 5' UTR. The conservation of this site is important for the element, as mRNAs which do not initiate far enough upstream will gradually erode their promoter with every mobilization, eventually leading to an element that does not contain a promoter and will therefore be unable to transcribe. Other factors, such as the sex-determining region Y (SRY) family of transcription factors (Tchénio *et al*, 2000) and runt-related transcription factor 3 (RUNX3) (Yang *et al*, 2003) have also been shown to have an effect on LINE-1 promoter activity through binding sites located in the 5' UTR.

In addition to the internal promoter that drives LINE-1 expression, the 5' UTR also contains an antisense promoter (ASP) located within the first 700 bp (Speek, 2001).

The role of this ASP is unclear at this time. It is speculated that it could interfere with transcription and therefore silences nearby genes. It is also thought that at least some of the chimeric mRNAs derived from the LINE-1 ASP may be fully functional and code for novel proteins. Searches in expressed sequence tag (EST) databases have uncovered multiple chimeric transcripts with cellular genes, including the proto-oncogene c-MET (Nigumann *et al*, 2002). There are also examples of the ASP of an upstream LINE-1 element driving transcription of a full-length cellular gene (Mätlik *et al*, 2006). In these cases the regulation of the gene is hypothesized to be more flexible, with the presence of an alternate promoter allowing for another level of control, and possibly contributing to the tissue-specific expression of certain genes.

With the LINE-1 sense promoter driving transcription from the +1 position and antisense transcript initiating near the +400-500 position, there is the possibility for the creation of double-stranded RNA (dsRNA) products that could become substrates for RNA interference (RNAi). Work by Yang and Kazazian described evidence of these dsRNAs being created and that they acted to limit LINE-1 retrotranspostion (Yang & Kazazian, 2006). They also demonstrated that the dsRNAs produced led to the degradation of LINE-1 mRNA though a Dicer-dependent mechanism.

The 5' UTR contains a CpG island that is subject to methylation at canonical CpG dinucleotides as well as at several non-CG sites (Woodcock *et al*, 1997). This methylation is thought to repress transcription of LINE-1 and is important for limiting LINE-1 associated genomic damage. The preservation of this mechanism of repression seems to be important because of the conservation of CpG sites within the LINE-1

promoter compared to the relative lack of CpG sites throughout the genome and the A/T rich nature of sequence into which LINE-1 preferentially inserts.

**ORF1 protein**

ORF1, the first of the two proteins encoded by LINE-1, is a 40 kDa polypeptide that has nucleic acid binding ability. It was first visualized from NTera2D1 and E21-2Ep teratocarcinoma cells, and was expressed at very high levels. In contrast, small amounts of ORF1 are expressed in 293 and HeLa cell lines (Leibold *et al*, 1990). Structural studies have revealed that ORF1 has a coiled-coil domain at the N-terminus (Holmes *et al*, 1992) which functions in the formation of higher order structures, and as an atypical RNA-recognition motif (RRM) (Khazina & Weichenrieder, 2009) that functions in the binding of nucleic acids (Hohjoh & Singer, 1996). The C-terminal domain of ORF1 is also thought to assist in nucleic acid binding (Khazina *et al*, 2011).

ORF1 proteins carry a positive charge at a neutral pH (Loeb *et al*, 1986), a common characteristic of nucleic acid binding proteins. This positive charge is capable of interacting with the negatively charged phosphate groups of nucleic acid backbones. However, there have been conflicting reports on the nucleic acid binding activity of ORF1 protein in regards to its sequence specificity. Using human ORF1 protein from cancer cell line extracts, Hohjoh and Singer described two fragments of LINE-1 RNA that shared some sequence homology that coprecipitated with ORF1 protein. There was no evidence of ORF1 protein binding if the target sequence was part of either a duplex of total RNA or an RNA/DNA hybrid duplex (Hohjoh & Singer, 1997). In contrast, Kolosha *et al.* found that mouse ORF1 protein purified from bacteria bound both the sense and antisense RNA from LINE-1, as well as a control transcript with no LINE-1

sequence generated from the tryptophan synthase alpha chain gene. These researchers then went on to determine that ORF1 protein was also capable of binding single-stranded DNA (ssDNA) independent of sequence, and double-stranded DNA (dsDNA), although the affinity for dsDNA was greatly reduced compared to single-stranded templates (Kolosha & Martin, 1997).

In addition to high affinity single-strand sequence binding, ORF1 protein also has nucleic acid chaperone activity. Addition of ORF1 protein to complementary oligonucleotides accelerates reannealing. It also facilitates the formation of the most thermodynamically stable duplex via strand exchange between oligonucleotides that contain single-strand DNA at one terminus or between oligonucleotides that contain internal mismatches (Martin & Bushman, 2001). The chaperone activities are concomitant with an increase in the melting temperature (Tm) of perfect duplexes and a decrease in the Tm of mismatched duplexes.

At this time the role that ORF1 plays in the lifecycle of mobile elements is unclear. ORF1 is absolutely required for the mobilization of LINE-1, (Moran *et al*, 1996) but not for Alu (Dewannieux *et al*, 2003). Attempts to elucidate the function of ORF1 have relied mostly on the analysis of point mutations, but from those experiments we have learned that the RNA binding activity of ORF1 protein contributed to the formation of the RNP. Disturbing its binding activity has a negative effect on LINE-1 retrotransposition.

**ORF2 protein**

ORF2 is the 150 kDa product of the second open reading frame of LINE-1. It consists of three conserved domains, an endonuclease (EN) domain, a reverse

transcriptase (RT) domain, and a cysteine-rich (Cys) domain. The endonuclease domain was characterized based on its homology to apurinic/apyrimidinic endonucleases, especially the critical active residues (Feng *et al*, 1996). The endonuclease activity is localized within the first 26 kDa of ORF2 and has the ability to create single-strand breaks at an A-T rich target site, as shown by the creation of relaxed circle DNA from a supercoiled substrate, leaving the cut DNA strand with a 3' hydroxyl group. The A-T rich nature of the target site specificity closely resembles the sites at which LINE-1 integrates into the genome. Longer incubations with supercoiled DNA and purified LINE-1 endonuclease domain led to the production of linear DNA, but the turnover time of the enzyme was quite low (Feng *et al*, 1996). Additionally, the EN domain can be inhibited by the presence of a poly-guanosine tract located in the 3' UTR of the LINE-1 mRNA (Cost *et al*, 2002). The ability to cut both strands of the target DNA has also been demonstrated by the formation of DNA double-strand breaks (DSBs) *in vitro* after transfection and/or with high levels of expression of LINE-1. The formation of the DNA DSBs was measured by the formation of γ-H2A histone family, member x (H2AX) foci. Formation of foci was dependent on the presence of ataxia telangiectasia mutated (ATM) and a functional LINE-1 endonuclease domain (Gasior *et al*, 2006). While the first strand is nicked at a consensus site of 3'-AA/TTTT-5', the second nick is made, on average, 7-20 bp away (Symer *et al*, 2002) with no apparent nucleotide specificity. The creation of DSBs by the LINE-1 endonuclease appears to be much more efficient than an actual insertion event. On average, 10-fold more breaks per cell can be seen, but almost all of those sites will be repaired without an insertion (Gasior *et al*, 2006). Based on results

from tagged LINE-1 retrotransposition assays, it is rare for more than one event to occur per cell.

The second conserved domain found in ORF2 contains reverse transcriptase (RT) activity and shares homology with reverse transcriptases from other retroelement families (Hattori *et al*, 1986; Malik *et al*, 1999; Xiong & Eickbush, 1988). Given the similarity to other reverse transcriptases, it is unsurprising that nucleoside analogue reverse transcriptase inhibitors suppress LINE-1 retrotransposition (Dai *et al*, 2011; Jones *et al*, 2008). Unlike other related retroviral RTs, the LINE-1 RT lacks RNase H activity. The ORF2 reverse transcriptase uses the exposed 3' hydroxyl to prime reverse transcription of LINE-1 (or another) RNA. The processivity of an RT refers to the amount of cDNA that can be retrocopied before the protein dissociates from the template. When compared to the Moloney Murine Leukemia Virus (MMLV) RT, LINE-1 RT was more processive, copying 620 nt per binding event. In the same study, the MMLV RT was only able to create products up to 120 nt long (Piskareva & Schmatchenko, 2006). For reference, the human immunodeficiency virus (HIV) RT is only able to synthesize products ranging in size from 50-100 nt before dissociating from its template (DeStefano *et al*, 1991). Template dissociation can occur due to several reasons. Most often, a thermostable hairpin or a homopolymer run of nucleotides stalls the RT and terminates the elongating chain. Both of these features were found at strong pause sites near the 3' end of the LINE-1 mRNA (Piskareva & Schmatchenko, 2006). An excess of enzyme facilitates the melting of secondary structure and the reinitiation of chain elongation. Even taking into account the high level of processivity, the low levels of ORF2 protein expression and

abundance of possible pause sites can help explain why full-length retrotransposons are rare.

The third and final domain found in ORF2 is a cysteine-rich domain (Cys) (Fanning & Singer, 1987). Although very little is known about this domain, it does contain a region that resembles a CCHC-type zinc knuckle. Nearly all retroviral nucleocapsid proteins contain a CCHC zinc knuckle; other proteins that contain the same structure have been shown to bind RNA (Muriaux & Darlix, 2010). It is possible that the Cys domain of LINE-1 functions in the same manner. Mutations in the Cys domain greatly reduce retrotransposition (Moran *et al*, 1996), but rigorous studies have yet to be completed to definitively attribute a mechanism of action to this domain.

**3' UTR**

Copies of LINE-1 end in a 3' UTR that contains a weak polyadenylation signal, as well as a run of adenine nucleotides that is most likely the result of reverse transcription of the LINE-1 mRNA polyadenine tail. Skipping of the LINE-1 polyadenylation signal and using one located downstream results in a transduction of genomic sequence located 3' of the element. Approximately 9% of genomic LINE-1 insertions have shown evidence of a 3' transduction event (Szak *et al*, 2002) ranging in size from a few bp to multiple kb (Goodier *et al*, 2000; Pickeral *et al*, 2000; Szak *et al*, 2003). These transduction events have the ability to mobilize regulatory sequences or even copy protein coding sequence to a unique location in the genome (Ejima, 2003; Moran, 1999). A stretch of guanine nucleotides in the 3' UTR is a potent inhibitor of the LINE-1 endonuclease. It is unknown if the inhibitory stretch of guanines serves a function during retrotransposition, but it has been speculated that it may inhibit a second

molecule of ORF2 until reverse transcription displaces the mRNA, delaying the creation of the DSB. Limiting DSB formation may serve two purposes. First, the damaging effect of DSBs to the cell would be lessened, and, second, by delaying the second strand nick, the ends of the DNA would remain in close proximity until reverse transcription was completed, possibly facilitating the resolution of the insertion (Cost *et al*, 2002).

While the 3' UTR is not required for retrotransposition (Moran *et al*, 1996) it may play a role in transporting the LINE-1 mRNA out of the nucleus. A binding site for nuclear export factor 1 (NXF1) is located within the 3' UTR (Lindtner *et al*, 2002). When the NXF1 binding element (NBE) found in LINE-1 was inserted into an RNA it was readily exported from the nucleus in *Xenopus* oocytes as well as mammalian cells.

**Target Primed Reverse Transcription**

LINE-1 is thought to insert into the genome by a process termed Target Primed Reverse Transcription (TPRT; Figure 2). This process has been most studied in the *Bombyx mori* R2 element, a non-LTR element that inserts into the 28s ribosomal RNA gene (Burke *et al*, 1987). Like LINE-1, the R2 element contains EN and RT activities (Burke *et al*, 1987; Xiong & Eickbush, 1988). A series of elegant experiments by Luan and colleagues first showed that the R2 element created a single-strand break at a precise location in the 28S gene and initiated reverse transcription from the 3'-OH that was created. They also demonstrated that the second strand break occurred after reverse transcription, creating a branched, DNA/RNA hybrid intermediate (Luan *et al*, 1993). While the complete TPRT reaction has not been biochemically demonstrated using LINE-1, three lines of reasoning point to insertion of LINE-1 elements via TPRT. First, LINE-1 insertions are flanked by target site duplications (TSDs), which arise during

Figure 2

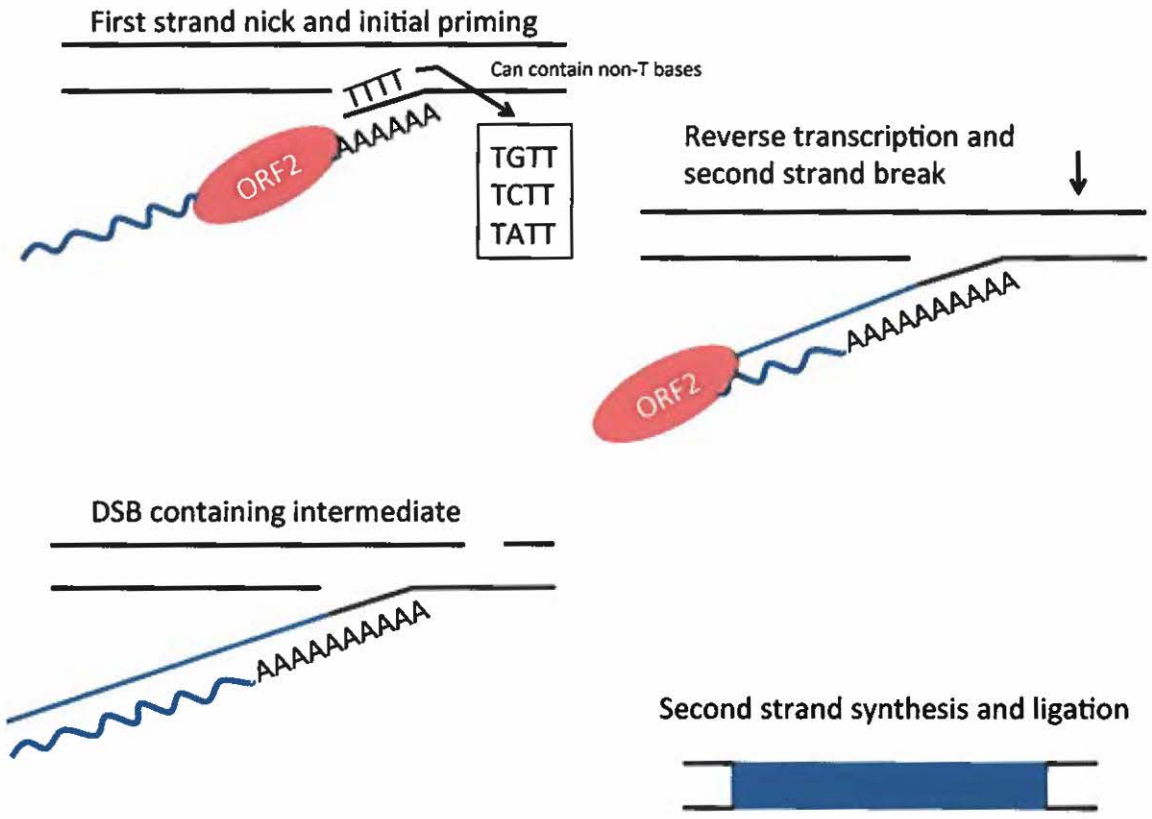



**Figure 2: Target Primed Reverse Transcription**

The proposed mechanism for LINE-1 insertion is TPRT. During TPRT the moblilized sequence anneals to a thymidine-rich stretch at an endonuclease cleavage site. The RT activity of ORF2 creates a cDNA copy of the element. A second strand break is created through an unknown mechanism, the second strand is synthesized, and the DSB is sealed, resulting in a new copy of the element.

TPRT due to the staggered nature of the DSB. These TSDs are a hallmark of TPRT resulting from the filling in of the single-stranded portions of the DSB. Second, there is complementarity between the thymidine-rich target site and the LINE-1 polyadenine tail, suggesting a priming event that leads to the reverse transcription of LINE-1 sequence. Lastly, and most compelling, partial TPRT reactions can be produced using purified ORF2 protein (Cost *et al*, 2002). In those assays, they detected the formation of branched molecules consisting of 3' LINE-1 sequence and the DNA target site that provides the 3'-OH for reverse transcription. Additionally, endonuclease cleavage of an A-T rich target site by the LINE-1 endonuclease domain (Cost *et al*, 2002) and reverse transcription of the LINE-1 mRNA (Kulpa & Moran, 2006) was demonstrated *in vitro* using purified proteins. Despite the growing evidence pointing to TPRT being the mechanism by which LINE-1 inserts, several steps of the process remain poorly understood. Second strand cleavage, the synthesis of the second strand of DNA, and the steps leading to the final integration require more study.

The current model of the TPRT reaction can be simplified to three steps: initiation, elongation, and resolution. During the initiation phase the A-T rich target site is nicked and the polyadenine tail of the LINE-1 mRNA anneals to the T rich portion of the target site. Although there is no direct evidence of annealing being required, the presence of an A-tail increases retrotransposition efficiency and internal initiation of reverse transcription is more common at sites containing a run of thymidine nucleotides. (Cost *et al*, 2002). A cDNA copy is made during the elongation phase using the 3'-OH created at the single-strand break. This copy can either proceed to the end of the mRNA, creating a full-length insertion, lose a variable portion of its 5' end, creating a truncation,

or there can be a twin priming event (Ostertag, 2001), leading to an inversion of the mRNA sequence (Figure 3).

**LINE-1 Mutagenesis**

LINE-1 is an element of genomic change. Its activity can lead to mutation through several different processes, potentially causing disease as a consequence (Figure 4). The most obvious source of mutation caused by LINE-1 is direct insertional mutagenesis, either by inserting a copy of itself, or one of the other sequences that utilize the LINE-1 retrotransposition machinery. Insertion of sequence into a gene can introduce nonsense codons, leading to truncated gene products or exon skipping. The LINE-1 sequence contains multiple splice donor and acceptor sites (Belancio *et al*, 2006), as well as premature polyadenylation signals (Perepelitsa-Belancio & Deininger, 2003; Roy-Engel *et al*, 2005). Aberrant gene products result from the use of these cryptic splice sites and spurious polyadenylation signals. Alternative gene products can also be created through the use of the sense and antisense LINE-1 promoters. Transcripts originating from the LINE-1 promoters are also capable of creating double-stranded RNA (dsRNA) and downregulating genes through the RNAi pathway. Other sequences mobilized by LINE-1 (Alu, SVA, pseudogenes) can also affect the genome in similar ways. Processed pseudogenes retrocopied by LINE-1 have, on occasion, been functional (Esnault *et al*, 2000; Maestre *et al*, 1995).

The first reports of disease caused by LINE-1 insertions were two cases of hemophilia caused by insertions into the Factor VIII gene (Kazazian *et al*, 1988). Currently, insertions into the dystrophin (Narita *et al*, 1993), adenomatous polyposis coli (APC) (Miki *et al*, 1992), β-globin (Divoky *et al*, 1996), retinitis pigmentosa 2 (RP2)

Figure 3

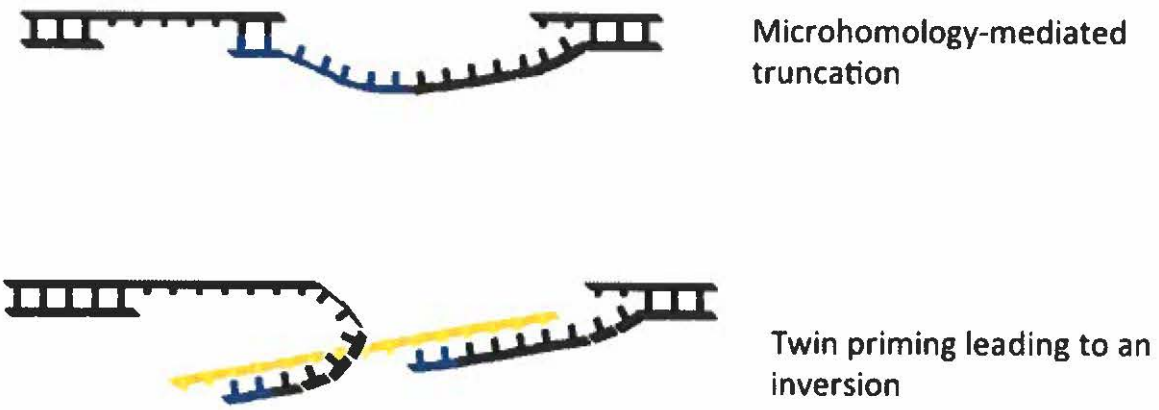



**Figure 3: Alternate finishes of a LINE-1 insertion**

Most LINE-1 insertions are not full length. The majority are 5' truncated. The truncated copies tend to have microhomology between the 5' truncation site and the flanking genomic region. A subset of insertions involves the inversion of LINE-1 sequence. In this model, a second priming occurs and when the insertion is resolved there is an inversion of the 5' sequence. Blue represents retrocopied sequence while yellow represents mRNA. Black is genomic DNA.

Figure 4

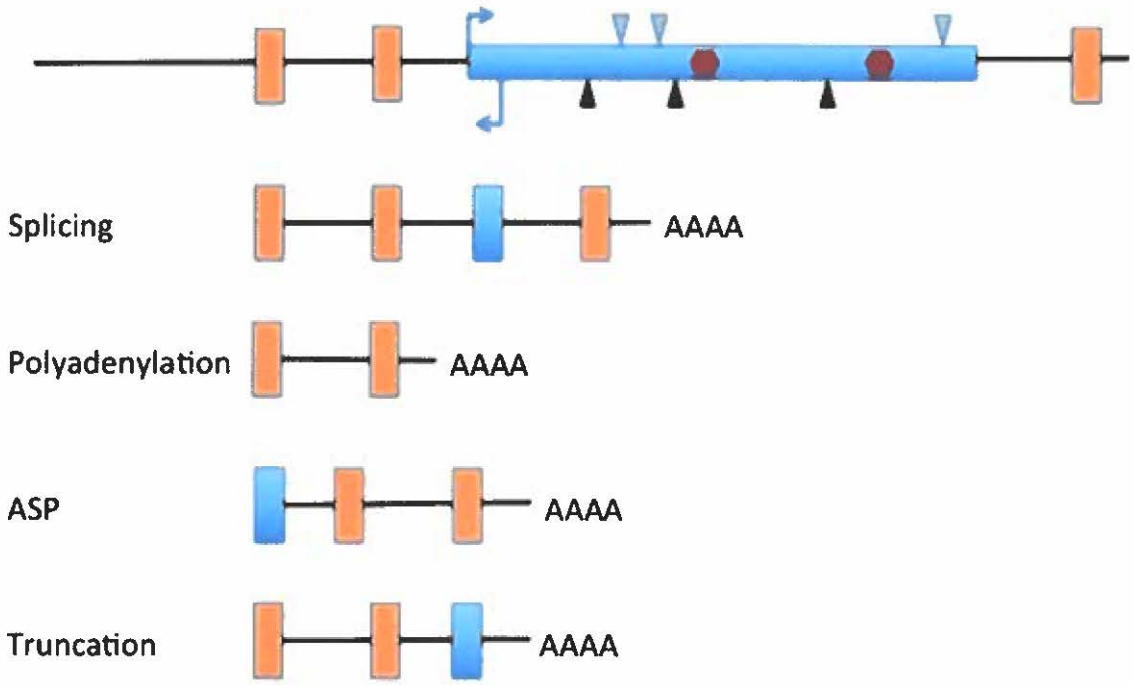



**Figure 4: Mutagenesis caused by a LINE-1 insertion**

LINE-1 insertions can affect genes in a number of ways. The splice sites and polyadenylation signals found within LINE-1 can cause abarrent splicing and premature polyadenylation of genes. The antisense promoter (ASP) can drive expression of novel gene products or deregulate a gene. Introduction of a nonsense codon can lead to a truncation of the gene product. Exons are represented by orange squares, LINE-1 by the blue rectangle. Splice sites are triangles; premature polyadenylation sites are red octagons. Opposing arrows mark the promoter and ASP.

(Schwahn *et al*, 1998), fukutin (FKTN) (Kondo-Iida *et al*, 1999), cytochrome b-245, beta (CYBB) (Meischl *et al*, 2000), V-Myc avian myelocytomatosis viral oncogene homolog (MYC) (Morse *et al*, 1988), and Factor IX (Li *et al*, 2001) genes have also been described. In the case of β-globin and RP2, the insertions were full-length copies of LINE-1, while the remainders were 5' truncated. Both of these full-length elements retained the ability to retrotranspose when placed upstream of a reporter cassette, indicating that multiple generations of LINE-1 insertions have the ability to be mutagenic (Kimberland *et al*, 1999).

LINE-1 can shape the genome without the insertion of sequences. The majority of double-strand breaks caused by LINE-1 endonuclease activity are most likely repaired through nonhomologous end joining (NHEJ) (Belancio *et al*, 2008). The other major DSB repair pathway in cells is homologous recombination, which repairs a smaller subset of LINE-1 induced breaks. DSBs repaired by NHEJ are prone to mutations stemming from inaccurate repair of the break site (Lieber, 2010). The damage done to the genome in this fashion may in fact be greater than those caused by insertional mutagenesis as the number of DSBs caused by a LINE-1 element transfected into cells is estimated to be an order of magnitude higher than the number of actual insertion events that occur in those same cells (Gasior *et al*, 2006).

The high numbers of similar sequences that make up mobile element families are very prone to recombination. The SINE Alu in particular has been implicated in a host of genomic rearrangements (Babcock, 2003; Belancio *et al*, 2010a) resulting in large-scale deletions (Gilbert *et al*, 2002) and duplications. These illegitimate recombination events

can be triggered by the activity of the LINE-1 endonuclease creating DSBs (Morales, personal communication).

**Controlling LINE-1**

Because of the numerous deleterious effects that LINE-1 can have on a cell, it is beneficial to limit its activity; cells have multiple methods of controlling LINE-1. Foremost, LINE-1 transcription is restricted in most cell types with the highest expression of LINE-1 occurring in the germ line. This repression is thought to stem from the fact that LINE-1 is normally heavily methylated at the CpGs located in its internal promoter (Woodcock *et al*, 1997). The methylation of the 5' UTR in inversely correlated to expression of ORF1 protein in a number of cell lines (Thayer *et al*, 1993). Four critical CpG sites are sufficient to silence LINE-1 (Hata & Sakaki, 1997). DNA methyltransferase I (DNMTI) is responsible for maintaining the repressive marks on CpG sites in the 5' UTR of LINE-1 (Kimura *et al*, 2003), and the position of these methylation sites appears to be more important than the overall level of methylation, indicating that the inhibition of LINE-1 by methylation occurs by the blocking of transcription factors at these four critical CpGs sites (Hata & Sakaki, 1997).

Although expression of LINE-1 can be seen in somatic cells (Belancio *et al*, 2010b), it is usually much less than that seen in the germ line. Cells that have undergone transformation also show high levels of LINE-1 expression. But even then there are a gradient of expression levels across different cancer cell lines (Belancio *et al*, 2010b). Recently it was shown that LINE-1 expression was derepressed during the reprogramming of fibroblasts to pluripotent stem (iPS) cells (Wissing *et al*, 2011). The authors of this study concluded that it was a decrease in methylation of the CpGs in the

promoter that led to an increase in expression. Although the methylation data supported their claim, they did not rule out a direct stimulation of LINE-1 transcription by one of the factors used to induce pluropotency in the fibroblasts.

Another method to regulate transcription is through the RNAi pathway. The activity of the ASP creates double-stranded RNA that is fed into the RNAi pathway. This induces LINE-1 mRNA degradation and a decrease in retrotransposon potential (Yang & Kazazian, 2006). The ASP is also thought to limit expression by competing with the sense promoter for transcription factors and the basal transcription machinery (Speek, 2001).

The presence of multiple polyadenylation and splice sites along the length of LINE-1 could also play a role in limiting the levels of retrotranspositionally competent LINE-1 mRNA. Northern blots performed with a probe to either the 5' UTR or the reporter cassette positioned at the 3' end of the element show a broad spectrum of LINE-1 mRNA species. When polyadenylation sites are mutated, nearby polyadenylation sites are used with greater frequency. LINE-1 containing ESTs also contain a variety of prematurely polyadenylation mRNA species (Perepelitsa-Belancio & Deininger, 2003). The other major process that limits the amount of full-length LINE-1 mRNA in the cell is splicing. Several splice donor and acceptor sites are utilized along the length of LINE-1, creating less than full-length mRNAs. There is evidence that spliced mRNAs are active in both humans and mice, demonstrated by insertions flanked by TSDs whose sequence matches that of the predicted spliced product (Belancio *et al*, 2006).

Although several different mechanisms act to limit the transcription of LINE-1 elements, many loci are capable of producing mRNA (Rangwala *et al*, 2009). The level

of expression of each individual locus is unlikely to be high, but the sum total of all the expression is meaningful. It is also worthwhile to note that expression of LINE-1 differs from person to person, and can potentially differ from cell to cell within the same individual.

In addition to limiting the amount of LINE-1 mRNA produced, cellular factors can influence LINE-1 directly. Members of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family were first shown to inhibit murine endogenous retroviruses MusD and intracisternal A-particle (IAP) (Esnault *et al*, 2005). The findings were advanced to show that APOBEC proteins A3A, A3B, A3C (although the inhibition is reduced), A3F, and A3G inhibit both LINE-1 and Alu retrotransposition (Bogerd *et al*, 2006; Lovsin & Peterlin, 2009; Niewiadomska *et al*, 2007). The APOBEC proteins are part of the innate antiretroviral defense mechanism and inhibit HIV by deaminating cytosine residues to uracil in the nascent proviral minus strand. Uracil containing viral DNA is then degraded in the cell (Mangeat *et al*, 2003). Some controversy surrounds these findings, as the inhibitory effect of the APOBECs does not rely on their subcellular localization or their deaminase activity (Bogerd *et al*, 2006; Niewiadomska *et al*, 2007). To date, the mechanism of the repression of LINE-1 by the APOBEC family of proteins is unknown.

Because LINE-1 causes DNA damage and associated genetic instability, it is easy to envision ways in which the DNA repair machinery would come into contact with retrotransposition intermediates. Cells deficient in excision-repair, complementing defective, in Chinese hamster, 1 (ERCC1) support LINE-1 to a greater level than cells that express ERCC1 (Gasior *et al*, 2008). ERCC1 functions in the nucleotide excision

repair (NER) pathway. It forms a heterodimer with xeroderma pigmentosum, complementation group F (XPF) and has two known functions, 5' bubble endonuclease and 3' flap endonuclease. It is probable that the 3' cDNA flap created as an intermediate of TPRT can be recognized by the ERCC1/XPF heterodimer. The excision of the flap would lead to the decrease seen in retrotransposition rate.

Other DNA repair proteins appear to facilitate LINE-1 retrotransposition. The first demonstration of a cellular protein being co-opted by the LINE-1 machinery to facilitate integration was ATM. LINE-1 could not undergo retrotransposition in cells devoid of ATM, but an isogenic cell line in which ATM was expressed allowed for retrotransposition (Gasior *et al*, 2006). The authors speculated that either ATM was required for the processing of the DSB intermediate formed during retrotransposition or that the kinase activity of ATM was required to phosphorylate either ORF1 or ORF2 and that modification was required for retrotransposition. Recently, contradictory data have been published that argue that ATM actually inhibited LINE-1 retrotransposition (Coufal *et al*, 2011). By crossing ATM knockout mice with mice containing a GFP tagged LINE-1 reporter cassette, the investigators concluded that ATM deficiency led to increased retrotransposition in the brain. They also documented the same effect in neural progenitor cells and colon cancer cells transduced using lentiviral hairpins to reduce ATM expression levels. The discrepancy between these two sets of data may be because the requirements for LINE-1 retrotransposition differ from cell to cell, where it is required for retrotransposition in fibroblasts and dispensable in neuronal and colon cancer cells. It is also possible that the hairpins used to decrease ATM expression in the latter study are causing off-target effects unrelated to LINE-1 and ATM. Toxicity due to

LINE-1 may also be higher in the fibroblast lines lacking ATM, leading to the death of the cells before selection of the LINE-1 retrotransposition event.

Other cellular proteins have also been shown to modulate LINE-1 retrotransposition. Using DT40 chicken B-cell lines, Suzuki *et al.* established that proteins in the NHEJ pathway assisted LINE-1 retrotransposition. In addition to a decrease in rate in NHEJ mutant cells, the authors of the study also described an increase in the average length of insertions in Ku70 -/- and Artemis -/- cells. Inhibition of NHEJ by addition of the DNA-dependent protein kinase, catalytic subunit (DNA-PKcs) inhibitor, NU7026, to HeLa cells also decreased LINE-1 retrotransposition (Suzuki *et al*, 2009). Because the level of retrotransposition in the mutant cells did not drop to zero, NHEJ was not absolutely required for retrotransposition, but appeared to play a role in the ultimate resolution of a subset of insertion events.

**LINE-1, DSBs, and toxicity**

The cell faces constant assault from the LINE-1 EN, creating DNA double-strand breaks. DSBs are one of the most cytotoxic lesions, which can lead to cellular senescence and apoptosis, as well as promote chromosomal rearrangements (Kass & Jasin, 2010). Two major pathways govern DSB repair in the cell, homologous recombination (HR) and NHEJ. HR is considered to be the more faithful repair pathway, conserving genetic information by using a homologous sequence as a template, whereas NHEJ has the potential to be non-mutagenic, but often results in small deletions at the break site (Jackson & Bartek, 2009).

Not surprisingly, the DSBs created by LINE-1 endonuclease cleavage led to an accumulation of cells in G2/M, induction of the pro-apoptotic gene Bax, and activation of

caspase 3 in a breast cancer cell line (Belgnaoui *et al*, 2006). The combination of those three factors strongly suggested that apoptosis was triggered in cells expressing LINE-1. Apoptosis is only one possible response to genomic damage. Cells can also respond to genetic damage by entering senescence, a permanent removal from the cell cycle (Jackson & Bartek, 2009). Expression of LINE-1, both full length and ORF2 alone, can cause cells to enter into a senescent state (Wallace *et al*, 2008) as measured by the expression of a senescence-specific β-galactosidase (Itahana *et al*, 2007).

**Tools for studying LINE-1 retrotransposition**

The sheer number of highly homogeneous LINE-1 elements in the human genome makes their study technically challenging. An important breakthrough in the study of retroelements occurred when a vector system was designed to allow for the selection of a single LINE-1 insertion event. This system involved the cloning of a reporter cassette at the 3' end of a LINE-1 element. The reporter gene is in the reverse orientation compared to the LINE-1 sequence and is interrupted by an intron in the sense orientation (Moran *et al*, 1996). Only upon successful transcription, splicing, and reverse transcription will a functional reporter gene be created (Figure 5). Expression of the reporter gene allows for the selection of cells harboring a new event with antibiotics. The system has also been adapted for use with Fluorescence-Activated Cell Sorting (FACS) protocols by interrupting a GFP gene with an intron. Successful retrotransposition events lead to GFP expression and green fluorescence.

The inclusion of a bacterial origin of replication between the reporter gene and the poly-A tail, combined with the use of a dual prokaryotic/eukaryotic promoter to drive expression of the reporter gene allows for the rescue of individual LINE-1 insertion

Figure 5

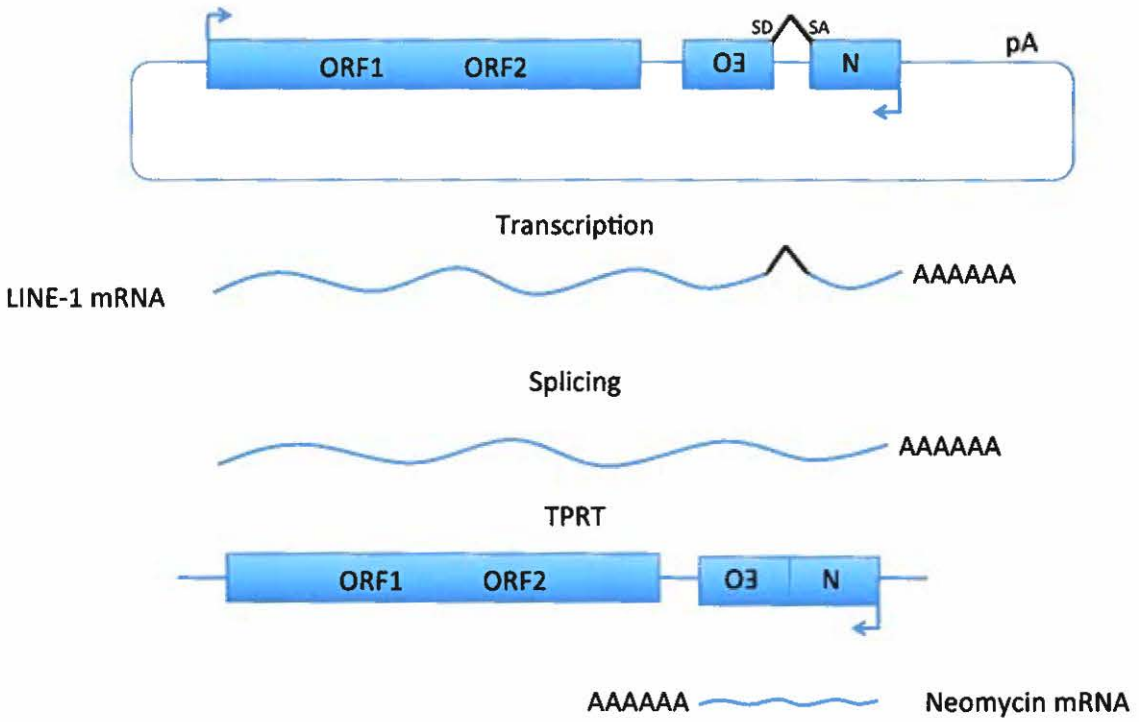



**Figure 5: The tagged LINE-1 retrotransposition assay**

To study retrotransposition a system was developed (Moran, Cell, 1996) that utilized an inverted neomycin resistance gene located 3' of the LINE-1. The neomycin resistance gene is interrupted by an intron in the sense orientation. Upon transcription, splicing, and successful retrotransposition, the resistance gene becomes active, conferring resistance to Geneticin. Arrows represent transcription start sites. SD and SA mark the locations of a splice donor and splice acceptor, respectively.

events. By digesting the genomic DNA with a restriction endonuclease that does not cut within the LINE-1 reporter coding region and then recircularizing the resulting DNA, a functional plasmid is created that can be subsequently transformed into bacteria and sequenced (Gilbert *et al*, 2002).

While insertions of these "tagged" LINE-1 elements do show all of the typical hallmarks of retrotransposition (TSDs, truncations, target site preference, etc.), they do represent an artificial system. For an insertion event to be scored, the entirety of the reporter gene must be reverse transcribed, and the element must land in a spot in the genome that is amenable to transcription. The rescue of insertion events might also be plagued by bias, with shorter insertions and insertions occurring nearer to restriction sites being overrepresented.

**Hypothesis and direction**

With the observed DNA damage caused by the LINE-1 endonuclease and the probable interactions with DNA damage sensor and repair proteins, we believe there is potential for other DNA repair pathways to regulate LINE-1. The current accepted model of the TPRT reaction creates DNA double-strand breaks that need to be repaired for survival. We found that cells that were stably transformed with LINE-1 ORF2 were subjected to the chronic presence of DNA DSBs. This constant genomic assault caused an increase in the rate of DSB repair and led to a reduction in LINE-1 and Alu retrotransposition.

TPRT also has the potential to create multiple insertion intermediates may be recognized as "abnormal" structures by the cell. One such intermediate is a branched molecule creating a small stretch of DNA/RNA hybrid duplex (Figure 2). The annealing

of the RNA on the DNA target site can create mismatches and as such possibly create a substrate recognizable by the mismatch repair (MMR) pathway.

## Chapter 2: Feedback inhibition of LINE-1 and Alu retrotransposition through altered double strand break repair kinetics

### Introduction

Of the 500,000 copies of LINE-1 in the genome, approximately 80 to 100 are full length and contain complete open reading frames. The second open reading frame of LINE-1 (ORF2) contains endonuclease (EN) and reverse transcriptase (RT) domains. Both of these domains are required for successful mobilization of LINE-1 and the sequences co-mobilized by the LINE-1 retrotransposition machinery. These sequences increase in copy number throughout the genome through a "copy-and-paste" mechanism termed Target Primed Reverse Transcription (TPRT). One required element for successful insertion through TPRT is the creation of a DNA double strand break (DSB). Multiple tissues and cell lines express LINE-1 and are therefore subjected to constant exposure to ORF2 protein (Belancio *et al*, 2010b), which has been shown to cause DNA DSBs (Gasior *et al*, 2006) and lead to multiple types of toxicity (Belgnaoui *et al*, 2006; Haoudi *et al*, 2004; Wallace *et al*, 2008). DSBs are highly toxic to cells and their repair is carried out through a combination of homologous recombination (HR) and non-homologous end joining (NHEJ). When cells are treated with low doses of the radiation-mimetic drug Zeocin they experience persistent DSBs. The cells that survive the treatment have a reduced potential for HR and an increased NHEJ activity (Delacôte *et al*, 2007). Increased NHEJ allows the cell to repair DSBs more quickly, as it is active in all cell cycles, whereas HR requires homologous sequence and is most active in

late S phase and early G2 (Kass & Jasin, 2010). However, repair by NHEJ can lead to mutations. A DSB that cannot be ligated must be processed. This can lead to nucleotide deletions, insertions, or substitutions (Kuhfittig-Kulle *et al*, 2007).

The correlation between the rates of LINE-1 retrotransposition and DNA damage and repair system has already been established. Although the nucleotide excision repair flap endonuclease ERCC1/XPF is known to have a limiting effect on LINE-1 retrotransposition (Gasior *et al*, 2008), LINE-1 EN generates more DSBs than actual retrotransposition events. Combined with its constitutive expression in multiple tissues and cell lines, it is possible that long term exposure to the LINE-1 EN and its associated DSBs may cause changes to the DNA damage response.

By creating stable cell lines that express either active or inactive ORF2 protein we have shown that cells have the ability to adapt to the constant creation of DSBs caused by the LINE-1 EN domain. This adaptation reduced the retrotransposition potential of both LINE-1 and Alu and abrogated the LINE-1-associated toxicity usually seen as a result of the acute overexpression of ORF2 protein. The cells also displayed an increase in the rate of DSB repair, indicating a possible mechanism by which the repression of retrotransposition and the reduction in toxicity occurred (Wallace *et al*, 2010).

**Results**

**LINE-1 expression inversely correlates with retrotransposition potential**

LINE-1 retrotransposition requires a full-length RNA. Spliced and prematurely polyadenylated RNAs are inactive. In order to quantify full-length, active LINE-1 mRNAs, we performed Northern blot analyses using HeLa, HCT116, and MCF7 cells. The highest expression of LINE-1 mRNA was in MCF7 cells and the amount of full-

length LINE-1 expression in HeLa cells was only 15% of that seen in MCF7 cells. No full-length LINE-1 mRNA was detected in the HCT116 cells (Figure 6).

We next measured the relative retrotransposition potential of each cell line using a transient transfection of a "tagged" LINE-1 element. Successful retrotransposition was determined by cell resistance to the antibiotic G418 (Moran *et al*, 1996). In contrast to our Northern blot findings; we were unable to visualize any retrotransposition events in the MCF7 cell line. Conversely, HeLa cells displayed an intermediate rate of LINE-1 retrotransposition and HCT116 cells produced twice as many colonies as the HeLa cells (Figure 7). One interpretation of these data suggests that the level of LINE-1 expression inversely correlated with retrotransposition potential. However, direct comparison of these different cell lines was problematic. While the correlation between the expression of LINE-1 and retrotransposition potential did not support the idea that cells can adapt to constitutive LINE-1 expression, there are other innate differences between these cells lines that may also have an effect on LINE-1 retrotransposition.

To address this concern we generated isogenic HeLa cell lines that differ only in their expression of full-length LINE-1 or the ORF2 protein. As a control, we also generated HeLa cell lines that express a mutant form of ORF2 protein, incapable of both endonuclease cleavage and reverse transcription, or the unrelated control protein, luciferase. To confirm expression of functional ORF2 protein in these cell lines, we assayed for the presence of p53 binding protein 1 (53BP1) foci, which are indicative of DNA double strand breaks, by immunofluorescence. Cells that expressed functional ORF2 protein showed significantly increased levels of 53BP1 foci compared to control

Figure 6

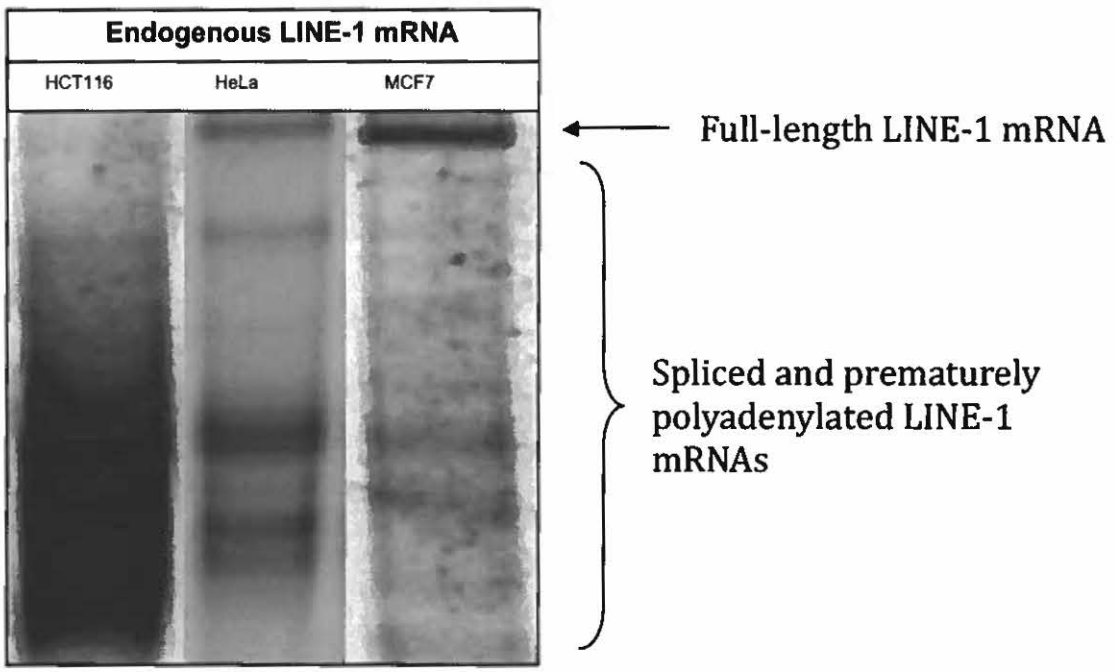



**Figure 6: Northern blot of endogenous LINE-1 mRNA in cancer cell lines**

Polyadenylated RNA was collected from HCT116, HeLa and MCF7 cancer cell lines and fractionated. LINE-1 mRNA was visualized using a probe specific to the 5' UTR. Full length LINE-1 mRNA is marked with an arrow. The bracket denotes smaller molecular weight products created by splicing or the premature polyadenylation of the LINE-1 transcript.

Figure 7

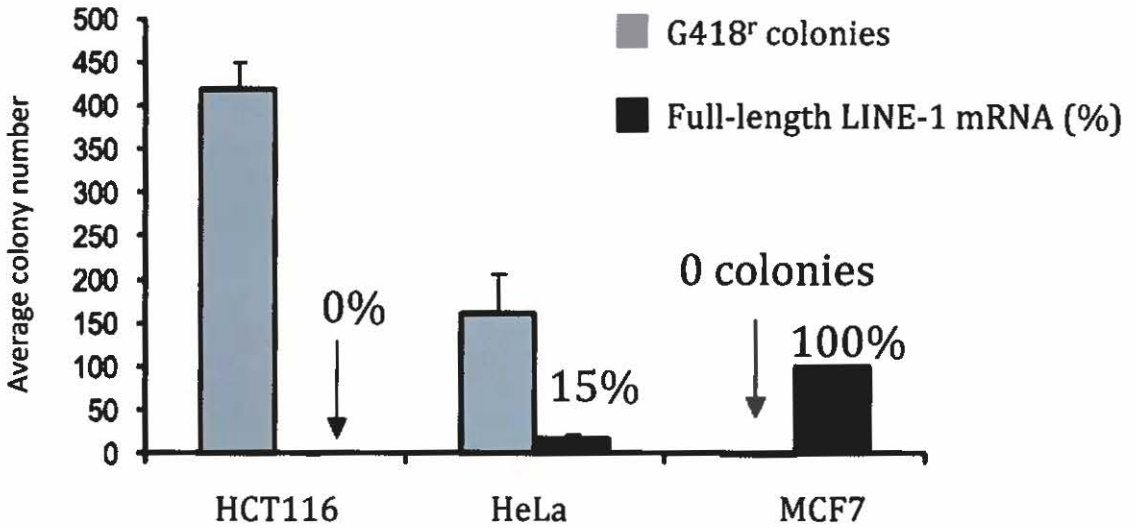



**Figure 7: Inverse correlation between full-length LINE-1 expression and LINE-1 retrotransposition potential**

Comparison of the level of full length LINE-1 mRNA and the retrotransposition potential of the three cancer cell lines: HCT116, HeLa and MCF7. The indicated cell lines were transfected with LINE-1 tagged with a neomycin reporter gene. Successful retrotransposition grants resistance to G418. Full-length RNA measurements are densitometry of the Northern blot in Figure 6. Error bars indicate standard deviations.

cells (Figure 8). We further confirmed that ORF2 protein was overexpressed in the cell lines by Western blot analysis (Figure 9).

To examine the effects of chronic LINE-1 expression on retrotransposition, we measured the retrotransposition rates of both LINE-1 (Moran *et al*, 1996) and Alu (Dewannieux *et al*, 2003) when driven by transient expression of ORF2 protein. In HeLa cells chronically expressing functional ORF2 protein, both LINE-1 and Alu retrotransposition were significantly reduced compared to cells expressing the mutated ORF2 protein or luciferase. The magnitude of repression was greater for Alu (8.6% of control vs 50.6%) (Figure 10).

We next tested if increased expression of LINE-1 would result in even further repression. To generate cells that expressed LINE-1 at an increased level we used a codon-optimized LINE-1 (OptL1) (Wallace *et al*, 2008). When LINE-1 and Alu retrotransposition were tested in these cells, both elements were almost completely inhibited (Figure 10).

Further analysis of the repression of Alu retrotransposition revealed that the repression is not specific to a single source of ORF2. When Alu retrotransposition was driven by either full-lengh LINE-1 or just ORF2 alone in HeLa ORF2 or HeLa mutant ORF2 cells, the repression remained similar (Figure 11).

To ensure that the repression of retrotransposition by the chronic expression of LINE-1 was not cell type specific we also generated HCT116 cell lines expressing either ORF2 or luciferase. LINE-1 and Alu retrotransposition were also suppressed in the ORF2 expressing HCT116 cells compared to the luciferase control (Figure 12).

Figure 8

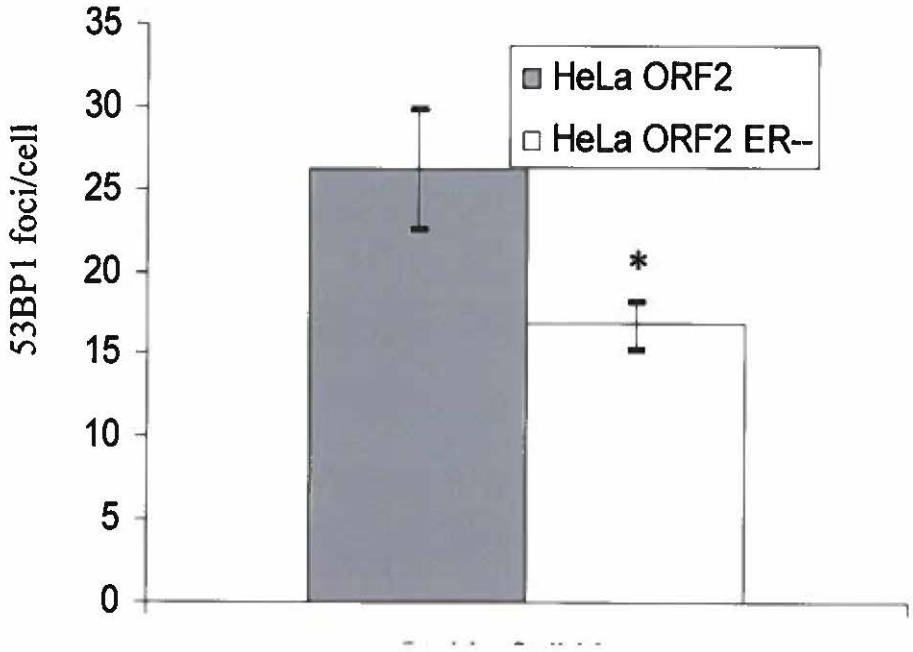



**Figure 8: HeLa cells constitutively expressing active ORF2 have increased numbers of DNA DSBs**

53BP1 foci were quantified in cells that have either ORF2 or mutant ORF2 (ORF2 ER--) stably integrated by immunofluorescence. Error bars represent standard deviation. The cell lines differ significantly in the number of 53BP1 foci present ($p<0.05$, two-tailed Student's t-test).

Figure 9

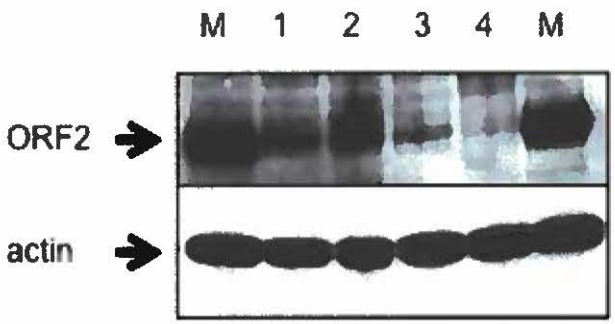



**Figure 9: Stable integration of ORF2 leads to overexpression**

Western blot analysis was performed on total cell lysates of HeLa cells transiently transfected with an ORF2 expression construct (M), stably transformed with mutant ORF2 (Lane 1), stably transformed with an ORF2 expression construct (Lane 2), stably transformed with a codon optimized full-length LINE-1 vector (Lane 3), or untransfected HeLa cells (Lane 4), using an antibody to the N-terminal portion of ORF2. The membrane was also probed with an actin antibody, which served as a loading control.

Figure 10

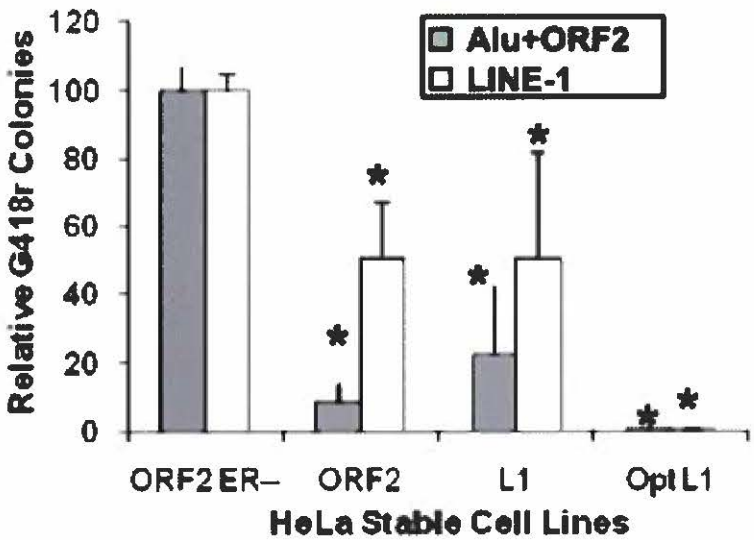



**Figure 10: Constitutive expression of LINE-1 leads to reduced LINE-1 and Alu retrotransposition**

LINE-1 and Alu retrotransposition assays were performed in the stably transformed HeLa cell lines indicated along the x-axis. ORF2ER-- is a mutant ORF2. The ORF2 cell line has been stably transformed with an ORF2 expression construct. L1 refers to HeLa cells stably transformed with a full-length LINE-1 construct. The OptL1 cell line is transformed with a LINE-1 construct that has been codon optimized. Error bars represent standard deviation and asterisks signify a significant difference from the mutant ORF2 control cell line ($p<0.05$) measured by the two-tailed Student's t-test.

Figure 11

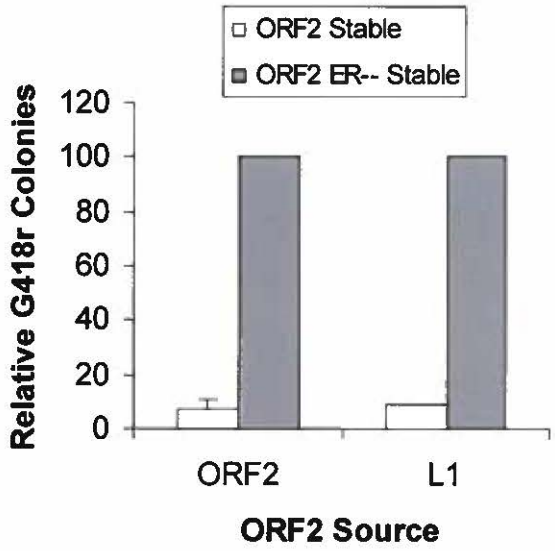



**Figure 11: The repression of Alu retrotransposition is not specific to a single source of ORF2**

Alu retrotransposition was driven by cotransfection of either full-length LINE-1 (L1) or by a construct containing only ORF2. ORF2 ER-- denotes the mutant ORF2 transformed HeLa cells. HeLa cells stably transformed with ORF2 are labeled ORF2 Stable. Successful Alu retrotransposition generates G418 resistance in the cell. Error bars represent standard deviation.

Figure 12

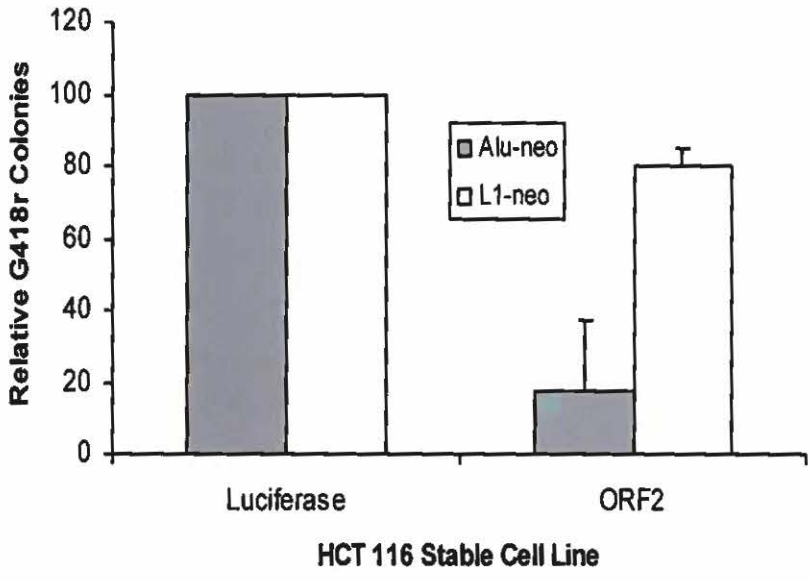



**Figure 12: Repression of retrotransposition by chronic expression of LINE-1 is not cell type specific**

HCT116 cell lines were generated stably expressing either ORF2 or luciferase. LINE-1 and Alu retrotransposition assays were carried out in these cells. Error bars represent standard deviation.

**Constitutive LINE-1 expression leads to decreased toxicity of high levels of transient ORF2 expression**

The constitutive expression of ORF2 protein led to persistent occurrence of DSBs. As a result, the cells expressing ORF2 may grow at a reduced rate or be less fit than cells not exposed to ORF2 protein and, therefore, display an apparent decrease in retrotransposition. We then observed the growth kinetics of HeLa ORF2 and HeLa ORF2 mutant cell lines by measuring cell densities over three days. Although the HeLa cell line expressing ORF2 was subjected to the constant generation of LINE-1 induced DSBs, its growth rate was not significantly different from cells expressing mutant ORF2 (Figure 13). These results indicate that the decrease seen in retrotransposition was not due to a decrease in the overall growth rate of the cells.

High levels of expression of LINE-1 can be extremely toxic to cells, resulting in a reduction of proliferation, senescence (Wallace *et al*, 2008), and apoptosis (Belgnaoui *et al*, 2006), which stems from the generation of DSBs (Gasior *et al*, 2006). An increase in sensitivity to LINE-1 expression may explain the reduction in retrotransposition seen in the ORF2 protein-expressing cell lines. To address this possibility, we measured the toxicity of transfected LINE-1 by cotransfecting it with a construct that confers resistance to G418. Cotransfection of ORF2 and the neomycin resistance vector resulted in significantly reduced numbers of colonies in the HeLa cells stably transfected with mutant ORF2 compared to the HeLa cell line stably expressing functional ORF2 (Figure 14). These results imply that the decrease in retrotransposition of the HeLa-ORF2 cells was not due to an increase in LINE-1 mediated toxicity. Instead, these findings suggest

Figure 13

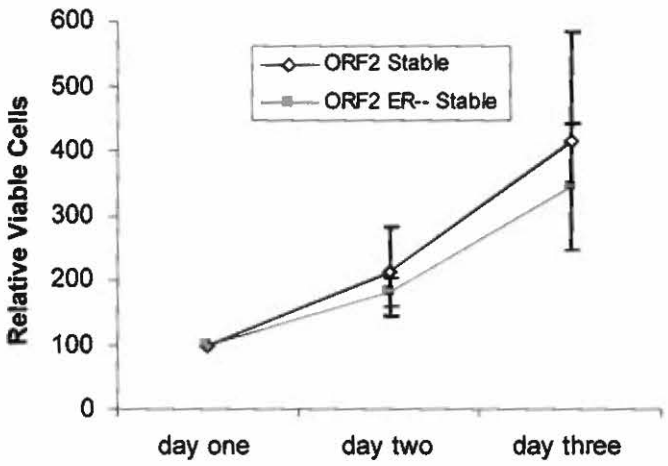
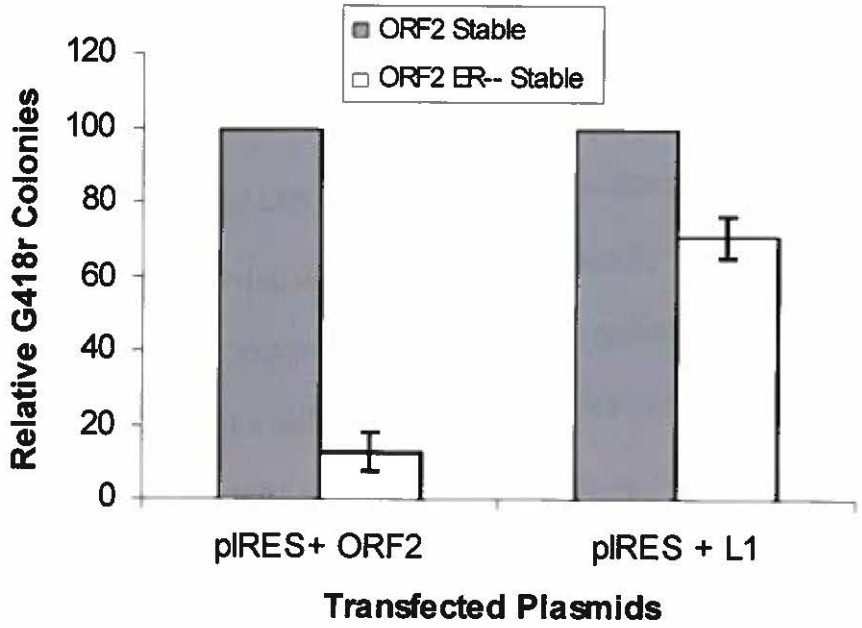



**Figure 13: Chronic expression of LINE-1 ORF2 does not affect HeLa cell proliferation**

$5x10^5$ HeLa ORF2 or HeLa ORF2 mutant (ER--) were seeded in T-75 tissue culture flasks. One, two, and three days after seeding, independent flasks were trypsinized and stained with trypan blue to mark dead cells. Live cells were counted using a hemocytometer. The number of live cells on day one was set to 100. Error bars represent standard deviations.

'igure 14




**igure 14: Persistent expression of LINE-1 ORF2 reduces toxicity associated with 'ansient LINE-1 expression**

To measure LINE-1-associated toxicity, a neomycin resistance plasmid, pIRES-GFP, was cotransfected into HeLa ORF2 and HeLa ORF2 mutant (ER--) cells with an (pression construct containing either full length LINE-1 (L1) or ORF2. Cells were 'own in the presence of G418 for 14 days. Colony formation was then used as a easure of LINE-1 related toxicity. Error bars represent standard deviation.

Figure 15

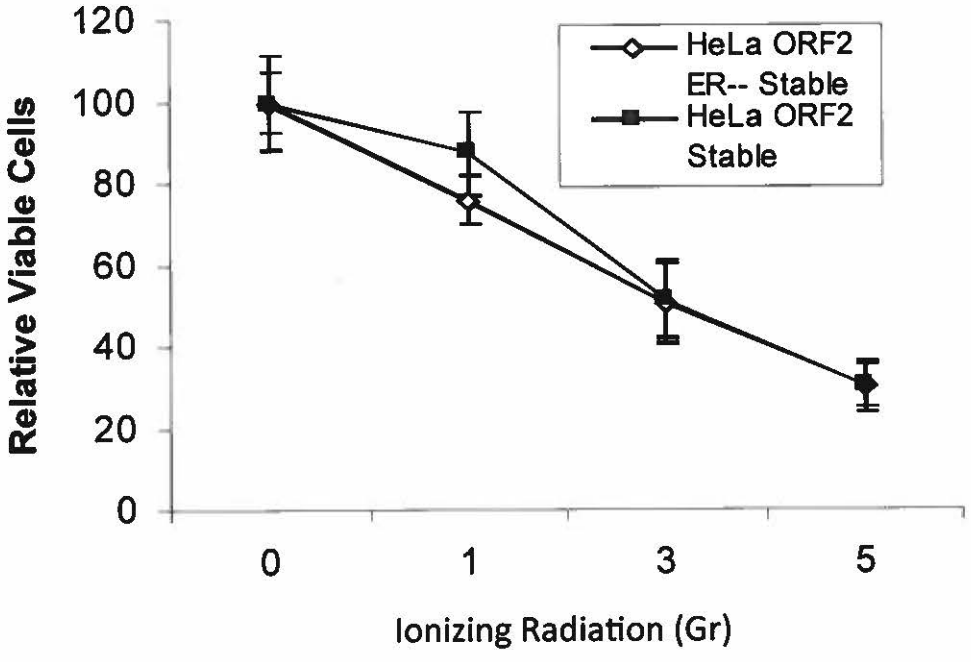



**Figure 15: Long-term expression of ORF2 does not alter the sensitivity to ionizing radiation in HeLa cells**

HeLa cells expressing ORF2 or a mutant ORF2 were exposed to variable levels of ionizing radiation. After a three-day recovery period the number of viable cells was counted using trypan blue exclusion and a hemocytometer. Error bars represent standard deviation.

Figure 16

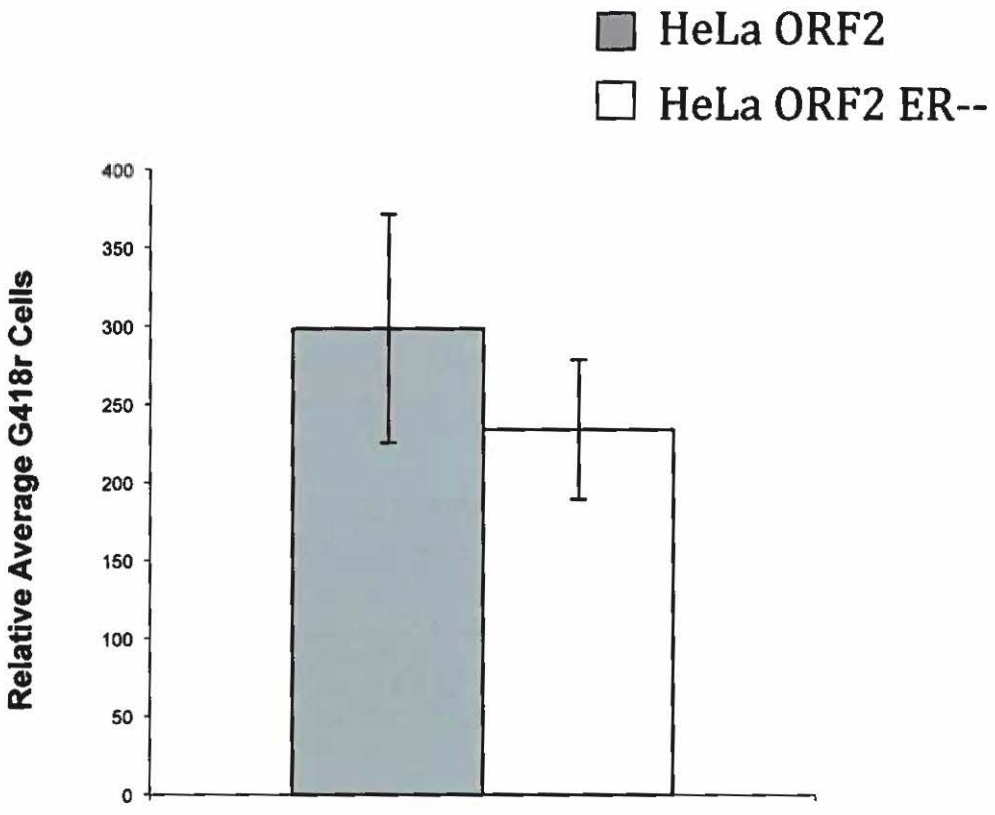
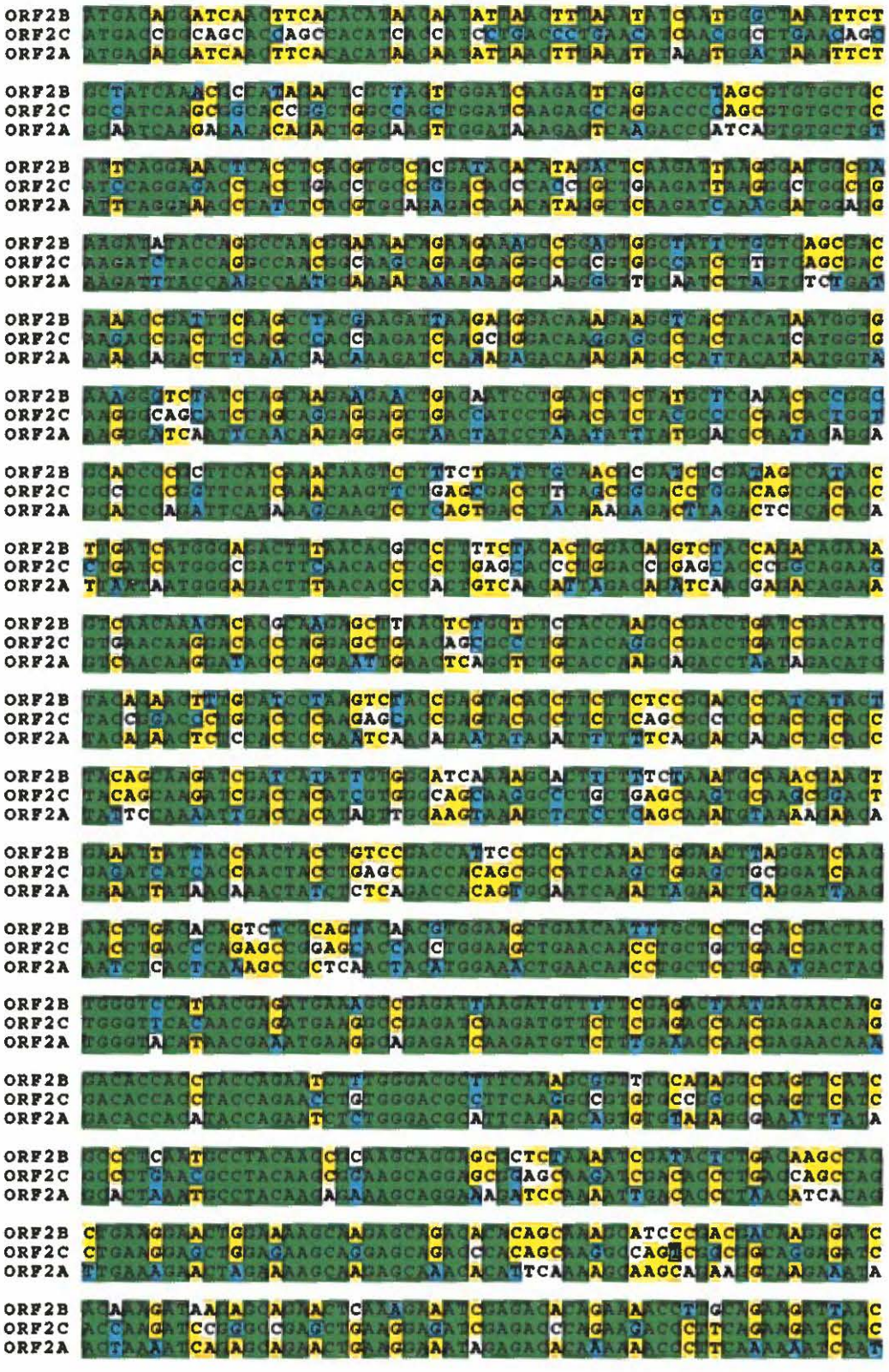
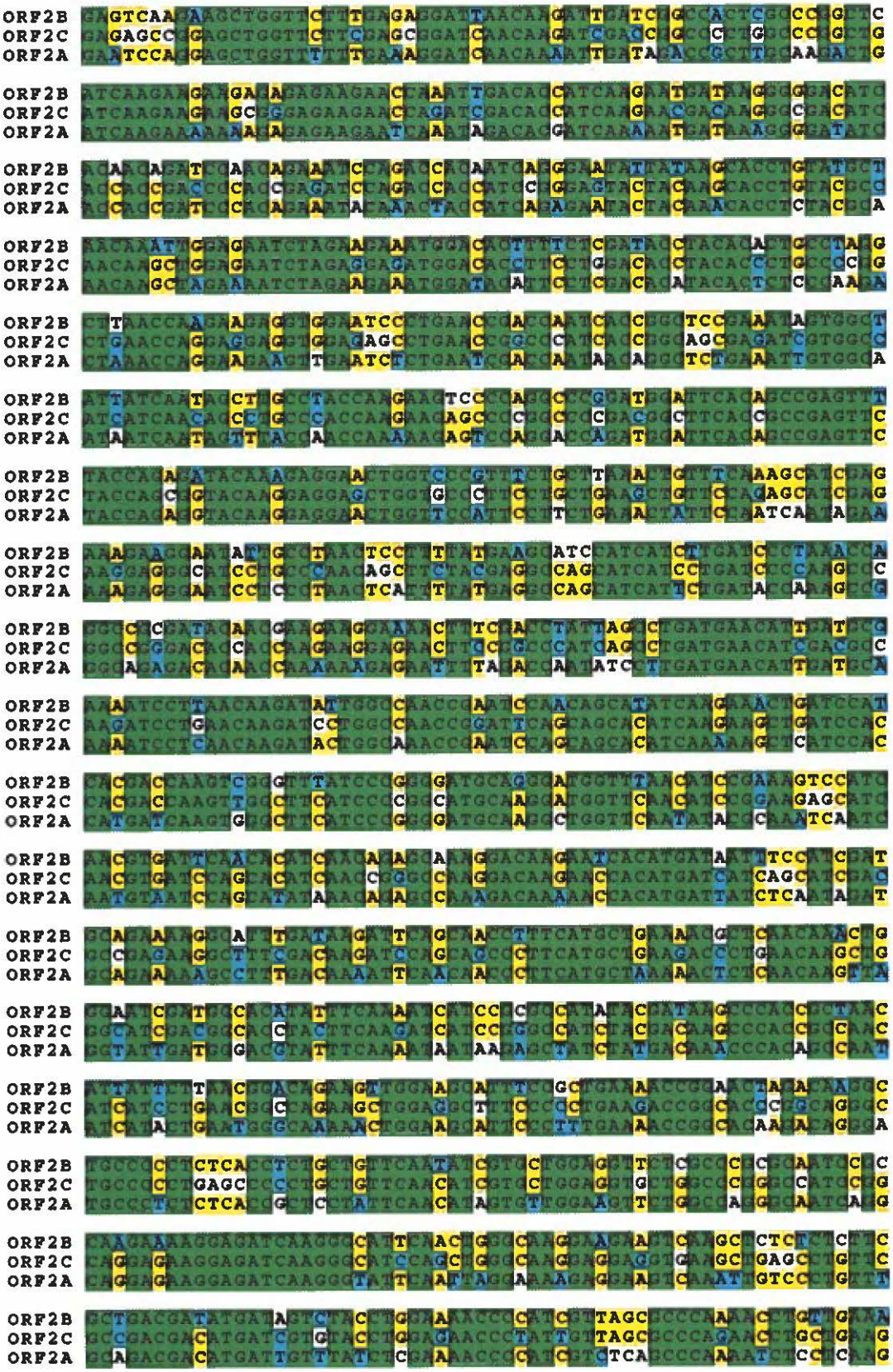


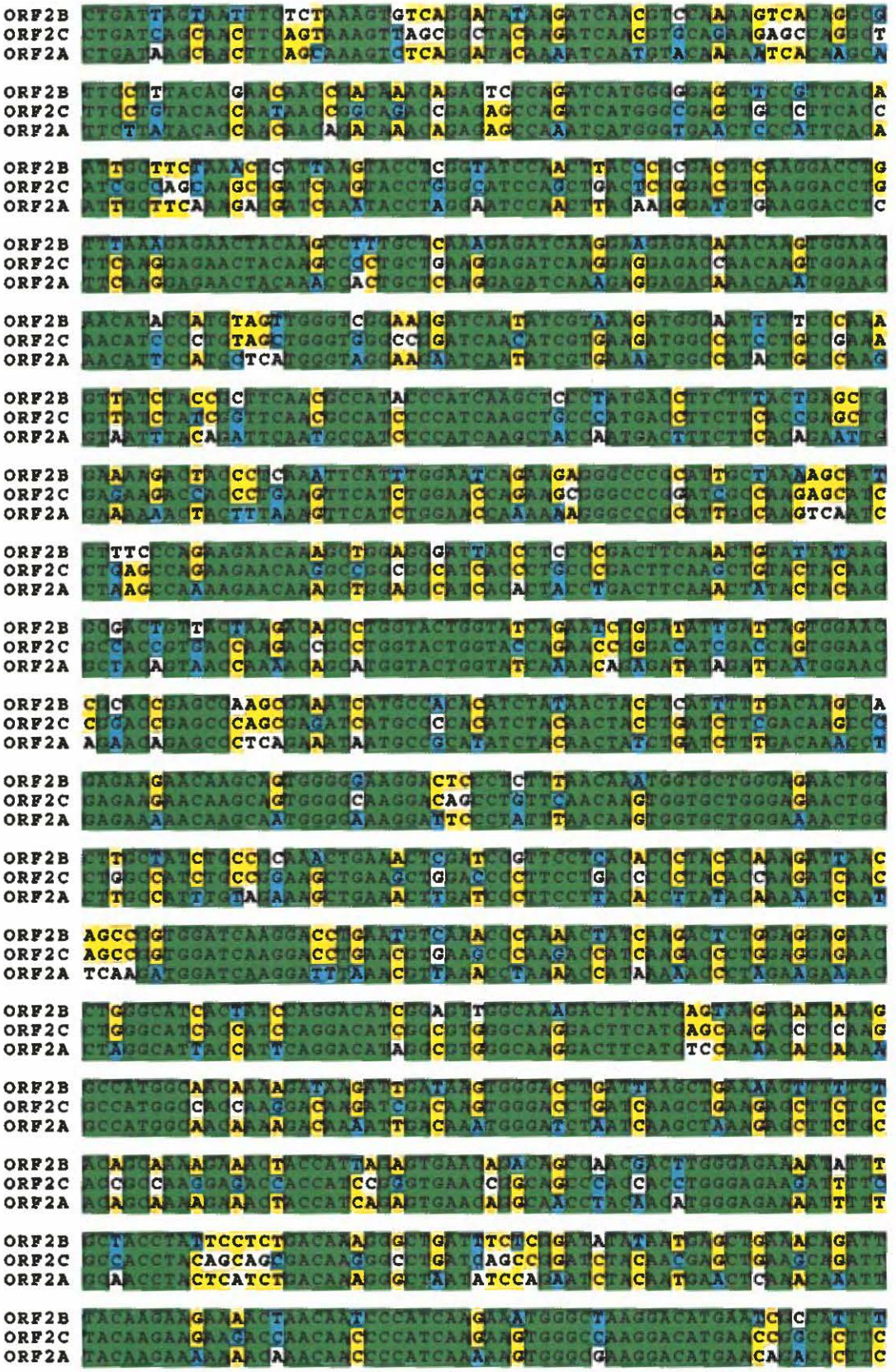

**Figure 16: Constitutive ORF2 expression does not affect HeLa cell sensitivity to ionizing radiation**

HeLa ORF2 and HeLa ORF2 mutant cell lines were exposed to one gray of ionizing radiation (IR) and then transfected with a neomycin resistance construct to measure colony-forming ability following IR-induced DSBs. Error bars represent standard deviation.

Figure 17




ORF2B
ORF2C
ORF2A


ORF2B
ORF2C
ORF2A

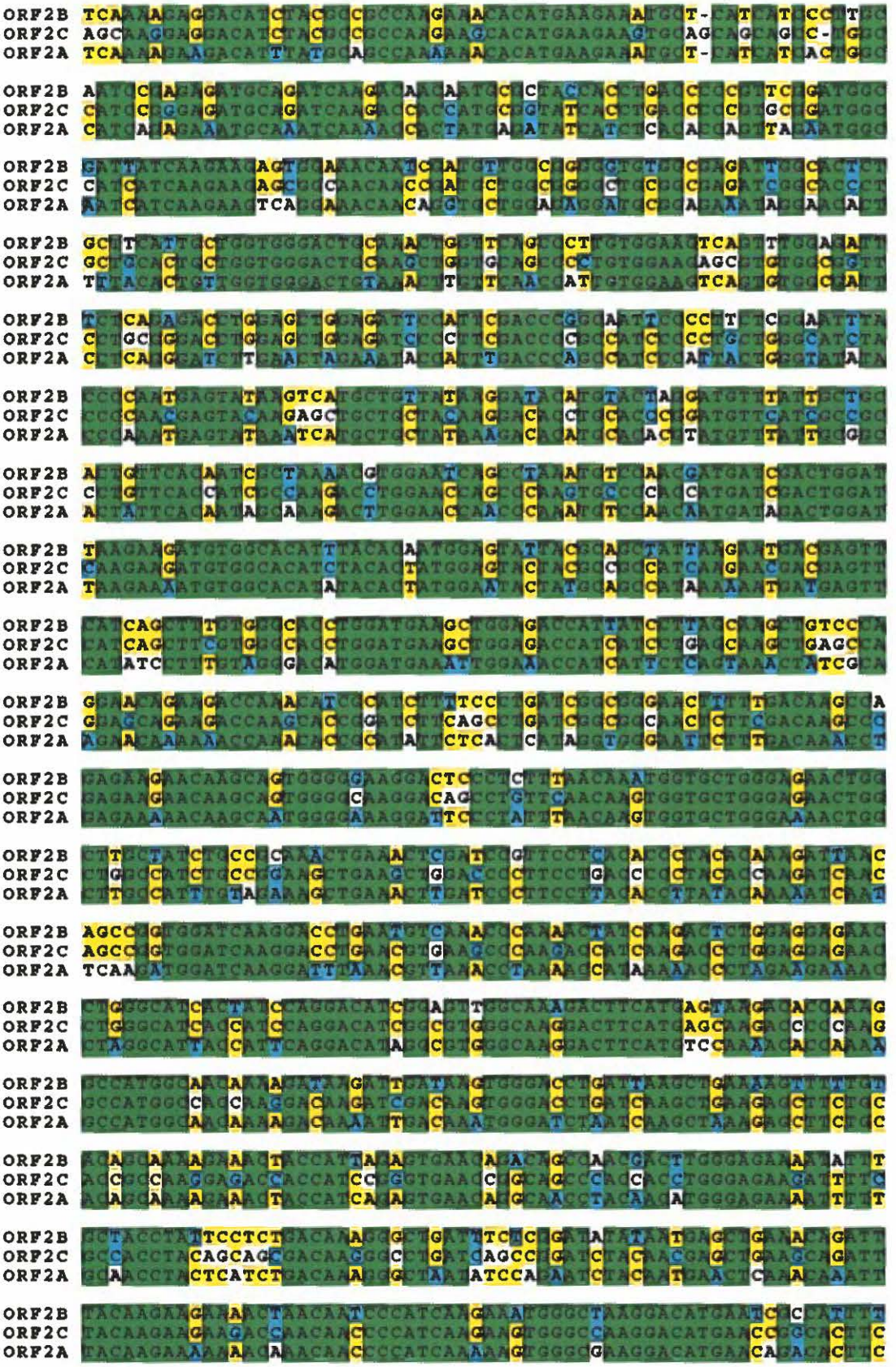

ORF2B
ORF2C
ORF2A

**Figure 17: Alignment of three ORF2s differing in nucleotide sequence**

The three ORF2 sequences (A, B, and C) were aligned with CLUSTALW. Nucleotides that are conserved between all three constructs are highlighted in green. Nucleotides that differ are highlighted in yellow if they match the consensus sequence and teal if they do not. BOXSHADE was used to color the alignment using Workbench 3.2 (workbench.sdsc.edu).

expressing HeLa cells was approximately 12-fold when driven by ORF2 A, and 33-fold and 5-fold when driven by ORF2 B and C respectively when compared to HeLa expressing mutant ORF2 (Figure 18). The ability of a cell, which has adapted to constitutive expression of LINE-1, to repress the activity of other ORF2 sequences differing by up to 22% indicates that the mechanism of repression is not based on the reduction of LINE-1 mRNA through RNAi.

We next sought to determine if the repression on LINE-1 and Alu extended to other elements mobilized by the LINE-1 retrotransposition machinery. U6 and 7SL RNA can be retrotransposed by ORF2 protein (Buzdin *et al*, 2002; Weber, 2006). Additionally, a LINE-1 related synthetic psuedogene consisting of ORF1 tagged with the retrotransposition reporter cassette has been shown to retrotranspose when driven by ORF2 (Alisch *et al*, 2006). Therefore, we used tagged versions of each element to determine if the repression of retrotransposition was specific to LINE-1 and Alu, or if the cells had a global repression of all retrotransposition. Retrotransposition assays in HeLa ORF2 and HeLa ORF2 mutant cells revealed that only the LINE-1-derived ORF1 pseudogene was inhibited. ORF1 was inhibited 27-fold in the HeLa ORF2 cells compared to the ORF2 mutant cells while both U6 and 7SL showed no significant change in retrotransposition (Figure 19).

**Constitutive LINE-1 ORF2 expression results in alterations in DSB repair kinetics**

LINE-1 creates DNA DSBs as an intermediate in the TPRT process (Gasior *et al*, 2006). Cells that are exposed to persistent, low levels of DSBs, alter their DSB repair pathways and cells with altered DSB repair kinetics gain a growth advantage (Delacôte *et al*, 2007). Additionally, several DSB repair pathways have been implicated in the

Figure 18

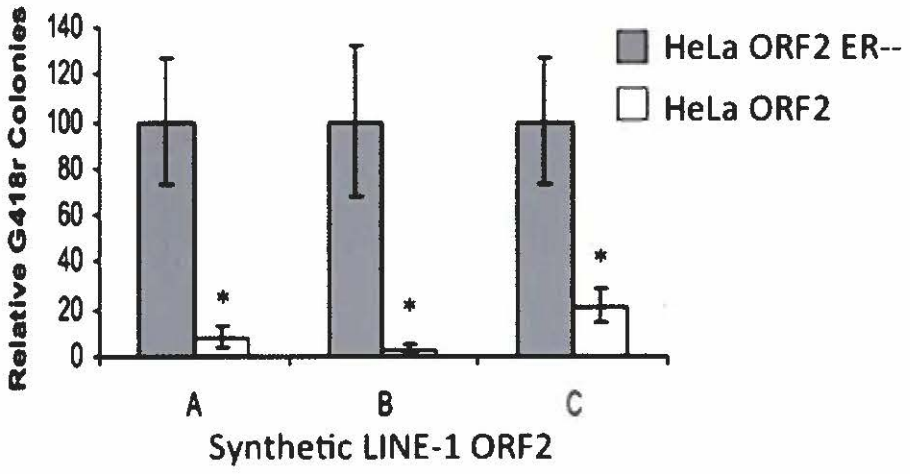



**Figure 18: Repression of retrotransposition is not altered by driving Alu with LINE-1 ORF2s differing in nucleotide sequence**

Three synthetic LINE-1 ORF2 constructs were used to drive Alu retrotransposition. The three constructs vary at the nucleotide level, but their amino acid sequences are identical. Error bars represent standard deviation. Asterisks signify statistical significance ($p<0.05$, two-tailed Student's t-test).

Figure 19

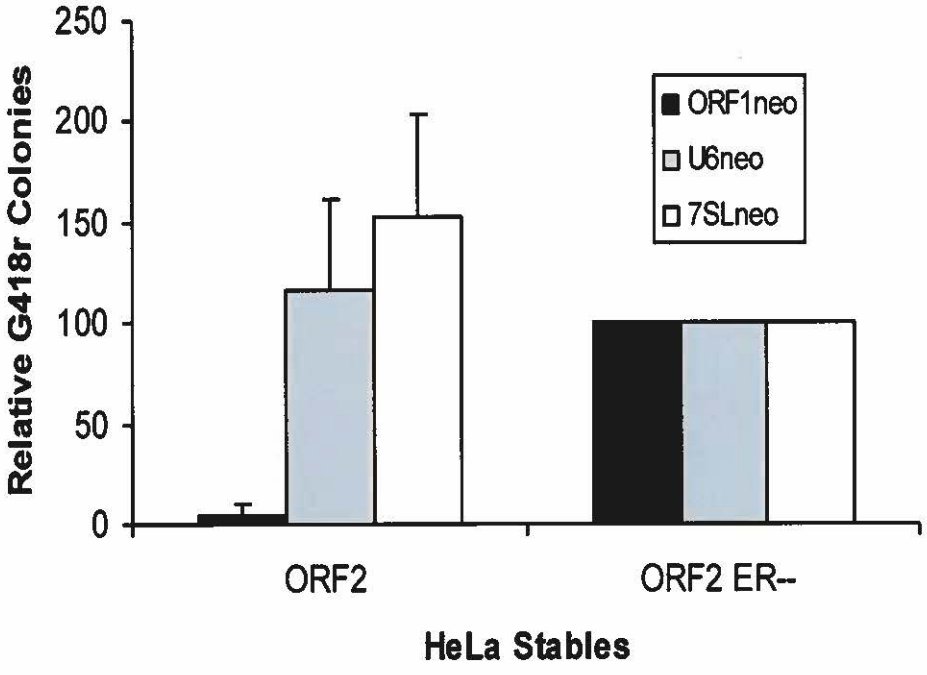



**Figure 19: Persistent LINE-1 expression does not reduce the retrotransposition of all elements mobilized by LINE-1**

Retrotransposition of neomycin reporter tagged 7SL, U6, and an ORF1 pseudogene constructs were measured in HeLa ORF2 and HeLa ORF2 mutant (ER--) cell lines. The retrotransposition of each construct was driven by cotransfection with an ORF2 expression construct. Error bars represent standard deviation.

retrotransposition life cycle (Gasior *et al*, 2008; Gasior *et al*, 2006; Suzuki *et al*, 2009). Given the intimate relationship between LINE-1 and the DNA damage response pathways we wanted to determine if DSBs created by chronic LINE-1 expression could also alter the DNA damage response. To study this we chose to examine the formation and disappearance of 53BP1 foci, an early indicator of DSBs and their repair, after the induction of DSBs by hydrogen peroxide ($H_2O_2$) treatment. Both HeLa ORF2 and HeLa ORF2 mutant cells had similar numbers of 53BP1 foci 45 minutes after treatment, as measure by immunofluorescence. However, 53BP1 foci formation in the HeLa ORF2 cell line returned to background levels after only 75 minutes, whereas the HeLa ORF2 mutant cell line required 135 minutes (Figure 20).

The increase in DSB repair kinetics in $H_2O_2$ treated cells led us to determine if DSBs from LINE-1 would also be repaired with increased kinetics. DSBs arise from a transfected LINE-1 once they have had a chance to be expressed and translated (Kroutter *et al*, 2009). Because of this lag period, we had to alter the timing of the 53BP1 immunofluorescence staining after transient transfection of LINE-1 ORF2 into HeLa ORF2 and HeLa ORF2 mutant cells. At 16 hours post transfection there were very few 53BP1 foci in either cell line. Twenty hours after transfection we observed a large increase in DSBs which returned to near background levels by 48 hours in the HeLa ORF2 cells, but remained high in HeLa ORF2 mutant cells (Figure 21). These data indicate that the increase in DSBs repair potential from persistent LINE-1 expression was not specific to LINE-1-induced breaks, but appeared to be a general increase in the ability of the cell to repair DSBs from multiple sources.

Figure 20

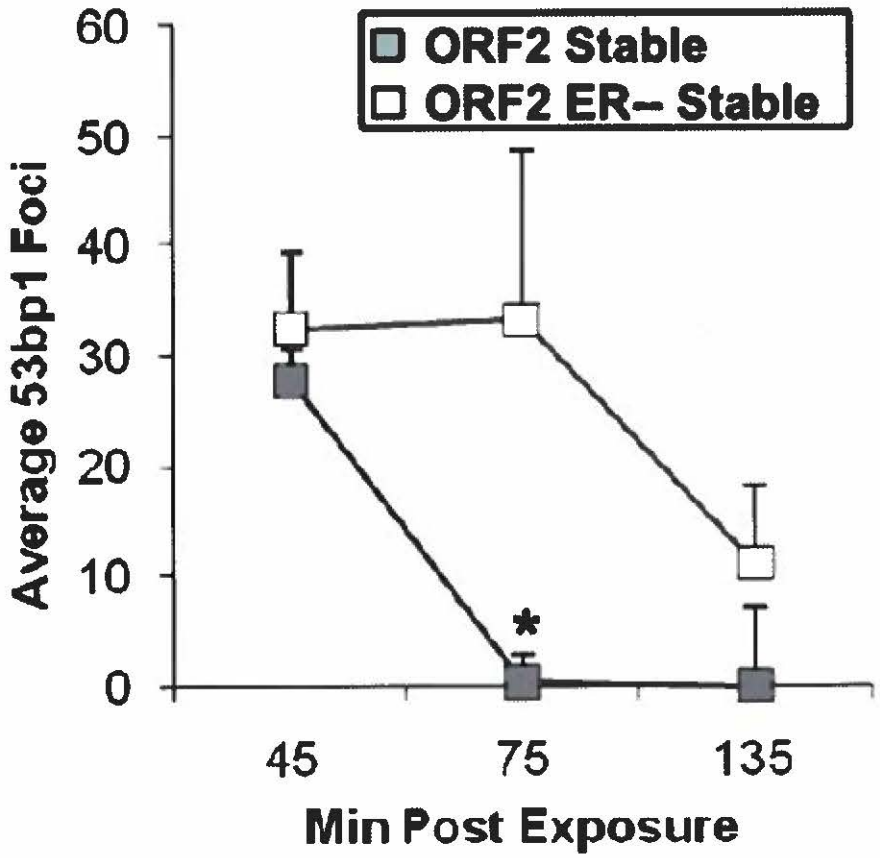



**Figure 20: HeLa cells persistently expressing LINE-1 ORF2 show increased repair of DSBs created by $H_2O_2$**

HeLa cells expressing ORF2 or mutant ORF2 (ER--) were treated with $H_2O_2$ for 10 minutes to create double strand breaks. Cells were fixed and permeabalized at 45, 75, and 135 minutes after exposure. Immunofluorescent 53BP1 foci were visualized to examine the rate of DSB repair. Average number of 53BP1 foci per cell is plotted with error bars representing standard deviation. ($p<0.05$, two-tailed Student's t-test).

Figure 21

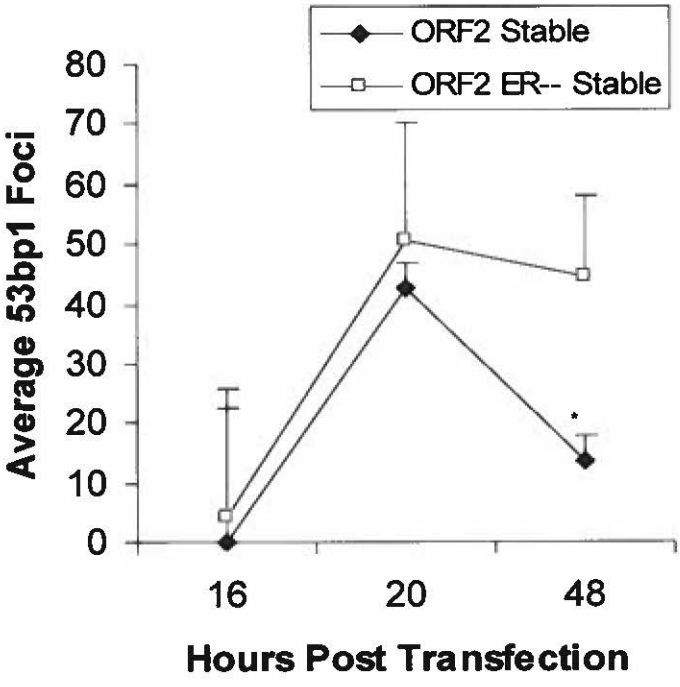



**Figure 21: Altered kinetics of DSB repair after transient LINE-1 expression in HeLa cells that have persistent LINE-1 expression**

At 16, 20, and 48 hours post transfection with ORF2, immunofluorescence-labeled 53BP1 foci were quantified in HeLa ORF2 and HeLa ORF2 mutant cells. The average number of 53BP1 foci per cell is plotted. Error bars represent standard deviation. Asterisks represent statistical significance between the two groups (two-tailed Student's t-test, $p<0.05$).

## Discussion

### Repression of retrotransposition from chronic exposure to LINE-1

Our data indicate that cells have the ability to adapt to persistent expression of LINE-1 elements, specifically ORF2, and that this adaptation is dependent on the dose of LINE-1. The chronic exposure to LINE-1 induced DSBs resulted in a strong repression of both LINE-1 and Alu retrotransposition as well as faster DSB repair kinetics all of which were dependent on the activity of ORF2. This decrease in retrotransposition potential correlated with our observation that the level of expression of full length LINE-1 mRNA in a cell is inversely proportional to its ability to support retrotransposition. Previous studies have found other mechanisms that control mobile elements, including methylation, RNA processing, APOBEC proteins, and the ERCC1/XPF endonuclease (Belancio *et al*, 2006; Gasior *et al*, 2008; Hata & Sakaki, 1997; Muckenfuss *et al*, 2006; Perepelitsa-Belancio & Deininger, 2003). The high copy number of LINE-1 may necessitate the large number of mechanisms to control its deleterious effects, and the wide range of pathways involved most likely reflects the importance of inhibition of LINE-1 in a variety of settings.

The feedback inhibition of LINE-1 serves as a fail-safe mechanism to counter the negative activity of ORF2. If LINE-1 expression is increased, perhaps due to a global hypomethylation, the cell is then able to increase the DSB repair to reduce the toxicity associated with LINE-1 expression. The inhibition of retrotransposition also minimizes the potential of insertional mutagenesis. It is possible that the increase in DSB repair also leads to an increase in mutation from error prone NHEJ.

### Mechanism of repression by chronic LINE-1 expression

We initially postulated that an RNAi mechanism was leading to the decrease in retrotransposition by reducing the amount of full-length LINE-1 mRNA available in the cell. However, this was not the case because Alu retrotransposition was repressed even when driven by ORF2 constructs with altered nucleotide sequences. Additionally, we did not observe the repression of retrotransposition with a LINE-1 sequence that differed by only a few point mutations that inactivated both the endonuclease and reverse transcriptase domains. It is highly unlikely that an RNAi mechanism would not affect such a highly similar sequence.

The most likely mechanism of repression is a response to the DSBs created by the LINE-1 endonuclease. Constitutive expression of LINE-1 is comparable to chronic exposure to DSB inducing chemicals, which results in an increase in DSB repair (Delacôte *et al*, 2007), without altering the sensitivity to IR or changing the growth rate of the cells. We speculate that the repression of retrotransposition stems from the clearing of retrotransposition intermediates before they can come to completion. A similar mechanism of controlling LINE-1 has also been suggested for proteins in the NHEJ pathway (Suzuki *et al*, 2009), where retrotransposition was limited by repairing the DSB before the LINE-1 can finish its insertion, thus inhibiting full-length insertions.

We have shown that constitutive LINE-1 expression induces a cellular change that both decreases the ability of both LINE-1 and Alu to undergo retrotransposition and increases the rate at which DSBs are repaired. This relationship between LINE-1 and the DNA damage response could serve to limit the toxicity caused by ORF2 activity and limit the LINE-1s mutation potential.

**Methods**

**Cell lines and culture conditions**

HeLa, HCT116, and MCF7 cell lines were maintained in minimum essential medium (Gibco) supplemented with 10% fetal bovine serum (Atlanta Biologicals), 1X sodium pyruvate (Invitrogen), and 1X non-essential amino acids (Invitrogen). Cells were maintained in humidified incubator at 37 °C and 5% $CO_2$.

Stable ORF2 expressing cell lines were created by transfecting 1 ug of untagged LINE-1, ORF2, mutant ORF2, or luciferase construct using Lipofectamine with Plus reagent according to manufacturer protocols (Invitrogen). Integration of the plasmid was selected for using 125 μg/mL hygromycin b (Invitrogen). Cells were grown in the presence of hygromycin b until colonies were visible. Colonies were then trypsinized, pooled, and expanded under selection to generate the stable cell lines.

**Northern blots**

The contents of four confluent T75 tissue culture flasks of each cell type were combined and total mRNA was extracted (TRIzol Reagent, Invitrogen). We then performed a chloroform extraction followed by isopropanol precipitation before poly-A selection (PolyATract mRNA Isolation System III, Promega). The poly-A selected RNA was precipitated, resuspended, and fractionated on a 2% agarose-formaldehyde gel.

We transferred RNA to a nylon membrane (Hybond-N; Amersham Biotech) by capillary transfer overnight at room temperature in a standard 5X sodium chloride/sodium citrate (SSC) solution. We crosslinked the RNA to the membrane with ultraviolet light and prehybridized it in 30% formamide, 1X Denhardt's solution, 1% SDS, 1M NaCl, 100 μg/ ml salmon sperm DNA and 100 μg/ml yeast tRNA at 60°C for at least 6 hours. Hybridization with a strand-specific probe was carried out overnight in the

same solution at 60°C. Several washes were done in high stringency conditions (0.1x SSC, 0.1% SDS) at 60°C for 10 minutes. We generated the strand specific probe from the 5' UTR of LINE-1 (MAXIscript T7 system; Ambion Inc.). The membrane was visualized using a Typhoon Phosphoimager (Amersham Biosciences) and quantified with the ImageQuant software (GE Healthcare).

**Retrotransposition assays**

Cells were seeded into T75 tissue culture flasks at a density of $5x10^5$ cells/flask. Transfections were performed the following day using Lipofectamine and Plus reagent (Invitrogen) according to manufacturer protocols using 1 μg of JM101/L1.3 (Drombroski, 1993) for LINE-1 retrotransposition analysis or 2 μg Alu-neo-tet (Dewannieux, Nat Gen, 2003) with 1 μg of the indicated source of ORF2 for Alu retrotransposition analysis. Retrotransposition events were selected for with 400 μg/mL Geneticin (Fisher Scientific) for 14 days. Colonies were then fixed and stained using 2.5% isopropanol, 5% acetic acid, and 0.5% crystal violet. Stained flasks were counted by hand.

**53BP1 foci visualization**

$1x10^4$ cells were plated onto a 96-well imaging plate (BD Falcon) 16 hours before visualization. After treatment cells were fixed with 3.7% formaldehyde in PBS for 10 minutes at room temperature and then permeabilized with 90% cold methanol for 10 minutes. After washing with PBS, cells were blocked with 1% BSA-PBS for 30 minutes at 4 °C and incubated with 2 μg/mL of 53BP1 antibody (Novus Biologicals) diluted in 1% BSA-PBS for one hour. After washing, cells were then incubated with 5 μg/mL Alexa 488 conjugated anti-rabbit secondary antibody (Molecular Probes) in 1% BSA-PBS and 5ug/mL Hoechst dye (Molecular Probes) for one hour at room temperature in

**Colony formation assay**

Cells were seeded into T75 tissue culture flasks at a density of $5x10^5$ cells/flask. The next day, 1 μg of a neomycin resistance plasmid (pIRES-EGFP, BD Biosciences) was transiently transfected (Lipofectamine and Plus reagent, Invitrogen). Cells were grown under 400 μg/mL Geneticin (Fisher Scientific) for 14 days. Colonies were then fixed and stained using 2.5% isopropanol, 5% acetic acid, and 0.5% crystal violet. Stained flasks were counted by hand.

**Ionizing radiation colony formation assay**

After exposure to one gray of ionizing radiation from a cesium-137 source, the colony formation assay described above was conducted.

**Ionizing radiation cellular proliferation assay**

The growth rate of cells exposed to ionizing radiation was performed as described above with minor alterations. The cell growth was measured at seven days after seeding instead of at 1, 2, and 3 days after seeding.

**Vectors and sequences**

The LINE-1 expression vectors used to create the stable cell lines and drive Alu retrotransposition were cloned into pBud CE 4.1 (Invitrogen) by ligation after HinDIII and BamHI digestion. The LINE-1s were driven by the CMV promoter and contained an SV40 polyadenynation signal.

The ORF2 sequence has been optimized to use the maximally efficient codons while retaining the amino acid sequence using the Codon Adaptation Index calculator found at www.evolvingcode.net.

**Statistical analysis**

In all cases, p-values were calculated using the two-tailed Student's t-test to compare experimental values to control values in each experiment using the statistical package in Microsoft Excel. Experiments were routinely done in triplicate with multiple repetitions.

## Chapter 3: Mismatch repair limits retrotransposition

### Introduction

### TPRT can create mismatched bases

The TPRT model of retrotransposition relies on RNA annealing to a genomic DNA sequence with a free 3' hydroxyl group to prime reverse transcription. It is clear from studies of the LINE-1 EN domain that sites other than AA/TTTT are cleaved by this enzyme (Repanas *et al*, 2007). However, the spectrum of sequences capable of being cleaved do not match the sequences actually utilized at insertion sites (Szak *et al*, 2002). The LINE-1 EN tolerates guanine in the target site, but guanine appears to be selected against when insertion sites from the genome are analyzed. We have previously shown that disruptions in an Alu A-tail hindered retrotransposition, especially guanine disruptions (Comeaux *et al*, 2009). It is surprising that guanine was the most disruptive, as G/T mismatches are predicted to be the least disruptive to base pairing (Ke & Wartell, 1993). Deviations from the stretch of thymidine nucleotides would create a mismatch between the RNA and DNA target site, potentially generating a substrate for the mismatch repair machinery. Both the twin priming and microhomology-mediated truncation models of TPRT resolution also predict priming events that can create base/base mismatches and insertion/deletion (indel) loops (Figure 3).

### Mismatch repair

The mismatch repair (MMR) system protects the cell from replication errors, spontaneous mutation, microsatellite instability (MSI) (Parsons *et al*, 1993), and lesions

direction. However, to date no helicases and only a single exonuclease have been found to be involved in eukaryotic MMR.

The MMR-associated exonuclease, ExoI, is a 5' to 3' exonuclease that also displays 5' flap endonuclease (Lee & Wilson, 1999) and RNaseH (Qiu *et al*, 1999) activities. Interaction of ExoI with MSH2 indicated that ExoI has a role in MMR. This interaction was first shown in yeast (Tishkoff *et al*, 1997) and then subsequently in humans (Schmutte *et al*, 1998). Later, ExoI was shown to interact with MLH1, PMS2 and MSH3, further identifying it as a critical part of an MMR-associated higher order complex (Schmutte *et al*, 2001).

With only one 5' to 3' exonuclease found to be participating in MMR reactions, the ability of the reaction to proceed in both directions was perplexing. Genschel *et al.* reported that ExoI mediated excision of a heteroduplex containing a break located 5' to the mismatch required only MutSα. However, when the single strand break was located 3' of the mismatch both MutSα and MutLα were required for resection (Genschel *et al*, 2002). In either case, ExoI was required for resection of the mismatched DNA. It was believed that there was either an undefined 3' to 5' exonuclease associated with MMR that required activation or recruitment by ExoI, or that ExoI contained cryptic 3' to 5' exonuclease activity. This disconnect of polarity would later be clarified by the discovery that the human MutLα complex contains endonuclease activity. This activity was dependent on the presence of a mismatch, MutS, PCNA, RFC, ATP, and a preexisting single strand break. In this model, MutLα cleaves upstream of the mismatch on the strand that already contains a break (Kadyrov *et al*, 2006). This meant that in the case of a preexisting 3' break, the mismatch would then be flanked by single strand

breaks on the same strand. ExoI could then resect the mismatch containing strand in a 5' to 3' direction.

Resection of a DNA strand by ExoI is controlled by interactions with MutS, MutL, and RPA. After binding a mismatch, MutS stimulates the exonuclease activity of ExoI. The creation of long stretches of single stand DNA recruits RPA, which then displaces the ExoI complex from the DNA. After the resected area proceeds past the mismatched DNA, MutS and MutL are no longer recruited to the site, further suppressing ExoI activity (Genschel & Modrich, 2003). This mechanism explains how ExoI can remove hundreds of base pairs on the 5' side of a mismatch, and why the resection terminates within a short stretch of nucleotides on the 3' side. In the case of a single strand break located 3' of the mismatch, PCNA acts to suppress ExoI activity from removing DNA sequence away from the mismatch (Dzantiev *et al*, 2004).

Multiple pathways have been shown to interact with the MMR system. Both of the major DSB repair pathways, HR and NHEJ, are influenced by MMR. During HR, MMR can mediate the recombination reaction by rejecting annealing between substrates that contain mismatches (Elliott & Jasin, 2001). Disruptions in the MMR pathway lead to increases in recombination between heterologous substrates. The MMR components, MutSα, MutSβ, and MutLα, were all shown to directly interact with Werner syndrome, RecQ helicase-like (WRN), with MutS heterodimers stimulating the helicase activity of WRN. The observed stimulation was even greater in the presence of a mismatch in the DNA substrate (Saydam *et al*, 2007). Another RecQ family helicase, RECQ1, was also found to interact with MMR proteins by unwinding mismatched DNA. In addition to the MutS stimulated helicase activity, RECQ1 stimulated the activity of ExoI (Doherty *et al*,

2005). Based on these data, the rejection of HR of divergent sequences by MMR occurs via the action of helicase proteins through the unwinding and possible degradation of the mismatched substrates. The initial annealing step in a TPRT reaction closely resembles that of strand invasion during HR. This structural similarity might be recognized through the same mechanism and, thus, act to limit retrotransposon activity.

In NHEJ, the interaction between MSH6 and Ku70, as well as the accumulation of MSH6 foci colocalized with γ-H2AX suggests a role for MMR in the repair of DSBs (Shahi *et al*, 2011). There is also evidence of direct association between MLH1 and mitotic recombination 11 (MRE11) (Her *et al*, 2002), a component of the Mre11-Rad50-Nbs1 (MRN) protein complex that is involved in NHEJ.

Hong *et al.* has previously presented evidence describing the recruitment of MMR factors to sites of DNA damage. They found that MSH3 and MSH2, in addition to MSH6, were recruited to sites of DNA damage induced by laser micro-irradiation. The recruitment of these proteins was dependent on interactions with PCNA. Additionally, MLH1 was also recruited to sites of DNA damage, but its recruitment was mediated by interaction with MSH2. This group subsequently showed that MSH2 and MSH3 were also recruited to sites of DNA damage caused by ultraviolet radiation. In those experiments, the presence of the nucleotide excision repair factor xeroderma pigmentosum, complementation group A (XPA) was necessary for MSH2 and MSH3 foci formation (Hong *et al*, 2008).

MMR is also thought to be involved in initiating the signaling of downstream effectors when DNA damage is present, triggering both the arrest of the cell cycle and apoptosis (Wu *et al*, 1999; Zhang *et al*, 1999). The mechanism of action is yet unknown,

but it is thought to signal through the ATM, ataxia-telangiectasia and rad3-related (ATR) kinase pathways and have both p53-dependent and p53-independent branches (Toft *et al*, 1999).

There is a clear connection between DNA repair pathways and the control of mobile elements (Gasior *et al*, 2008; Gasior *et al*, 2006; Suzuki *et al*, 2009). Additionally, there are multiple steps in the TPRT reaction that have the potential to create mismatches. Using isogenic cell lines that differ in their expression of MLH1, and ability to perform mismatch repair by extension, we show that the MMR pathway can regulate mobile element activity. By rescuing *de novo* insertion events we found that the repression occurs at the initial priming step of TPRT, therefore changing the frequency of non-thymidine bases found at insertion sites. Finally, using oligonucleotides that mimic preinsertion complexes, we demonstrate that the mismatch repair sensor protein MSH2 binds at DNA/RNA duplex sites containing a mismatch, further implicating MMR in the control of retroelements.

**Results**

**A-tail disruptions at the 3' end of Alu inhibit retrotransposition**

We have previously shown that Alu elements with disrupted A-tails retrotranspose at a lower rate than Alus with perfect, homogenous A-tails (Comeaux *et al*, 2009). New Alu insertions preferentially prime near the 3' terminus of the RNA (Wagstaff, submitted), thus disruptions near the 3' end will create mismatches between the RNA and DNA target site. By inserting the sequence $(GAA)_4G$ near the 5' end, in the middle, and near the 3' end of the A-tail (Figure 23) we wished to determine if the location of the disruption could affect Alu retrotransposition. As the disruption was

Figure 23

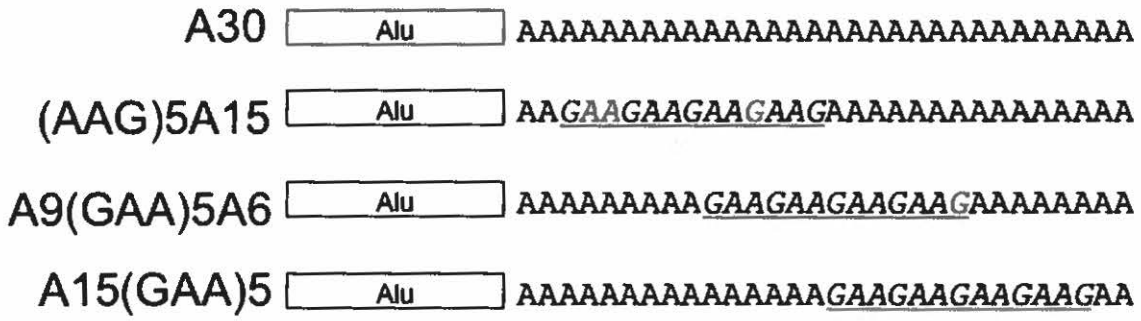



**Figure 23: Disruptions in an Alu A-tail**

To determine if the position of disruptions has an affect on Alu retrotransposition we created Alu expression constructs with disruptions in the 5', middle, and 3' end of the A-tail. Areas of heterogeneity are underlined.

moved further toward the 3' end of the RNA, there was a decrease in the rate of Alu retrotransposition in HeLa cells. Having a disruption near the 5' end of the A-tail reduced retrotransposition 2.4-fold (+/- 0.13) compared to an Alu with an A-tail consisting of 30 adenines. Disruptions in the middle and at the 3' end of the A-tail reduced retrotransposition 10.5-fold (+/- 0.01) and 8.6-fold (+/- .06) respectively (Figure 24).

**MLH1 deficiency in HCT116 colon cancer cells allows for increased LINE-1 retrotransposition**

HCT116 cells are a colon cancer cell line that is naturally deficient in MLH1 (as well as the binding partner PMS2) (Papadopoulos *et al*, 1994). These cells have been shown to exhibit microsatellite instability and an increased mutation rate as well as an inability to carry out mismatch repair (Parsons *et al*, 1993). Two separate pairs of isogenic cells lines were used to examine the effect of MLH1 on LINE-1 retrotransposition in HCT116 cells. HCT116+MLH1 cells have been complemented with MLH1 cDNA and show a 4.3-fold (+/- 0.15) decrease in the LINE-1 retrotransposition potential of the HCT116+Vector control cells ($p=0.0009$) (Figure 25). A western blot probing for MLH1p confirms that the HCT116+Vector cell line does not express detectable levels of MLH1 whereas the HCT116+MLH1 cell line does express MLH1 protein (Figure 26).

Because MMR proteins are known to interact with and signal checkpoints (Wu *et al*, 1999; Zhang *et al*, 1999), we wished to assess the relative colony forming potential of each cell line. We measured colony formation after transfection with a G418 resistance

Figure 24

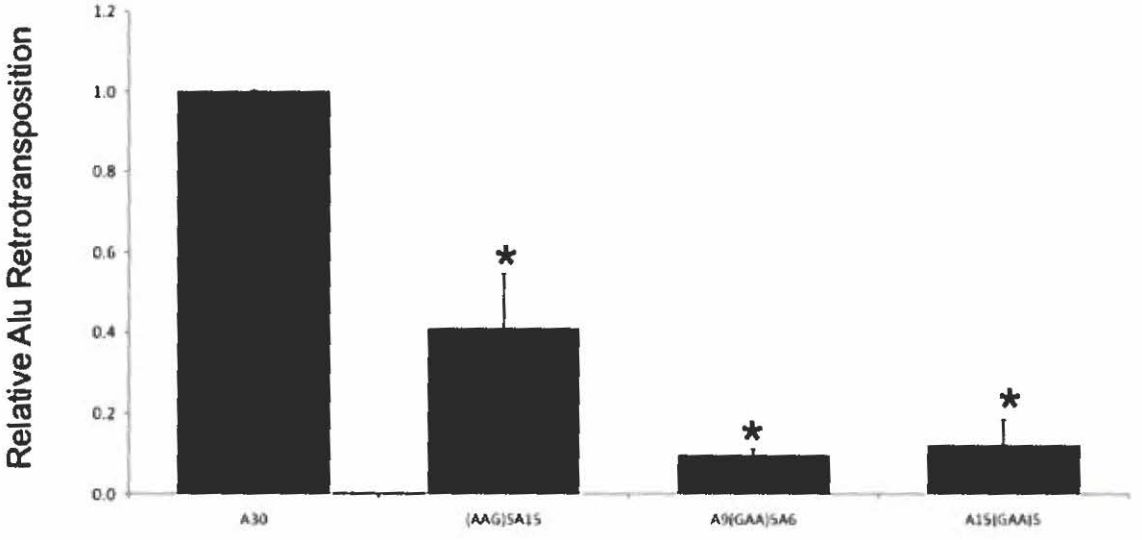



**Figure 24: Disruptions near the 3' end of an Alu A-tail are disruptive to Alu retrotransposition**

HeLa cells were seeded and cotransfected with the indicated Alu retrotransposition construct tagged with the neomycin reporter gene and an ORF2 expression construct. Cells harboring a retrotransposition event are selected for using G418. After 14 days, colonies are stained with crystal violet and the colonies are counted. Error bars represent standard deviation. Asterisks represent statistical significance ($p<0.05$, two-tailed Student's t-test).

Figure 25

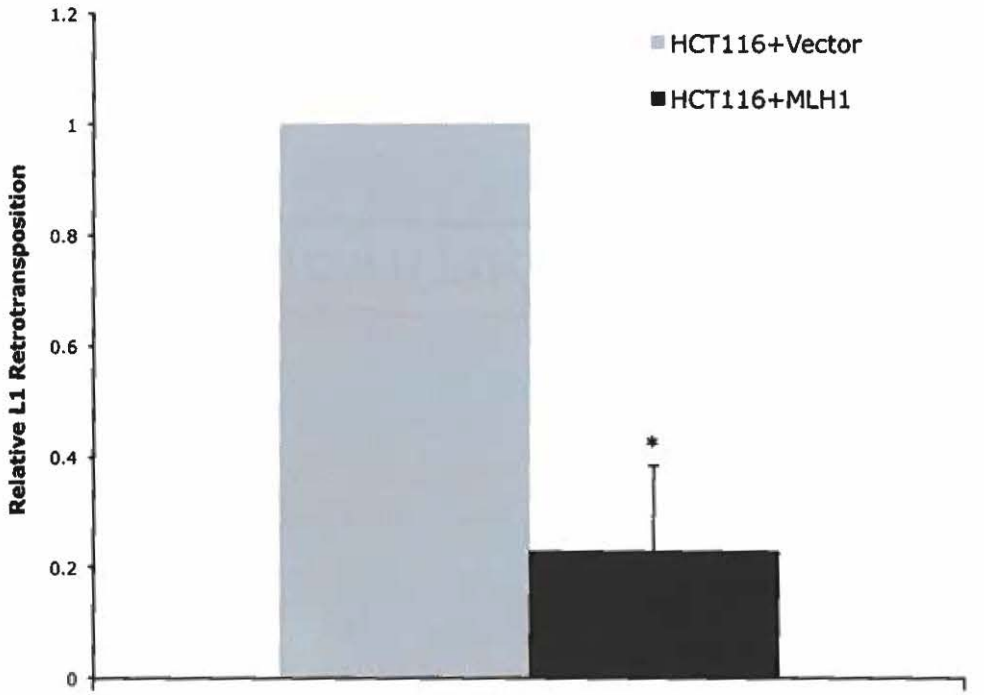



**Figure 25: LINE-1 retrotransposition is reduced in HCT116 cells complemented with MLH1 cDNA**

Cells were seeded and transfected with LINE1.3CMV-neo. After selection for 14 days with G418 the resulting colonies were stained with crystal violet and counted. Error bars represent standard deviation. Astericks represent statistical significance ($p<0.05$, two-tailed Student's t-test).

Figure 26

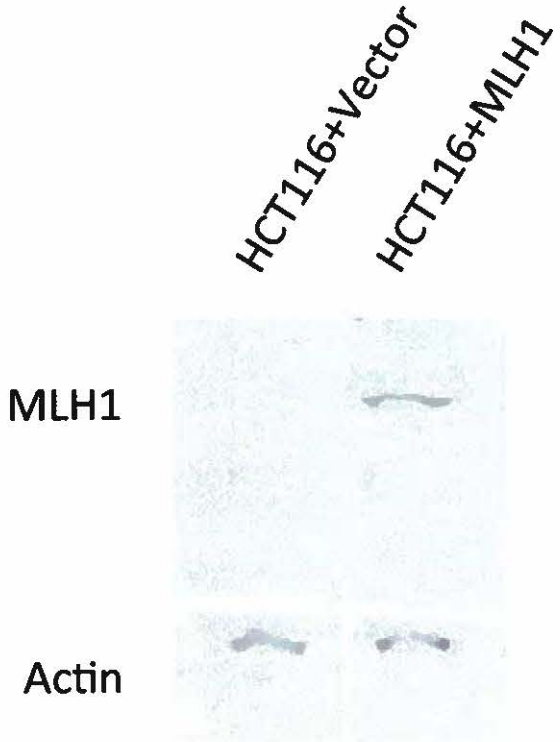



**Figure 26: Only HCT116+MLH1 cells express MLH1**

Whole cell extract from approximately $1x10^6$ cells was collected in Laemmli buffer. 30 uL of extract was subjected to Western blot analysis of MLH1 and Actin, which served as a loading control, using Alexa-fluor conjugated secondary antibodies and the Li-Cor Odyssey Pro IR scanner.

Figure 27

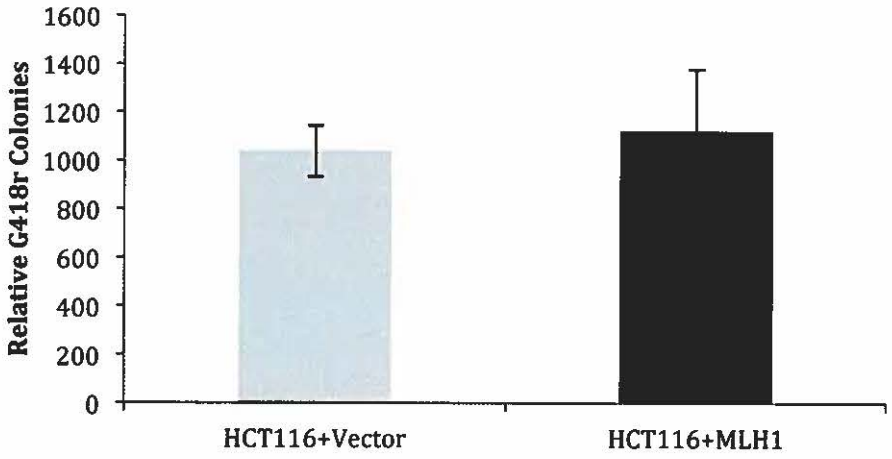



**Figure 27: Colony formation in HCT116+Vector and HCT116+MLH1 cell lines**

Colony formation was measured by transfecting pIRES, which contains a neomycin resistance gene. After transfection, cells containing the plasmid were selected for using G418 for 14 days. Cell were fixed, stained with crystal violet, and counted. Error bars represent standard deviation.

construct (Figure 27). On average the HCT16+Vector cell line formed 1040 +/- 106 colonies while the HCT116+MLH1 cell line formed 1121 +/- 256 (p=0.39).

Transient, high-level expression of LINE-1 is toxic to cells (Wallace *et al*, 2008). Due to this fact we sought to determine if the reduction in LINE-1 retrotransposition could be attributed to an increase in cell death after introduction of LINE-1. By cotransfecting a LINE-1 construct with a neomycin containing plasmid, we assessed the LINE-1-related toxicity between the MLH1 positive and negative cell lines. LINE-1 was slightly more toxic to HCT116+Vector cells (Figure 28). An average of 277 +/- 39 colonies formed in the HCT116+Vector cells whereas the HCT116+MLH1 cell line formed and average of 417 +/- 76 colonies. These results indicated that the lack of MMR did not lead to an increased sensitivity to LINE-1-induced toxicity and that the observed decrease in retrotransposition could not be explained by cell death. In fact, the increased toxicity in HCT116+Vector cells meant the level of repression in HCT116+MLH1 cells was actually underestimated.

To confirm the HCT116+MLH1 cell line was in fact exhibiting a MMR positive phenotype, we utilized a previously published microsatellite instability (MSI) assay (Vo *et al*, 2005). This assay takes advantage of the instability of dinucleotide repeats in MMR negative cells by fusing an out of frame GFP coding region with a (CA)17 dinucleotide repeat preceding it. If the cells exhibit no MSI the GFP coding region remains out of frame and is not expressed. A mutation in the microsatellite restores the reading frame and leads to GFP expression. The cells containing MLH1 exhibited less MSI than the vector-containing cells (75 +/- 18 vs 266 +/- 33) (Figure 29).

Figure 28

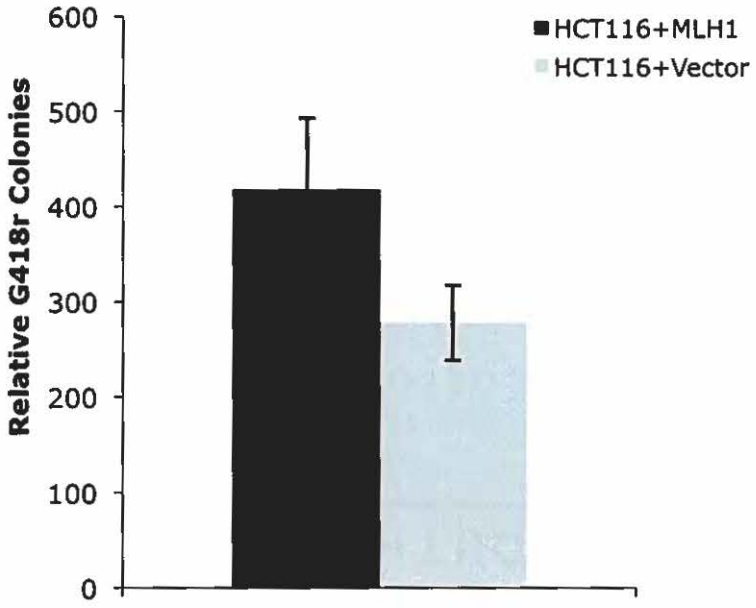
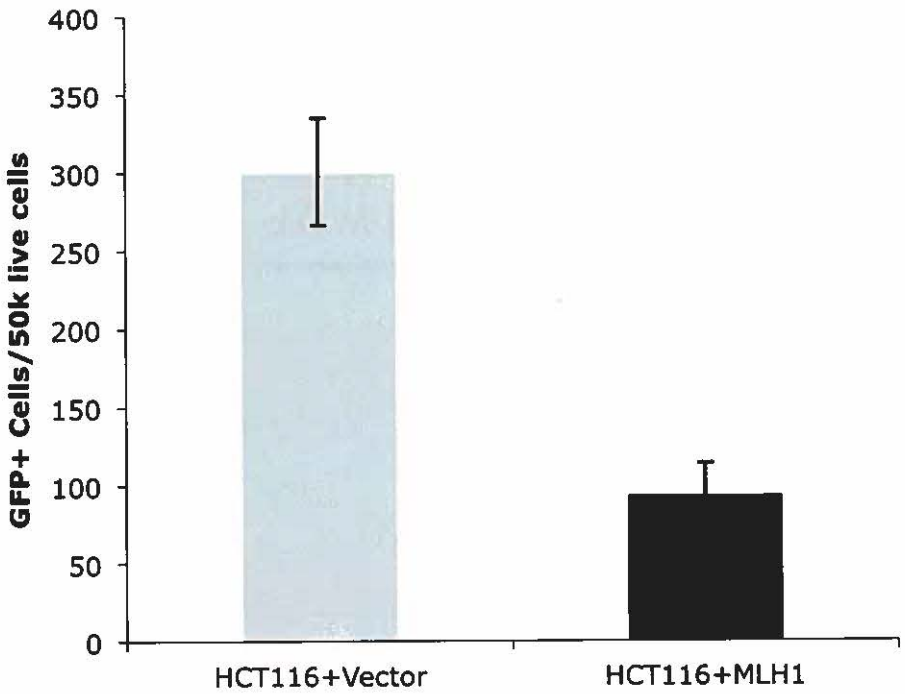



**Figure 28: LINE-1 toxicity is similar in HCT116+Vector cells and HCT116+MLH1 cells**

Untagged LINE-1 was cotransfected with pIRES. After transfection, cells that integrated the vector and survived the LINE-1 mediated toxicity were selected for using G418 for 14 days. Cells are fixed, stained with crytal violet and counted to determine relative toxicity of LINE-1. Error bars are standard deviation.

'igure 29




**'igure 29: HCT116+MLH1 cells display reduced MSI**

Microsatellite instability was measured by transfection of an out of frame GFP :onstruct preceded by a (CA)17 dinucleotide repeat. Mutations that placed GFP back nto frame lead to fluorescence. Error bars are standard deviation.

It has been reported that overexpression of MLH1 can lead to a dominant negative effect and increased mutation frequency (Shcherbakova *et al*, 2001), ultimately leading to apoptosis (Zhang *et al*, 1999). To rule out the possibility of increased mutation frequency leading to a secondary mutation in the cDNA complemented cell lines controlling the decrease in LINE-1 retrotransposition, we performed the assay in HCT116 cells that have been complemented with chromosome 3 (Koi *et al*, 1994). The expression of chromosome 3, which contains MLH1 and its endogenously associated regulatory sequences, most likely results in normal expression of MLH1 and not overexpression. Because this cell line is resistant to neomycin, we could not use the standard tagged LINE-1 construct. Instead we utilized a LINE-1 construct tagged with blasticidin in the same manner as the neomycin tag. Correcting the MLH1 deficiency using a chromosome led to a 3.3-fold (+/- 0.16) decrease in the frequency of LINE-1 retrotransposition ($p=2.08x10-6$) (Figure 30).

**MEFs deficient in MLH1 show increased retrotransposition**

To ensure that the observed difference in retrotransposition was not due to clonal selection in the HCT116 cell lines, we further tested the effect of MLH1 deficiency in an isogenic pair of mouse embryonic fibroblast (MEF) cell lines. Both copies of MLH1 have been knocked out by gene targeting in the MEF cell line MC2 (Prolla *et al*, 1998). The complementary cell line MC2+WT contains a full-length cDNA copy of the hMLH1 gene (Buermeyer *et al*, 1999). The MEF cell lines yielded poor colony formation when using the standard drug-resistance tagged LINE-1 retrotransposon vectors. To investigate the LINE-1 retrotransposition rate in the MEF, we utilized a LINE-1 construct tagged with EGFP followed by FACS analysis at six days post transfection. The cells were

Figure 30

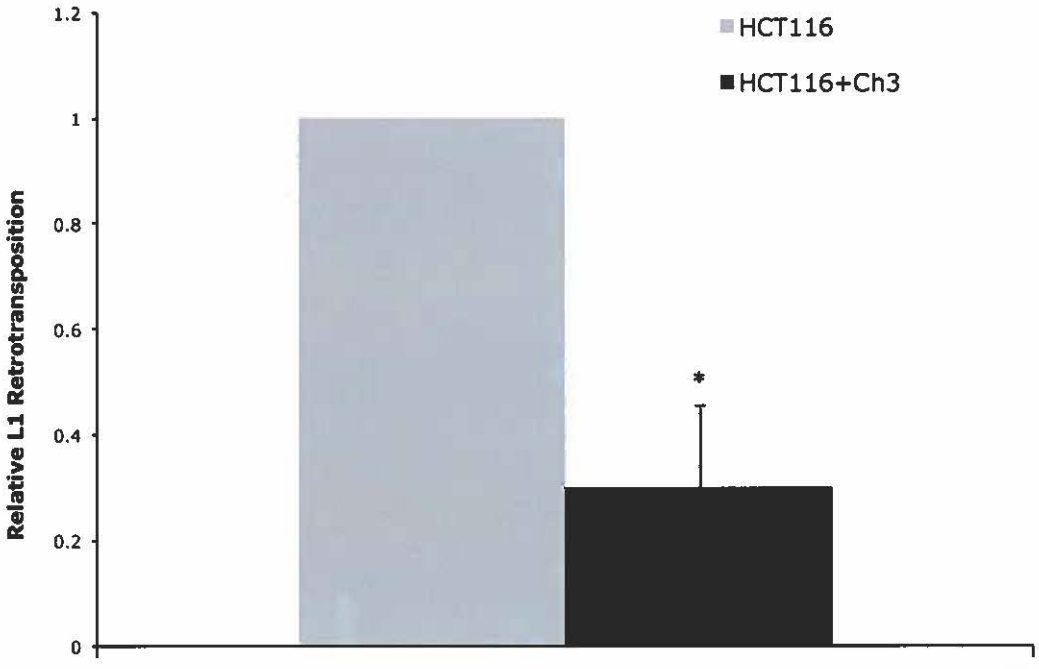



**Figure 30: Retrotransposition is reduced in HCT116 cells complemented with chromosome 3**

Uncomplemented HCT116 cells or HCT116 cells complemented with chromosome 3 (HCT116+Ch3) were transfected with a LINE-1 construct tagged with a blasticidin reporter cassette. After 14 days of selection the colonies were stained and counted. Error bars represent standard deviation. An asterisk signifies statistical significance ($p<0.05$, two-tailed Student's t-test).

naintained under puromycin selection after transfection to select for the presence of the etrotransposition vector. At six days post transfection we observed a 3.5-fold (+/- 0.3) eduction in EGFP positive cells in the MEFs that expressed MLH1 compared to cells hat did not express MLH1 (p=0.025) (Figure 31).

**ILH1 expression can be modulated in 293T-Lα cells**

The 293T-Lα cell line contains the MLH1 gene under the control of a etracycline-regulated promoter (Cejka *et al*, 2003). In the presence of tetracycline ILH1 gene expression is turned off. By maintaining specific amounts of doxycycline in he media, the level of MLH1 protein can be tightly regulated. We maintained cells in ;0, 0.8, 0.2 and 0.0 ng/mL doxycycline. These levels resulted in a gradient of MLH1 rotein expression (<1%, 4%, 21%, and 100% expression respectively) (Figure 32, igure 33).

**ILH1 inhibits L1 retrotransposition in a dose-dependant manner**

Because this cell line is resistant to G418, the LINE-1 retrotransposition rate xperiments were carried out using a LINE-1 construct tagged with reporter cassette that rants resistance to blasticidin upon successful retrotransposition. In the absence of loxycycline the cells fully express MLH1 protein and LINE-1 retrotransposition is nhibited 2.6-fold (+/- 0.04). The cells cultured in the presence of 0.2 ng/mL of loxycycline showed a 2.3-fold (+/- 0.02) reduction in LINE-1 retrotransposition otential. Culturing cells in 0.8 ng/mL doxycycline led to only a slight 1.3-fold (+/- 0.15) lecrease in retrotransposition as compared to the cells treated with 50 ng/mL of loxycycline in which MLH1 expression was undetectable (Figure 34).

Figure 31

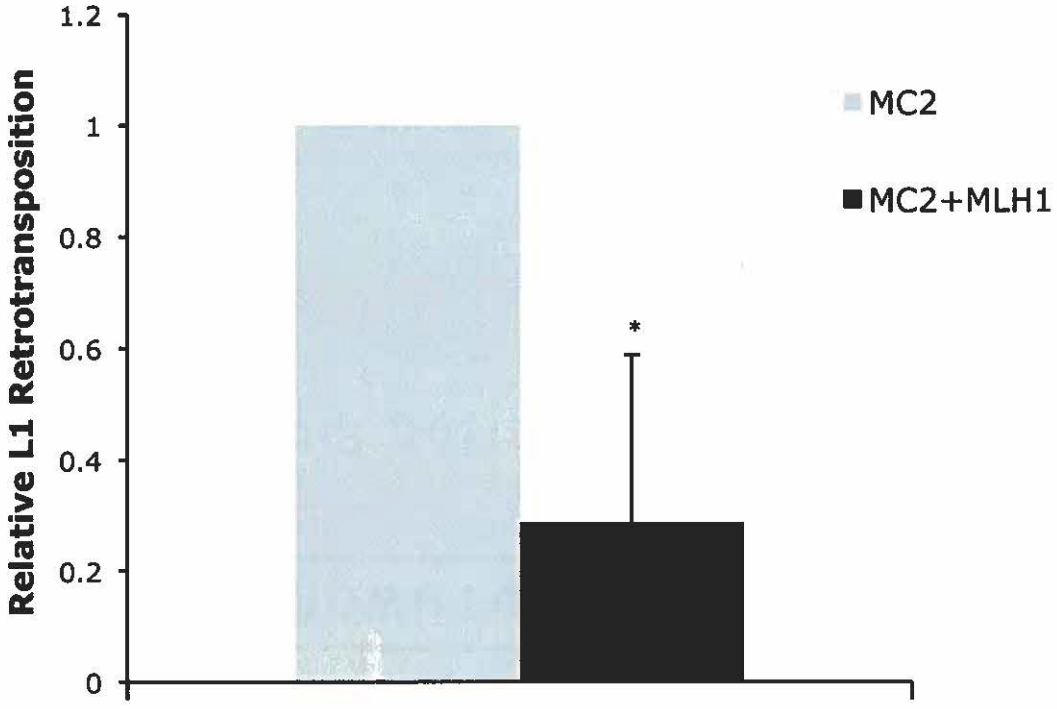



**Figure 31: MEF complemented with hMLH1 show decreased LINE-1 retrotransposition**

Retrotransposition was measured using LINE-1 tagged with EGFP in MLH1 negative (MC2) and MLH1 complemented (MC2+MLH1) MEFS. Cells not containing the vector were selected against for 6 days. Cells were then subjected to FACS analysis to determine LINE-1 retrotransposition frequency. Error bar represents standard deviation. Asterisk signifies statistical significance ($p<0.05$, two-tailed Student's t-test).

Figure 32

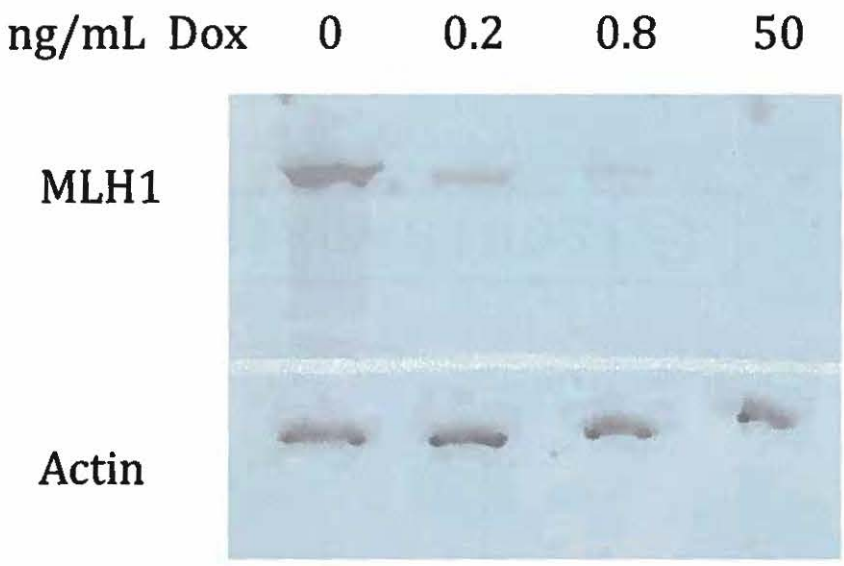



**Figure 32: Western blot of doxycycline regulation of MLH1 in 293T-Lα**

293T-Lα cells were cultured in the presence of the indicated amount of doxycycline (dox). Whole cell extracts were loaded on a polyacrylamide gel and probed for MLH1, with actin serving as a loading control.

Figure 33

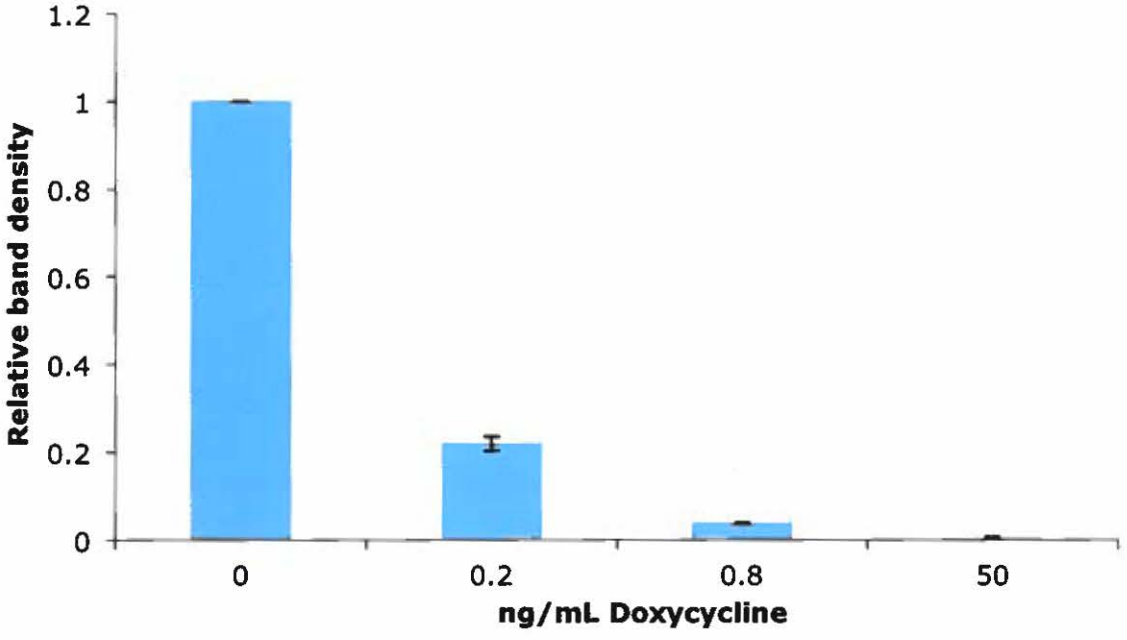



**Figure 33: Quantification of MLH1 expression in 293T-Lα with doxycycline**

MLH1 Western blot images from 293T-Lα doxycycline gradient cell lines were quantified using ImageJ. The untreated cell line was set to a relative density of 1 and actin was used to normalize signals across all blots. Error bars are standard deviation.

Figure 34

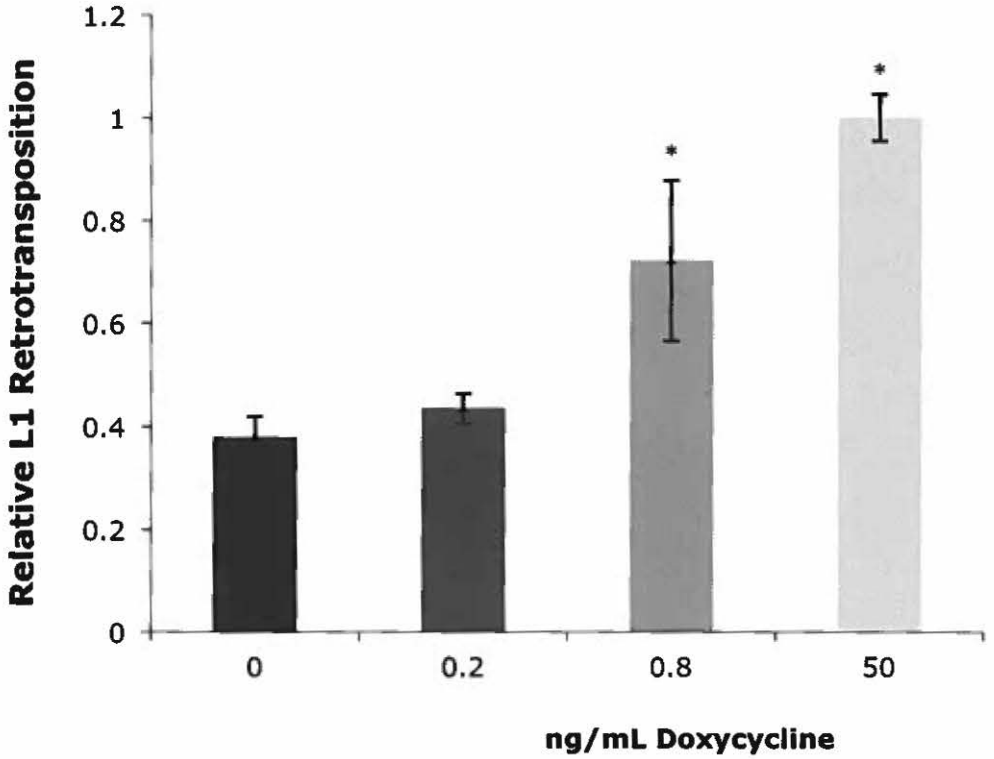



**Figure 34: LINE-1 retrotransposition in 293T-Lα cells after doxycycline treatment**

293T-Lα cells were cultured in the indicated amount of doxycycline for 8 days to modulate MLH1 expression. Retrotransposition was measured with LINE-1 tagged with a blasticidin reporter cassette. Cells were reseeded into blasticicin-containing medium after transfection to select for retrotransposition events until colonies formed. Colonies were stained with crystal violet and counted. Error bars represent standard deviation. Asterisks represent statistical significance when compared to the 0 ng ($p<0.05$, one-way ANOVA, Tukey's multiple comparison test).

**Characteristics of rescued LINE-1 insertions**

The inclusion of a bacterial origin of replication and a dual prokaryotic/eukaryotic promoter on the neomycin resistance gene in the 3' UTR of the LINE-1 construct allows the rescue of new insertions. The genomic DNA can be digested with an enzyme that does not cut within the LINE-1 sequence and then religated to create a functional plasmid. Rescuing *de novo* insertions allows us to examine several different characteristics of new LINE-1 insertions. By taking advantage of islands of differential sequence between the synthetic LINE-1 rescue construct and endogenous LINE-1 we can design sequencing primers along the entire length of the element to locate the 5' end of the insertions. To sequence the 3' end of the insertion the data from the 5' end is used to estimate the probable genomic insertion site and a primer is designed to flank it.

Studies of the LINE-1 endonuclease domain have show that many different thymidine-rich target sites are capable of being cleaved (Repanas *et al*, 2007) (Figure 35). However, the cleavage sites utilized for insertion do not match the spectrum of possible LINE-1 endonuclease cleavage sites, with guanine in the target site being underrepresented at insertion loci in both endogenous insertions (Symer *et al*, 2002; Szak *et al*, 2002) (Figure 36) and tagged insertions (Gilbert *et al*, 2005) (Figure 37). To determine if mismatches could be leading to the repression of LINE-1 insertions by the MMR machinery we rescued (Gilbert *et al*, 2005) *de novo* insertions from the genomes of HCT116+MLH1 and HCT116+Vector cells as well as analyzed a previously published data set of LINE-1 insertions rescued from HCT116 cells (Symer *et al*, 2002). If MMR is capable of recognizing and removing insertion initiation complexes that contain a

Figure 35

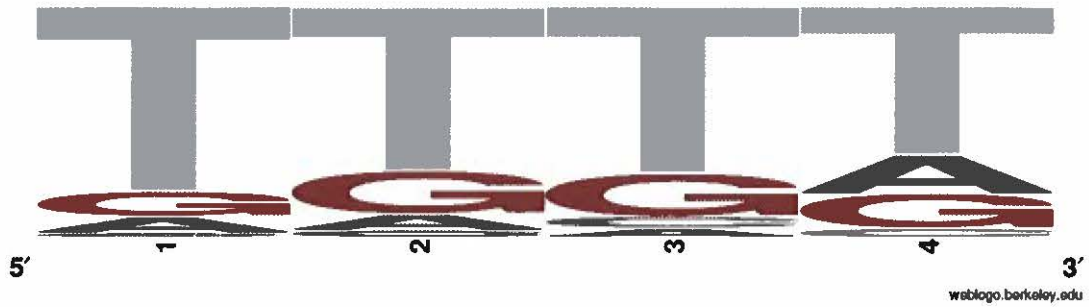



**Figure 35: LINE-1 endonuclease cleavage site preference**

A WebLogo (weblogo.berkley.edu) of the published (Repanas *et al*, 2007) preference of the LINE-1 endonuclease. Only the four bases that comprise the priming site are shown. The frequency of each nucleotide at each position is represented by its height. Guanine residues are shown in red.

Figure 36

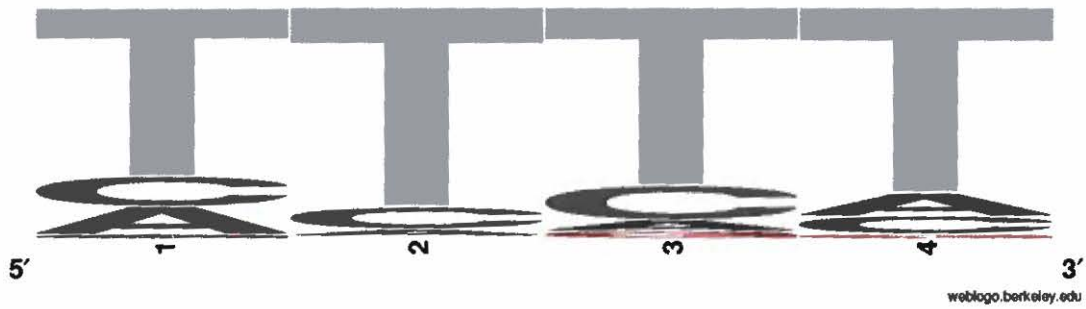



**Figure 36: Endogenous LINE-1 insertion site preference**

A frequency WebLogo (weblogo.berkley.edu) was created using TSD data from 1794 LINE-1 insertion sites (Szak *et al*, 2002). Only the four bases thought to be involved in priming are shown. Guanine residues are colored red.

Figure 37

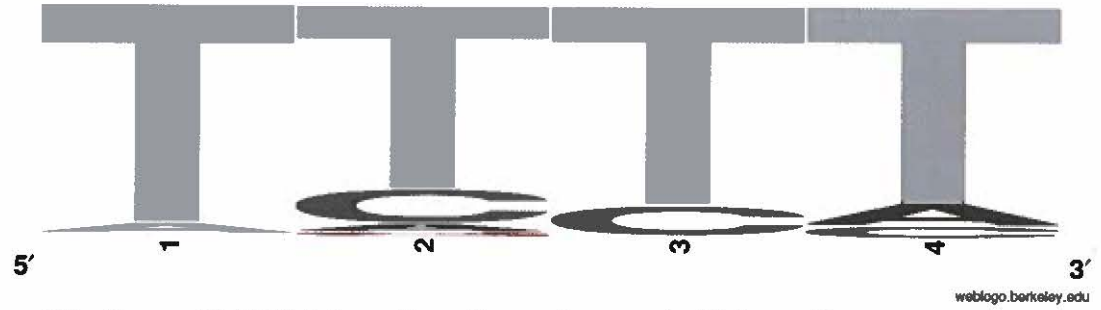



**Figure 37: Tagged LINE-1 insertion site preference in HeLa cells**

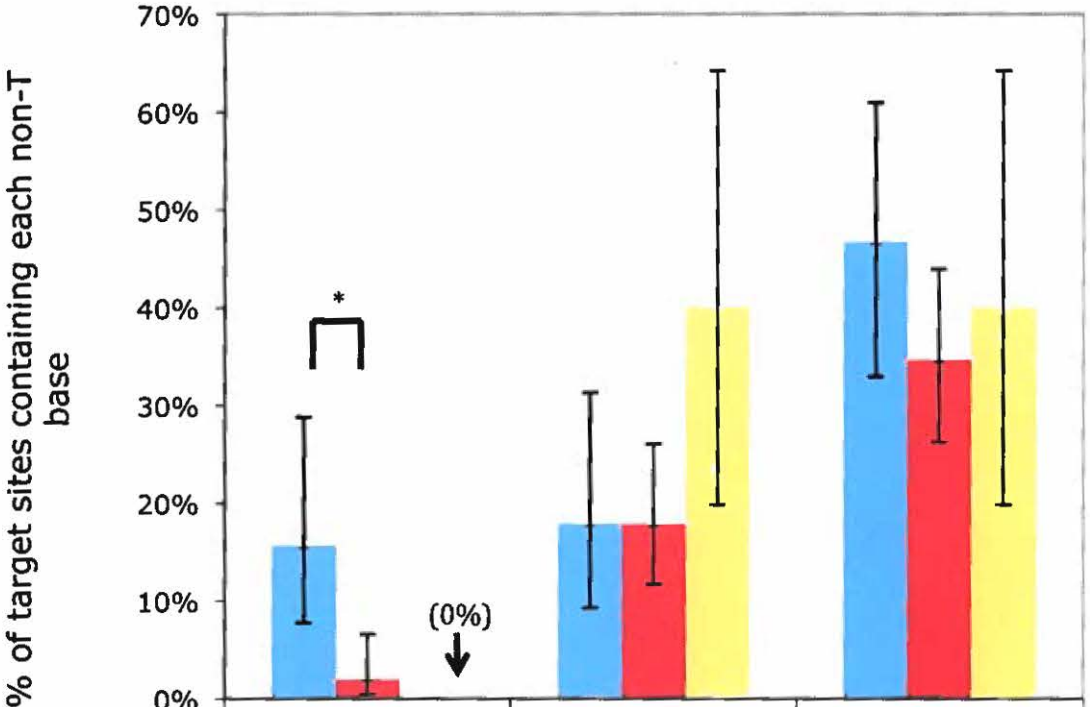
Data from tagged LINE-1 insertion sites were mined (Gilbert *et al*, 2005) to determine the priming site. A frequency WebLogo of the four priming bases was created using weblogo.berkely.edu. Guanine residues are colored red.

mismatch, rescues from cell lines that have functional MMR should show a decrease in the proportion of deleterious mismatches at the priming site.

LINE-1 insertion sites rescued from MMR negative HCT116 and HCT116+Vector cells exhibited an increased proportion of insertion sites containing a guanine when compared to insertion sites rescued from MMR competent HeLa cells (Gilbert *et al*, 2005) or the complemented HCT116+MLH1 cells (7/45 vs 2/107 vs 0/15 respectively). The frequency of other non-thymidine bases found within insertion sites did not differ significantly between the two HCT116 cell lines or from what was previously reported in tagged LINE-1 insertions from HeLa cells (Figure 38).

Sequencing both the 5' and 3' junctions of the insertions also allowed us to measure total insert length, A-tail length, target site duplication (TSD) size, and the proportion of inserts that are full-length, inverted, or truncated. The average insert length did not differ significantly between the HCT116+Vector and HCT116+MLH1 cells (1495 +/- 1785 and 1356 +/- 1063 respectively) (Figure 39). The lengths of the HCT116 inserts did also not differ significantly from insertions previously rescued in HeLa cells (1092 +/- 1643). The length of the A-tails of new inserts was also not significantly different between HCT116+Vector, HCT116+MLH1, and HeLa (72 +/- 40, 61 +/- 45, and 66 +/- 32) (Figure 40). TSD size was also similar in each cell line. TSDs from HCT116+Vector cells averaged 8 +/- 7 base pairs. HCT116+MLH1 TSDs were 6 +/- 7 base pairs (Figure 41). Lastly, there was no appreciable change in the proportion of truncated, full-length, or inverted elements (Table 1). Insertions from both HCT116+Vector and HCT116+MLH1 exhibited all of the hallmarks of standard LINE-1 retrotransposition event with the exception of an increase in guanine in the target site.

Figure 38




| | G | A | C |
|---|---|---|---|
| HCT116+Vector | 16% | 18% | 47% |
| HeLa | 2% | 18% | 35% |
| HCT116+MLH1 | 0% | 40% | 40% |

**Figure 38: Analysis of non-T bases in rescued target sites**

LINE-1 retrotransposition events were rescued and sequenced. The insertion site was determined by examining the filled and empty genomic sites. The frequency of each non-T base is plotted independently. Error bars represent 95% confidence intervals around each proportion. An asterisk represents statistical significance ($p<0.05$, Fisher's exact test).

Figure 39

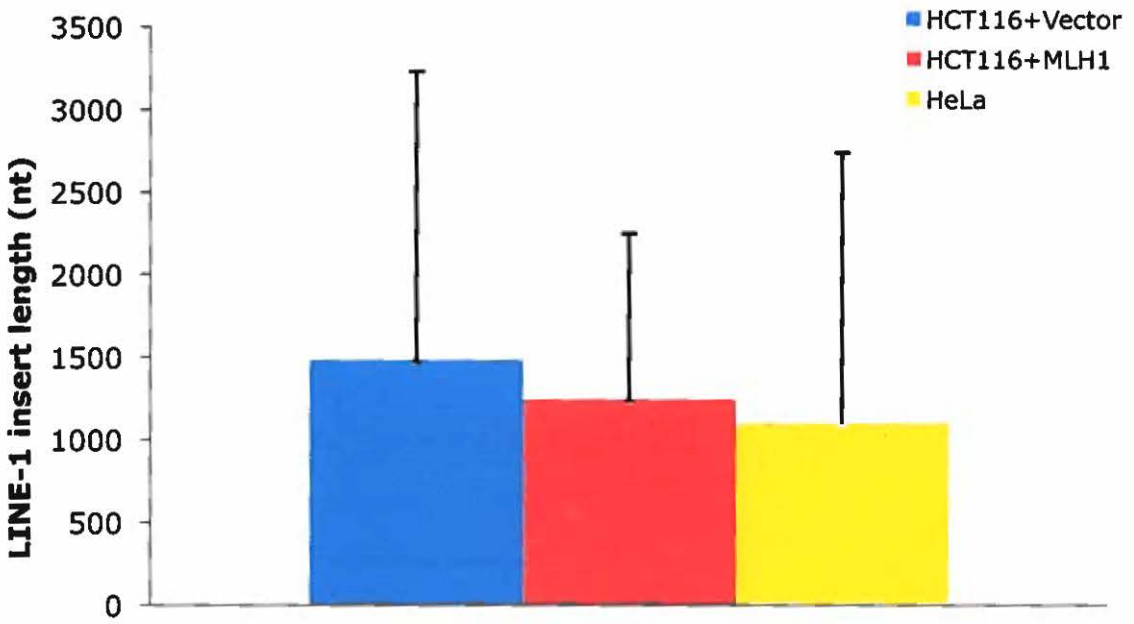



**Figure 39: Amount of LINE-1 sequence retrotranscribed**

The length of each rescued insertion was calculated by subtracting the length of the neomycin resistance gene and bacterial origin of replication. The resulting length represents just LINE-1 sequence. Error bars are standard deviation.

Figure 40

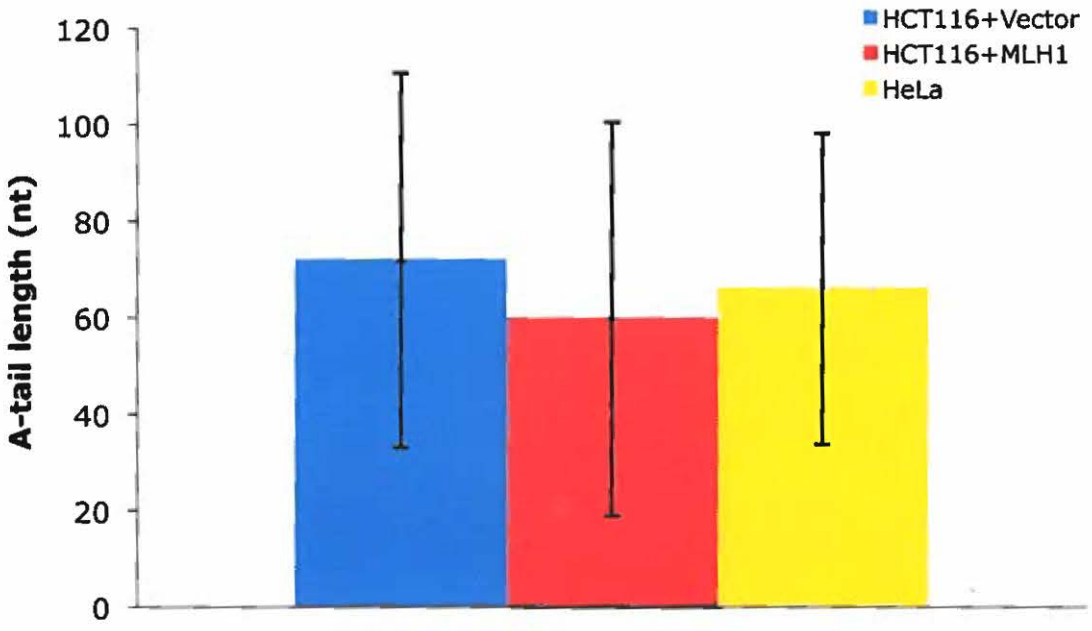



**Figure 40: A-tail length of rescued insertions**

The length of the A-tail was determined from either sequencing out from near the LINE-1 polyadenylation signal or from sequencing in from genomic sequence 3' of the insertion site. An A-tail was defined as the longest stretch of uninterrupted adenine residues. Error bars are standard deviation.

Figure 41

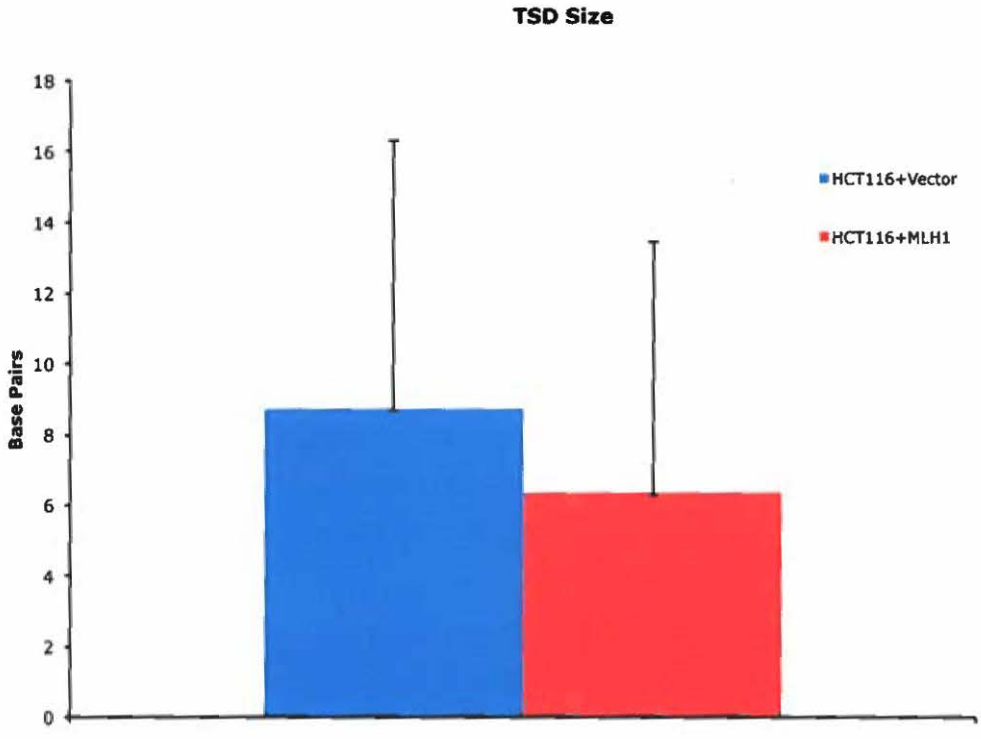



**Figure 41: Target site duplication size of rescues from HCT116+Vector and HCT116+MLH1**

Sequencing both ends of *de novo* LINE-1 insertions allowed us to determine the size of the target site duplications created by the staggered nature of the DSB. Only the areas of perfect homology were counted as the TSD. Error bars are standard deviation.

Table 1

| Full-length | Truncated | Inverted | Total | Cell Line | Reference |
|---|---|---|---|---|---|
| 6 | 75 | 19 | /100 | HeLa | (Gilbert *et al*, 2005) |
| 2 | 22 | 6 | /30 | HCT116 | (Symer *et al*, 2002) |
| 2 | 15 | 1 | /18 | HCT116+Vector | This study |
| 0 | 14 | 1 | /15 | HCT116+MLH1 | This study |

**Table 1: Proportions of different classes of LINE-1 insertions**

Each insertion was quantified as full-length, truncated, or inverted. Full-length insertions had to start at the +1 nucleotide of the LINE-1 construct. Inverted elements had to contain enough inverted sequence to unambiguously describe the inversion. Truncated elements were any that did not reach the +1 nucleotide or did not contain an inversion.

**MSH2 recognizes RNA/DNA structures that mimic LINE-1 retrotransposition intermediates**

Our data point to the cellular MMR machinery being able to recognize mismatches between the retroelement RNA and the DNA at the endonuclease target site. Repression of retrotransposition by the mismatch repair pathway would presumably require the recognition of mismatches by one of the two MutS heterodimers, which both contain MSH2. MSH2 is the most abundantly expressed MMR protein in HeLa cells (Chang, 2000), and complexes containing MSH2 directly recognize mismatched bases (Palombo *et al*, 1996). Direct surveillance of the priming step is the most straightforward mechanism of detection by MMR. To determine if the MMR sensor proteins are capable of recognizing LINE-1 initiation structures, we developed a LINE-1 Associated Bead-Oligo Capture Assay (LABOCA). The assay involves creating DNA/RNA duplexes containing a biotin tag. After incubation with HeLa cell nuclear extracts, we probed for MSH2 binding (Figure 42). MSH2 did not bind to a perfect DNA/DNA duplex (Lane 1) but can be seen bound to the DNA/DNA duplex containing a A/G mismatch (Lane 2). Binding was also observed to a DNA/RNA duplex that did not contain a mismatch (Lane 3). DNA/RNA duplexes containing a G/A mismatch displayed increased binding compared to the perfect duplex DNA/RNA construct, but were less than that of a mismatch DNA/DNA duplex (Lane 4).

**Discussion**

This study collectively demonstrated that mismatch repair acts to limit both LINE-1 and Alu retrotransposition. The repression of retrotransposition appeared to be at the priming stage of TPRT, where non-T/A base pairs can be formed. Alu elements

Figure 42

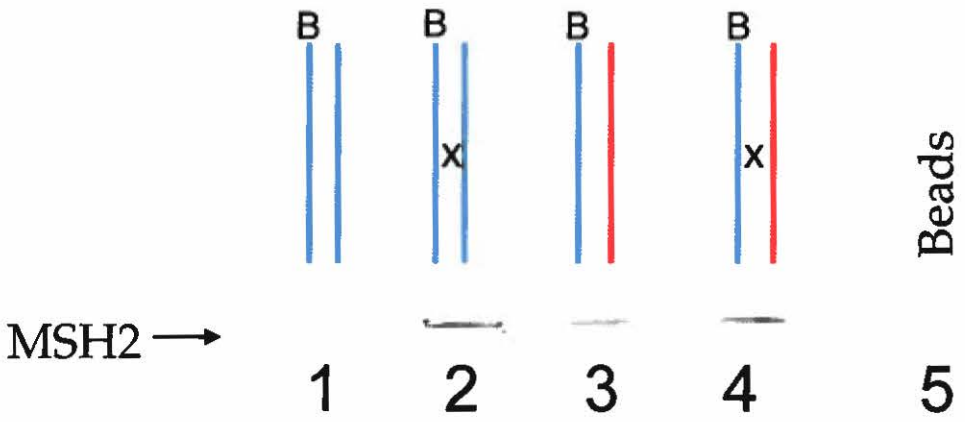



**Figure 42: LINE-1 Associated Bead-Oligo Capture Assay**

Oligos were designed to either contain a G/A mismatch or be perfect duplexes. After annealing they were bound to streptavidin magnetic beads and incubated with HeLa cell nuclear extract. Bound proteins were separated on a polyacrylamide gel and probed for MSH2 via western blotting. Blue lines are DNA oligos and contain a 5' biotin tag (B), red lines are RNA oligos. Mismatches are represented by x.

insert into the genome with their A-tail intact. This allows for the creation of different base/base mismatches. The most detrimental mismatch in an Alu A-tail creates a G/T mispair. These results agree with previous studies in the mismatch repair field that have shown a G/T mismatch is very well recognized and quickly repaired (Bishop *et al*, 1989).

Target sites for LINE-1 integration in MMR negative cells contained more guanines in the priming strand than MMR positive cells. Because LINE-1 elements are polyadenylated, this creates a G/A mismatch at the priming site. The change in target site composition makes it more closely resemble the spectrum of target sites capable of being cleaved by the LINE-1 endonuclease (Repanas *et al*, 2007). The difference in the magnitude of repression between LINE-1 and Alu can be explained by the ability of the mismatch repair pathway to have a differential response to mismatches (Bishop *et al*, 1989). Studies using purified MSH2/MSH6 have demonstrated that the repair of G/T and C/T mismatches is twice that of an A/G mismatch in duplex DNA (Genschel *et al*, 1998).

Previous work examining the mismatch repair pathway has been focused on the recognition and elimination of mismatches arising in DNA/DNA duplexes. The ability of human MSH2-containing complexes to recognize DNA/RNA duplexes represents a new substrate for the MMR machinery. Recently it was shown that MMR in bacteria and yeast act on duplexes containing mispaired ribonucleotides (Shen *et al*, 2011). This ability to recognize these duplexes across very different organism indicates that sensing mismatches in RNA/DNA duplexes is an important mutation-avoidance mechanism.

Regulation at the priming step is advantageous for the cell as endonuclease mediated insertion events begin with priming and faithful repair at this step would leave no trace of the attempted retroelement insertion. We did not find evidence of

microhomology mediated truncation or twin priming being regulated by MMR, but we cannot completely rule out possible MMR surveillance of those insertion intermediates. MSH2 binding to perfectly matched DNA/RNA duplexes opens the possibility that all retroelement insertions can be recognized by mismatch repair. It is possible that all retroelement insertions can be recognized and acted on by the mismatch repair machinery, not just those containing a mismatch.

The ability to repress LINE-1 insertion limits mutagenesis, but probably has no effect on the creation of DNA DSBs. However, "old" Alu elements that have been previously rendered incapable of retrotransposition by mutations in their A-tails (Comeaux *et al*, 2009) may become reactivated in cells lacking MMR activity. The high copy number of Alu makes this possibility an even greater threat to the cells.

The mechanism of MMR surveillance of retrotransposition is unknown at this time, but based upon our data we present two possibilities. First is a steric inhibition model where the formation of the mismatch between the DNA and RNA recruits mismatch repair factors and blocks access of ORF2 to the free 3' hydroxyl group and thus obstructs reverse transcription. Because we observed the repression in MLH1 positive and negative cells we believe that an alternate model of repression is more feasible. In an alternative model, the mismatch is recognized by an MSH2-containing heterodimer and recruits MutL. The endonuclease activity of MutL could then create a single strand break 5' of the endonuclease cleavage site, removing the initiation complex together. Alternatively, the recruitment of ExoI could then degrade the target site or utilize the RNaseH activity of ExoI and destroy the LINE-1 mRNA (Figure 43).

Figure 43

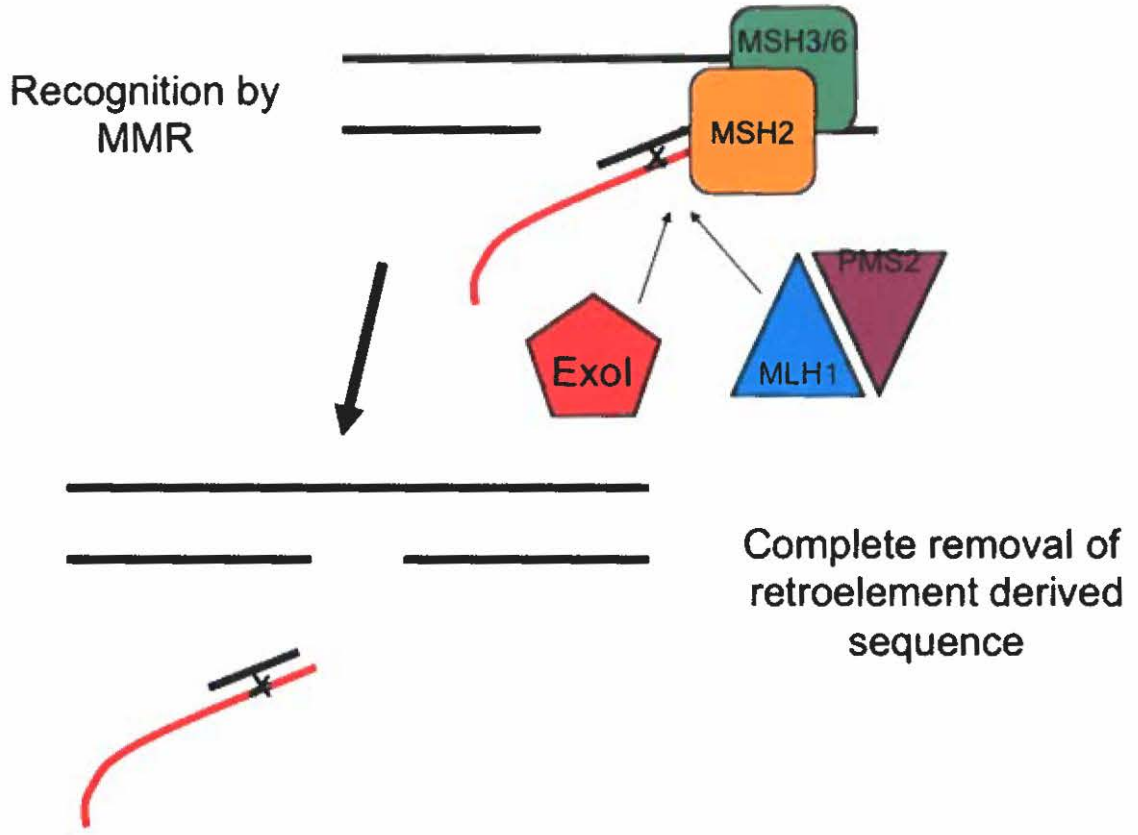



**Figure 43: Model of repression of LINE-1 and Alu by MMR**

The creation of a mismatch at the priming site recruits a MutS heterodimer. MutS recruits MutL, which may create a single strand break upstream of the LINE-1 induced break, causing the pre-insertion complex to dissociate from the DNA. Alternatively, the exonuclease activity of ExoI can degrade the small DNA flap created or truncate the RNA through its RNaseH activity.

Genetic instability is a hallmark of HNPCC tumors. Cells that lack the control of MMR have an increased mutation rate and display MSI (Parsons *et al*, 1993). In addition to their already high mutation rate, we showed that HNPCC tumors lacking in MMR might have an increased retrotransposition rate, further increasing the genetic burden. It remains to be seen if HNPCC tumors lacking MMR do display increased levels of retrotransposition compared to surrounding, MMR-containing tissues.

**Methods**

**Cells and culture conditions:** HCT116 cell lines were maintained in MEM (Gibco) containing 10% FBS (Gibco). The MEF cell lines were grown in DMEM (Gibco) supplemented with 10% FBS (Gibco). 293T-Lα cells were grown in DMEM supplemented with 10% tetracycline screened FBS (Clontech), hygromycin, zeocin, and penicillin/streptomycin. All cell lines were grown in a humidified incubator at 37 °C and 5% $CO_2$.

**Retrotransposition analysis**: HCT116 – cells were seeded at a density of $5x10^5$ per T-75 culture plate. 1 μg of tagged LINE-1 vector (L1.3CMV-neo) was transfected using Lipofectamine2000 (Invitrogen) according to manufacturer directions. Twenty-four hours after transfection LINE-1 retrotransposition events were selected for using 400 μg/mL G418 (HCT116+MLH1 and HCT116+Vector) or 5 μg/mL blasticidin (HCT116+Ch3 and HCT116) and maintained until colonies formed. 293T-Lα - cells were seeded at a density of $4x10^6$ and transfected with 4 μg LINE-1-Blasticin construct using Lipofecatmine2000. Two days post transfection $5x10^5$ cells were reseeded and insertions were selected for using 5 μg/mL blasticidin until colonies formed. Colonies were fixed, stained with crystal violet, and counted. Alu – HeLa cells were seeded at a

μg of HeLa cell nuclear extract (CelLytic NuCLEAR Extraction Kit, Sigma) was incubated with the duplexes in assay buffer (20 mM Tris pH 8, 2 mM MgCl2, 100 mM NaCl, poly dI:dC, protease inhibitor cocktail) at room temperature with gentle rotation for 45 minutes. After incubation, the complexes were washed via magnetic separation with assay buffer three times. The bound proteins were separated from the oligos by boiling in Laemmli buffer and subjected to Western blot analysis for MSH2 (Calbiochem). MSH2 antibodies were diluted 1:1000 in blocking buffer and incubated at 4 °C overnight. Alexa-fluor-800CW conjugated secondary antibodies were used for visualization using a Li-Cor Oddessy Pro IR scanner at a dilution of 1:5000 in blocking buffer. The sequences of the oligos are shown in Table 4.

**Microsatellite instability assay**: HCT116+MLH1 and HCT116+Vector cells were transfected with pCA-OF18 which contains an out of frame EGFP preceded by (CA)17. Two days post transfection the cells were subjected to FACS analysis on a BD LSR II using the FACS Diva 6.1.3 software platform. $5x10^4$ cells per experiment were analyzed.

**Weblogo creation**: Frequency Weblogos (http://weblogo.berkeley.edu/logo.cgi) were created using only the four base pairs used for priming the TPRT reaction. The endonuclease target site data was simulated using data from Repanas *et al.* The priming sites of genomic LINE-1s were compiled from analysis performed by Szak *et al.*

**Statistical analysis:** Two-tailed Student's t-tests were performed using Microsoft Excel's built in statistical package. One-way ANOVA was done in Prism 5 for Windows. Fisher's exact tests were completed using GraphPad's online calculator found at http://www.graphpad.com/quickcalcs/contingency1.cfm. Experiments were routinely done in triplicate with multiple replicates.

Table 2

| Primer Name | Sequence |
|---|---|
| L1opt-neo-rescue related primers | |
| Neo | TCGCCTTCTTGACGAGTTCT |
| 4960 | TATTCCCCTTCCTGTGTCCA |
| 4590 | GGGTCAAATGGTATTTCTAGTTCA |
| 4388 | TTCTCTGATGGCCAGTGATG |
| 4139 | TGCAGAAGCTCTTTAGCTTGA |
| 3915 | GTTTTCCCAGCACCACTTGT |
| 3474 | CCATTTGTTTGTCTCCTCTTTG |
| 3043 | CCAGTTTTTGCCCATTCAGT |
| 2730 | TTCGGTTTGCCAGTATCTTG |
| 2273 | TCCCCTTTATCATTTTTGATCG |
| 2240 | ATCAGTCTTGCAAGCGGTCT |
| 1217 | CAGCACACTGATGGGTCTTG |
| 815 | TCCTTCAGGATGTTGAAGATTG |
| 397 | CGTTCATCTCGTCCTCCATT |
| Genomic sequencing primers for insertions in HCT116+MLH1 cells | |
| m17-18 | TGGCAAATGCTAAAACATGG |
| m21-09 | CGGGCTCAAGTAATCTGCTC |
| m21-10 | TGTTCAATGGGTGCTTGAAA |
| m21-16 | TGATGGCATCCACTCTGAAG |
| m111-1 | TGTCATTACGGAATAGATGCTG |
| m111-2 | AGGCCCCATTAACCACTACC |
| m211-1 | GCTCAAAGAGGGTGAGCAAT |
| m211-2 | TTCCACCCATATCAAATAATTCAA |
| m211-4 | CCAAGGTGGCCTAGATTCAA |
| m316-4 | GGGCTTCTTCAATCTGTGGA |
| m316-6 | CACTTTCTCCTGCCTCAACC |
| m720-4 | TGTTAAACAGAGACCTCATTCCTG |
| m720-5 | CAAAGCTATGATGCCCAACTT |
| m215-8 | GTGCAGTGTGCTTTTGGTGT |
| m215-11 | CATCATTCTCATTCCTTCTTTTGA |
| Genomic sequencing primers for insertions in HCT116+Vector cells | |
| v18-07 | TTTGCCTTTAGGGAAATTTAACA |
| v18-09 | TGTGAAAATGCTTCAAATCTGC |
| v18-23 | CGTAAAGTTGTTGTGGGTCAAA |
| v19-03 | TTTTGTGGAATGCTCCTTCA |
| v1020-02 | TGTGTTAGGCAGCAGTCAGG |
| v1020-14 | CCACTGAATTCAAAGGTTTTCA |
| v1020-15 | GAAGACTCTTTTCCATCAATTTCTTT |
| v1020-16 | GCCACATTTTCTTTAACCATTCA |
| v1020b-09 | TATTGAACCAGCCTTGCATC |
| v1020b-12 | GGAAACATGCACCACCTCAT |
| v1020b-16 | CCTCACAGCCAAGCACATTA |
| v1020b-19 | ACCCAAGAATCATTCAGGAGCAGGT |

| | |
|---|---|
| v1112-01 | TTCCTTCTTGCCTCATTGCT |
| v211-01 | CACTCAGGTTTTTGGGCAAT |
| v811-01 | TGCAGAAGCTCTTTAGCTTGA |
| v1020-10 | CATTTTACTCAATCTGAGCAATACAA |

**Table 2: Primers used for sequencing rescued LINE-1 insertions**

L1opt-neo-rescue primers are specific for the LINE-1 rescue construct and take advantage of sequence changes from the consensus sequence of a genomic L1HS subfamily element where possible. These primers sequence L1opt-neo-rescue from 3' to 5'. The HCT116 related primers are specific to the genomic locus of each insertion.

Table 3

| Clone | Target Site Duplication | TSD Size | 5' junction | Priming Site | Poly A tail | Breakpoint location | Notes |
|---|---|---|---|---|---|---|---|
| m17-18 | | 0 | 4610 | TTAA/GT | 0 | 10:104346443 | EN independent, 19 bp deletion (GTTGGGGAAGATGAAAAAC) |
| m21-09 | | 0 | 3719 | TTAT/AA | 154 | 2:157123835 | untemplated TG, 3 bp deletion (AAA) |
| m21-10 | | 0 | 3849 | TTTA/TT | 80 | 6:20678726 | 6 bp homology (TTGTCA) |
| m21-16 | AACAAATATT | 10 | 4728 | TATT/AA | 51 | 13:44693563 | untemplated CC |
| m111-1 | CCCCTCCCTTTTC | 13 | 4943 | TTTC/AC | 29 | 6:18633816 | duplication (6485-6133), template switch to U6 with homology at each end |
| m111-2 | TTAGAGTCTTTTTTTTTTTT | 20 | 5011 | TTTT/AA | 153 | 19:31042032 | 1 bp homology (T) |
| m211-1 | ACATCTGAGAATTTTCT | 17 | 4761 | TTCT/AT | 47 | 12:77557039 | 5 bp homology with 1 mismatch (ACATcT) |
| m211-2 | CATATTTCTT | 10 | 4322 | TCTT/GT | 130 | 3:96850778 | 5 bp homology (CATAT), G->C mutation in TSD |
| m211-4 | | 0 | 4028 | TCTT/AA | 57 | 1:241635186 | 8 bp deletion (AAAAAATA), 2 bp homology (TA) |
| m316-4 | | 0 | 3120 | TTCT/AT | 50 | 7:127121885 | 5 bp deletion (ATCGT), 2 bp homology (GA) |
| m316-6 | | 0 | 3499 | TTTT/AA | 54 | 8:61490873 | 355 bp deletion (see data sheet) |
| m720-4 | ATGAAATCTT | 10 | 5007 | TCTT/AA | 87 | 4:30621241 | |
| m720-5 | | 0 | 3254 | TTTT/CA | 68 | 12:96976267 | 8 bp homology (CAAAGAAA) |
| m215-8 | TTA | 3 | 4542 | TTTA/AT | 30 | chr1:8080929 | 2 bp homology (TT) |
| m215-11 | AACTTCCTATTT | 12 | 4550 | ATTT/AA | 36 | chr2:182863284 | 1 bp homology (A) |
| | | | | | | | |
| v18-07 | TTAAATAATT | 10 | 4265 | AATT/AA | 71 | 5:82029091 | 1 bp homology (T) |
| v18-09 | ATAAAGTTCATTTTTT | 16 | 4723 | TTTT/AC | 98 | 5:91733754 | Inversion (4722,4705-4723) |
| v18-23 | TTTAATTTCTTT | 12 | 4404 | CTTT/AA | 37 | 8:117725595 | |
| v19-03 | | 0 | 4318 | TCAG/TT | 11 | 5:169213298 | 2 bp homology (TT), 7 bp deletion (ACTTCAG) |
| v1020-02 | GGGACACAGCCTTTCT | 16 | 53 | TTCT/AA | 141 | 3:154513109 | |
| v1020-14 | | 0 | 61 | TTTT/AA | 61 | 7:120267247 | TTTTT DR? |
| v1020-15 | TATATTGAAAATCTT | 15 | 1 | TCTT/AT | 85 | 6:153029357 | |
| v1020-16 | TTGTATGTCTTCTTTT | 16 | 3553 | TTTT/GA | 119 | 4:43303665 | 151 bp of Ch2 seq at 5' end, template switch, DNA repair, artifact of rescue? |
| v1020b-09 | | 0 | 4511 | TTCT/AA | 73 | 4:54459805 | Chimera associated with 23 kb deletion, 34 bp homology, GG in A-tail |
| v1020b-12 | 293 bp | 293 | 5047 | TTTT/CT | 84 | 14:39491780 | 10 bp homology |
| v1020b-16 | | 0 | 4842 | AGTT/AT | 0 | 7:116163466 | no A-tail, microhomology on both ends, 2 bp deletion (TT) at 5' end |
| v1020b-19 | | 0 | 4704 | TTCT/GA | 22 | 3:98627672 | 5 bp deletion (GATGA), 4 bp homology (TGTG) |
| v1112-01 | CTTTTCCTTTATTTTAT | 17 | 1 | TTAT/AG | 70 | 8:100896695 | |
| v211-01 | TGGTTTTGTTACTTTG | 16 | 5035 | TTTG/AA | 101 | 5:97522576 | 6 bp homology (TGGTTT) |
| v811-01 | | 0 | 4051 | TGTT/AA | 65 | 1:109513130; 12:65935479 | Rearrangement between Ch1 and Ch12, untemplated at 3' end (TTAATATATA), 3 bp homology (CTT) |
| v1020-10 | GAATTTATTTCTT | 13 | 3754 | TCTT/AA | 103 | 6:94673422 | |
| v1020-05 | | | 4304 | | | 6:71782543 | untemplated CAAT, extra piece of Ch 6 ligated during cloning |
| v1020b-07 | | | 4487 | | | 8:132782353 | translocation in 5' flank |

**Table 3: Rescued insertions from HCT116+MLH1 and HCT116+Vector**

Shown are the characteristics of each rescued LINE-1 insertion from HCT116+MLH1 (clones that start with m) and

[illegible] between the 5' end of the element and the flanking genomic DNA is

Table 4

| Oligo name | Sequence (5' -> 3') |
|---|---|
| Biotin-T-DNA | /Biotin/cgcgtagatcgatgttttttttttcgcgcagagctcgcgc |
| DNA-comp | gcgcgagctgtgcgcgaaaaaaaaaacatcgatctacgcg |
| Biotin-G-DNA | /Biotin/cgcgtagatcgatgttttt**G**ttttcgcgcagagctcgcgc |
| RNA-comp | gcgcgagcucugcgcgaaaaaaaaaacaucgaucuacgcg |

**Table 4: Oligonucleotides used to create LABOCA substrates**

The biotin tag is labeled as /Biotin/. The nucleotide that creates a mismatch is in bold uppercase.

# Chapter 4

## Discussion

### Constant LINE-1 expression alters DSB repair

The mutagenic potential of LINE-1 is two fold. First, LINE-1, as well as other sequences mobilized by LINE-1, can insert into genes or into gene regulatory elements. This can lead to both disrupted gene expression and mutated gene products (Figure 4). The second mechanism of genomic instability associated with LINE-1 expression is the creation of DNA DSBs (Gasior *et al*, 2006). DSBs are a toxic lesion that must be repaired by the cell in order to survive (Wallace *et al*, 2008). However, even the repair process can introduce mutations in the form of non-allelic HR or nucleotide deletions at sites of NHEJ. It is critical for the cell to maintain genetic integrity; therefore controlling LINE-1 activity is of utmost importance.

The repression of retrotransposition (Wallace *et al*, 2010) represents a cellular protective measure to limit the damage caused by LINE-1 expression. This mechanism explains why normal cells can tolerate high levels of LINE-1 expression without succumbing to either senescence or apoptosis. It also allows cells to resist mutations that might otherwise push them toward transformation and cancer progression.

It has been reported that constant exposure to low levels of the antibiotic zeocin subjects the cell to prolonged DSBs (Delacôte *et al*, 2007). These zeocin-treated cells display increased DSB repair kinetics, which concurs with our observation of the same

phenomenon in cells that chronically express LINE-1. An increase in DSB repair also protects the integrity of the genome.

It is possible that in the cell's attempt to limit LINE-1-associated damage, it permits an increased number of smaller mutations. NHEJ is associated with mutations, usually in the form of small deletions, occurring when the ends of DSBs cannot be precisely ligated back together (Kass & Jasin, 2010). In this case, small deletions might be less deleterious to the cell than multiple kbs of inserted sequence.

**MMR limits LINE-1 and recognizes insertion intermediates**

Based on the observation that disruptions located near the 3' terminus of an Alu element inhibited retrotransposition, and that Alu elements preferentially prime near their 3' end, we speculated that mismatches could be formed at the TPRT initiation site. As such, we reasoned that those mismatches could become a substrate for the MMR pathway.

We demonstrated in multiple cell lines lacking MLH1, and therefore MMR deficient, an increased rate of LINE-1 retrotransposition. Insertion sites rescued from cells lacking MMR displayed an increased frequency of guanine nucleotides at the priming site. The presence of guanines in the priming site would create a mismatch between the DNA and the poly-A tail of the element. It is unclear why guanine residues would be selected against specifically, as MMR displays similar repair affinities for C/A and G/A mismatches (Bishop *et al*, 1989), yet cytosine and adenine at the priming site are tolerated (Figure 38). The presumed G/T mismatch formed in our Alu retrotransposition assays, however, is strongly recognized by MMR (Fang & Modrich, 1993; Genschel *et al*, 1998).

The inclusion of guanine in the LINE-1 insertion sites makes the insertion site preference more closely resemble the endonuclease site preference. The A-T richness is maintained, but sites with guanine are clearly capable of being cleaved by the LINE-1 endonuclease at an intermediate level (Repanas *et al*, 2007). The removal of LINE-1 insertion complexes at these sites could help explain the order of magnitude difference between the number of DSBs created per cell and the number of insertion events.

Only recently has it been shown that MMR could recognize mispaired ribonucleotides in a DNA strand in bacteria and yeast (Shen *et al*, 2011). The ability of MSH2 to recognize DNA/RNA duplexes represents a new substrate for MMR in human cells.

It is also advantageous for the cell to limit LINE-1 insertions at the priming step. Regulation early in the TPRT process would allow for a greater chance of limiting mutation by removing pre-insertion complexes before they had a chance to complete. Again, if a DSB is formed at these sites, the repair by NHEJ may lead to small deletions, so the mutation would not be completely avoided. It is also possible in this case that MMR can limit the size of the insertion, creating small insertions that would be undetectable in our system unless they were large enough to allow for the expression of the resistance tag.

The data presented here address the possibility that the genetic instability of tumors lacking MMR has been wholly underestimated. In addition to the increased mutation rate associated with MMR deficiency, we have shown an increase in the rate of LINE-1 insertions. This could, in turn, lead to even more mutations and possibly contribute to carcinogenesis and cancer progression. As more mutations occur, the cell

could become even more unstable, creating a positive feedback loop. It is interesting to note that if MMR deficiency allows for mismatches at retroelement insertion sites, it is possible that inactive Alu elements with battered A-tails (Comeaux *et al*, 2009) could be reactivated, adding to the mutation rate (Figure 44).

**A unified model of DNA repair and LINE-1**

DNA repair plays a vital role in the control of LINE-1. In Chapter 2 we showed that by exposing cells to LINE-1 we could elicit a response that increased the rate of DSB repair and decreased the rate of retrotransposition, likely due to the chronic exposure to DSBs created by the LINE-1 endonuclease. This represents a cellular protective measure to maintain genomic integrity under conditions of high LINE-1 expression. In Chapter 3 we showed that mutations in DNA repair genes, in this case MMR, could lead to an increase in LINE-1 retrotransposition by tolerating insertions that contain a mismatch at their initial priming site.

Together these data expand on the previous studies involving DNA repair and LINE-1 and present two related branches of regulation. First is the simple observation that an increase in DSB repair leads to the repression of LINE-1. The specific pathway(s) involved are as yet undefined, but the repression is reliant on the formation of DSBs (Wallace *et al*, 2010). Here, a comparison can be made between our study and the work by Delacote *et al* that showed an increase in NHEJ activity in cells exposed to chronic DSBs (Delacôte *et al*, 2007).

The second branch involves the opposite response. Cells with decreased DNA repair displayed an increase in LINE-1 activity. In addition to the work presented here involving MMR, previous studies have shown disruptions in both NER (Gasior *et al*,

2008) and NHEJ (Morrish *et al*, 2002) led to an increase in LINE-1 retrotransposition. The correlation between NER dysfunction and increased rates of LINE-1 retrotransposition is further strengthened by recent work showing that six out of twenty non-small cell lung cancers contained somatic LINE-1 insertions (Iskow *et al*, 2010). Lung cancers commonly have defects in the NER pathway (Cheng *et al*, 2000) and loss of NER correlates with increased predisposition to cancer (Friedberg *et al*, 2000; Ide *et al*, 2000) and a poor outcome (Matakidou *et al*, 2007).

Additionally, the work presented in Chapter 3 provides an explanation for why exposure to heavy metals causes a stimulation of LINE-1 (Kale *et al*, 2005). DNA repair pathways are frequent targets of heavy metals (Beyersmann & Hartwig, 2008). In particular, cadmium binds to the MSH2/MSH6 heterodimer and inhibits DNA binding (Wieland *et al*, 2009). Expression of MSH6 is downregulated by exposure to cadmium, possibly through the inhibition of transcription factors (Hsu *et al*, 2010). Cadmium also disrupts the NER pathway by disrupting the binding of capability of XPA (Hartmann & Hartwig, 1998) and the cellular localization of XPC (Schwerdtle *et al*, 2010).

Stimulation, or loss of repression, of LINE-1 presents a double-edged sword. In one case increased LINE-1 expression can lead to an increase in mutation rate. This could accelerate disease progression or aid in cellular transformation. Alternatively, cells with increased LINE-1 activity could be subject to cell death stemming from the creation of DSBs (Wallace *et al*, 2008). The dose of LINE-1 and the time course of expression would be major determining factors in the cellular response.

As an interesting contrast to the model presented above, treatment with IR leads to an increase in LINE-1 insertions (Farkash *et al*, 2006). If the formation of DSBs alone

were enough to cause a decrease in LINE-1 retrotransposition then IR would be thought to decrease instead of stimulate activity. However, the DSBs created by IR are more acute than the DSBs created by long term expression of LINE-1 and their quick repair does not allow for the required cellular changes to modulate LINE-1 activity (Figure 45).

**Future directions**

We have demonstrated that an increase in retrotransposition rate is associated with MMR deficiency *in vitro*. Translating our results *in vivo* will require the deep sequencing of cancer genomes deficient in MMR and comparing the rate of retrotransposition to normal tissues. The use of both LINE-1-targeted and whole genome second-generation sequencing techniques will be instrumental in accomplishing this goal. This will also allow us to examine whether MMR acts to limit retrotransposition by creating very short insertions that would be undetectable in our present tagged LINE-1 retrotransposition system.

While the selection against guanine residues in the priming site is evident, the mechanism behind the repression is not. Further work will aim to assess whether the MMR pathway has the same base/base mismatch affinities for DNA/RNA duplexes as it does for DNA/DNA duplexes.

Testing the effect of MMR on Alu retrotransposition directly is also of importance. Creating MMR-deficient cell lines that lack neomycin resistance will allow the same assays completed here with LINE-1 to be repeated using Alu. A novel Alu rescue vector has also recently been developed to allow the characterization of the insertions sites (Wagstaff, submitted).

This dissertation has outlined two new mechanisms employed by cells in the control of retroelements. This adds to the already impressive list of pathways involved in repressing the activity of LINE-1 and illustrates the importance of limiting LINE-1 activity.

Figure 45

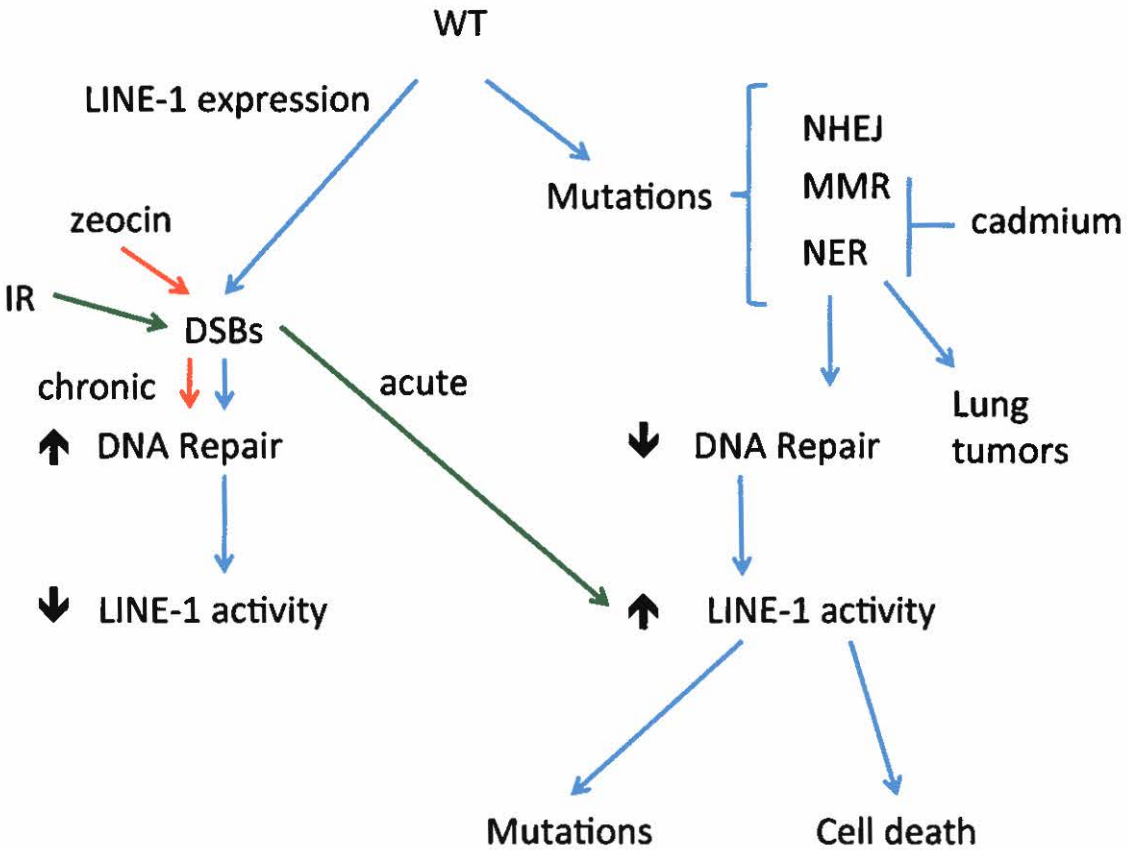



**Figure 45: DNA repair and LINE-1**

DNA repair is linked to LINE-1 through several pathways. The cell can regulate LINE-1 by increasing DNA repair in response to chronic exposure to LINE-1 induced DSBs. Cells also increase DNA repair in response to chronic DSBs created by treatment with zeocin. When DNA repair is impaired either through mutations or exposure to environmental toxins the cells experiences a decrease in DNA repair potential affecting several different pathways. This decrease in DNA repair leads to an increase LINE-1 retrotransposition. Increases in LINE-1 activity can lead to increased mutation rates or cell death.

**List of References**

Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev*, Vol. 20, pp. 210-224.

Athanikar JN, Badge RM, Moran JV. (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Research*, Vol. 32, pp. 3846-3855.

Babcock M. (2003) Shuffling of Genes Within Low-Copy Repeats on 22q11 (LCR22) by Alu-Mediated Recombination Events During Evolution. *Genome Research*, Vol. 13, pp. 2519-2532.

Becker KG, Swergold GD, Ozato K, Thayer RE. (1993) Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum Mol Genet*, Vol. 2, pp. 1697-1702.

Belancio VP, Deininger PL, Roy-Engel AM. (2009) LINE dancing in the human genome: transposable elements and disease. *Genome Med*, Vol. 1, p. 97.

Belancio VP, Hedges DJ, Deininger P. (2006) LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res*, Vol. 34, pp. 1512-1521.

Belancio VP, Hedges DJ, Deininger P. (2008) Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res*, Vol. 18, pp. 343-358.

Belancio VP, Roy-Engel AM, Deininger PL. (2010a) All yall need to know 'bout retroelements in cancer. *Seminars in Cancer Biology*, pp. 1-11.

Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. (2010b) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res*.

Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A. (2006) Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int*, Vol. 6, p. 13.

Beyersmann D, Hartwig A. (2008) Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. *Arch Toxicol*, Vol. 82, pp. 493-512.

Bhattacharyya NP, Skandalis A, Ganesh A, Groden J, Meuth M. (1994) Mutator phenotypes in human colorectal carcinoma cell lines. *Proc Natl Acad Sci USA*, Vol. 91, pp. 6319-6323.

Bishop DK, Andersen J, Kolodner RD. (1989) Specificity of mismatch repair following transformation of Saccharomyces cerevisiae with heteroduplex plasmid DNA. *Proc Natl Acad Sci USA*, Vol. 86, pp. 3713-3717.

Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci USA*, Vol. 103, pp. 8780-8785.

Buermeyer AB, Wilson-Van Patten C, Baker SM, Liskay RM. (1999) The human MLH1 cDNA complements DNA mismatch repair defects in Mlh1-deficient mouse embryonic fibroblasts. *Cancer Res*, Vol. 59, pp. 538-541.

Burdett V, Baitinger C, Viswanathan M, Lovett ST, Modrich P (2001) In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc Natl Acad Sci U S A* **98:** 6765-6770

Burke WD, Calalang CC, Eickbush TH. (1987) The site-specific ribosomal insertion element type II of Bombyx mori (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol Cell Biol*, Vol. 7, pp. 2221-2230.

Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E (2002) A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1. *Genomics* **80:** 402-406

Cejka P, Stojic L, Mojas N, Russell AM, Heinimann K, Cannavó E, di Pietro M, Marra G, Jiricny J. (2003) Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. *EMBO J*, Vol. 22, pp. 2245-2254.

Chang DK. (2000) Steady-state Regulation of the Human DNA Mismatch Repair System. *Journal of Biological Chemistry*, Vol. 275, pp. 18424-18431.

Cheng L, Spitz MR, Hong WK, Wei Q. (2000) Reduced expression levels of nucleotide excision repair genes in lung cancer: a case-control analysis. *Carcinogenesis*, Vol. 21, pp. 1527-1530.

Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. (2009) Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? *Genome Res*, Vol. 19, pp. 545-555.

Cost GJ, Feng Q, Jacquier A, Boeke JD. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J*, Vol. 21, pp. 5899-5910.

Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y, Carson CT, Macia A, Moran JV, Gage FH. (2011) Telomerase and Retrotransposons: Reverse Transcriptases That Shaped Genomes Special Feature Sackler Colloquium: Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci USA*, Vol. 108, pp. 20382-20387.

Dai L, Huang Q, Boeke JD. (2011) Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochem*, Vol. 12, p. 18.

De Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. (2011) Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet*, Vol. 7, p. e1002384.

Delacôte F, Deriano L, Lambert S, Bertrand P, Saintigny Y, Lopez BS. (2007) Chronic exposure to sublethal doses of radiation mimetic Zeocin selects for clones deficient in homologous recombination. *Mutat Res*, Vol. 615, pp. 125-133.

DeStefano JJ, Buiser RG, Mallaber LM, Myers TW, Bambara RA, Fay PJ. (1991) Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled. *J Biol Chem*, Vol. 266, pp. 7423-7431.

Dewannieux M, Esnault C, Heidmann T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*, Vol. 35, pp. 41-48.

Divoky V, Indrak K, Mrug M, Brabec V, Huisman T, Prchal J (1996) A novel mechanism of beta thalassemia: The insertion of L1 retrotransposable element into beta globin IVS II. *Blood* **88**

Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN. (2007) Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol Cell Biol*, Vol. 27, pp. 4685-4697.

Doherty KM, Sharma S, Uzdilla LA, Wilson TM, Cui S, Vindigni A, Brosh RM, Jr. (2005) RECQ1 helicase interacts with human mismatch repair factors that regulate genetic recombination. *J Biol Chem* **280:** 28085-28094

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N. (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet*, Vol. 6.

Dzantiev L, Constantin N, Genschel J, Iyer RR, Burgers PM, Modrich P. (2004) A defined human system that supports bidirectional mismatch-provoked excision. *Molecular Cell*, Vol. 15, pp. 31-41.

Ejima Y. (2003) Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. *Hum Mol Genet*, Vol. 12, pp. 1321-1328.

Elliott B, Jasin M (2001) Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. *Mol Cell Biol* **21:** 2671-2682

Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O. (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, Vol. 433, pp. 430-433.

Esnault C, Maestre J, Heidmann T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*, Vol. 24, pp. 363-367.

Fang WH, Modrich P. (1993) Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. *J Biol Chem*, Vol. 268, pp. 11838-11844.

Fanning T, Singer M. (1987) The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res*, Vol. 15, pp. 2251-2260.

Farkash EA, Kao GD, Horman SR, Prak ETL. (2006) Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res*, Vol. 34, pp. 1196-1204.

Feng Q, Moran JV, Kazazian HH, Boeke JD. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, Vol. 87, pp. 905-916.

Flanagan SA, Krokosky CM, Mannava S, Nikiforov MA, Shewach DS. (2008) MLH1 deficiency enhances radiosensitization with 5-fluorodeoxyuridine by increasing DNA mismatches. *Molecular Pharmacology*, Vol. 74, pp. 863-871.

Friedberg EC, Bond JP, Burns DK, Cheo DL, Greenblatt MS, Meira LB, Nahari D, Reis AM. (2000) Defective nucleotide excision repair in xpc mutant mice and its association with cancer predisposition. *Mutat Res*, Vol. 459, pp. 99-108.

Fujita H, Kato J, Horii J, Harada K, Hiraoka S, Shiraha H, Sakaguchi K, Shiratori Y. (2007) Decreased expression of hMLH1 correlates with reduced 5-fluorouracil-mediated apoptosis in colon cancer cells. *Oncol Rep*, Vol. 18, pp. 1129-1137.

Gasior SL, Roy-Engel AM, Deininger PL. (2008) ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst)*, Vol. 7, pp. 983-989.

Gasior SL, Wakeman TP, Xu B, Deininger PL. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol*, Vol. 357, pp. 1383-1393.

Genschel J, Bazemore LR, Modrich P (2002) Human exonuclease I is required for 5' and 3' mismatch repair. *J Biol Chem* **277:** 13302-13311

Genschel J, Littman SJ, Drummond JT, Modrich P. (1998) Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *J Biol Chem*, Vol. 273, pp. 19895-19901.

Genschel J, Modrich P. (2003) Mechanism of 5'-directed excision in human mismatch repair. *Molecular Cell*, Vol. 12, pp. 1077-1086.

Gilbert N, Lutz S, Morrish TA, Moran JV. (2005) Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol Cell Biol*, Vol. 25, pp. 7780-7795.

Gilbert N, Lutz-Prigge S, Moran JV. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell*, Vol. 110, pp. 315-325.

Goodier JL. (2004) A potential role for the nucleolus in L1 retrotransposition. *Hum Mol Genet*, Vol. 13, pp. 1041-1048.

Goodier JL, Mandal PK, Zhang L, Kazazian HH. (2010) Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum Mol Genet.*

Goodier JL, Ostertag EM, Kazazian HH. (2000) Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum Mol Genet*, Vol. 9, pp. 653-657.

Goodier JL, Zhang L, Vetter MR, Kazazian HH. (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol Cell Biol*, Vol. 27, pp. 6469-6483.

Haoudi A, Semmes OJ, Mason JM, Cannon RE. (2004) Retrotransposition-Competent Human LINE-1 Induces Apoptosis in Cancer Cells With Intact p53. *J Biomed Biotechnol*, Vol. 2004, pp. 185-194.

Hartmann M, Hartwig A (1998) Disturbance of DNA damage recognition after UV-irradiation by nickel(II) and cadmium(II) in mammalian cells. *Carcinogenesis* **19:** 617-621

Hata K, Sakaki Y. (1997) Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene*, Vol. 189, pp. 227-234.

Hattori M, Kuhara S, Takenaka O, Sakaki Y. (1986) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature*, Vol. 321, pp. 625-628.

Her C, Vo AT, Wu X. (2002) Evidence for a direct association of hMRE11 with the human mismatch repair protein hMLH1. *DNA Repair (Amst)*, Vol. 1, pp. 719-729.

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

Hohjoh H, Singer MF. (1997) Ribonuclease and high salt sensitivity of the ribonucleoprotein complex formed by the human LINE-1 retrotransposon. *J Mol Biol*, Vol. 271, pp. 7-12.

Holmes SE, Singer MF, Swergold GD. (1992) Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *J Biol Chem*, Vol. 267, pp. 19765-19768.

Hong Z, Jiang J, Hashiguchi K, Hoshi M, Lan L, Yasui A. (2008) Recruitment of mismatch repair proteins to the site of DNA damage in human cells. *Journal of Cell Science*, Vol. 121, pp. 3146-3154.

Hsu T, Tsai H-T, Huang K-M, Luan M-C, Hsieh C-R. (2010) Sublethal levels of cadmium down-regulate the gene expression of DNA mismatch recognition protein MutS homolog 6 (MSH6) in zebrafish (Danio rerio) embryos. *Chemosphere*, Vol. 81, pp. 748-754.

Ide F, Iida N, Nakatsuru Y, Oda H, Tanaka K, Ishikawa T. (2000) Mice deficient in the nucleotide excision repair gene XPA have elevated sensitivity to benzo[a]pyrene induction of lung tumors. *Carcinogenesis*, Vol. 21, pp. 1263-1265.

Iskow RC, Mccabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE. (2010) Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons. *Cell*, Vol. 141, pp. 1253-1261.

Itahana K, Campisi J, Dimri GP. (2007) Methods to detect biomarkers of cellular senescence: the senescence-associated beta-galactosidase assay. *Methods Mol Biol*, Vol. 371, pp. 21-31.

Iyer RR, Pluciennik A, Burdett V, Modrich PL. (2006) DNA mismatch repair: functions and mechanisms. *Chem Rev*, Vol. 106, pp. 302-323.

Jackson SP, Bartek J. (2009) The DNA-damage response in human biology and disease. *Nature*, Vol. 461, pp. 1071-1078.

Jacob S, Aguado M, Fallik D, Praz F. (2001) The role of the DNA mismatch repair system in the cytotoxicity of the topoisomerase inhibitors camptothecin and etoposide to human colorectal cancer cells. *Cancer Research*, Vol. 61, pp. 6555-6562.

Jones RB, Garrison KE, Wong JC, Duan EH, Nixon DF, Ostrowski MA. (2008) Nucleoside Analogue Reverse Transcriptase Inhibitors Differentially Inhibit Human LINE-1 Retrotransposition. *PLoS ONE*, Vol. 3, p. e1547.

Kadyrov FA, Dzantiev L, Constantin N, Modrich P (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* **126:** 297-308

Kale SP, Moore L, Deininger PL, Roy-Engel AM (2005) Heavy metals stimulate human LINE-1 retrotransposition. *Int J Environ Res Public Health* **2:** 14-23

Kass EM, Jasin M. (2010) Collaboration and competition between DNA double-strand break repair pathways. *FEBS Lett*, Vol. 584, pp. 3703-3708.

Kazazian HH, Moran JV. (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet*, Vol. 19, pp. 19-24.

Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, Vol. 332, pp. 164-166.

Ke SH, Wartell RM (1993) Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA; determination by temperature-gradient gel electrophoresis. *Nucleic Acids Res* **21:** 5137-5143

Khazina E, Truffault V, Büttner R, Schmidt S, Coles M, Weichenrieder O. (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol*, Vol. 18, pp. 1006-1014.

Khazina E, Weichenrieder O. (2009) Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc Natl Acad Sci USA*, Vol. 106, pp. 731-736.

Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W, Kazazian HH. (1999) Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet*, Vol. 8, pp. 1557-1560.

Kimura F, Seifert H-H, Florl AR, Santourlidis S, Steinhoff C, Swiatkowski S, Mahotka C, Gerharz C-D, Schulz WA. (2003) Decrease of DNA methyltransferase 1 expression relative to cell proliferation in transitional cell carcinoma. *Int. J. Cancer*, Vol. 104, pp. 568-578.

Koi M, Umar A, Chauhan DP, Cherian SP, Carethers JM, Kunkel TA, Boland CR. (1994) Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Research*, Vol. 54, pp. 4308-4312.

Kolosha VO, Martin SL. (1997) In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc Natl Acad Sci USA*, Vol. 94, pp. 10155-10160.

Kondo-Iida E, Kobayashi K, Watanabe M, Sasaki J, Kumagai T, Koide H, Saito K, Osawa M, Nakamura Y, Toda T. (1999) Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). *Hum Mol Genet*, Vol. 8, pp. 2303-2309.

Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM (2009) The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet* **5:** e1000458

Kubo S, Seleme MdC, Soifer HS, Perez JLG, Moran JV, Kazazian HH, Kasahara N. (2006) L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci USA*, Vol. 103, pp. 8036-8041.

Kuhfittig-Kulle S, Feldmann E, Odersky A, Kuliczkowska A, Goedecke W, Eggert A, Pfeiffer P. (2007) The mutagenic potential of non-homologous end joining in the absence of the NHEJ core factors Ku70/80, DNA-PKcs and XRCC4-LigIV. *Mutagenesis*, Vol. 22, pp. 217-233.

Kulpa DA, Moran JV. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum Mol Genet*, Vol. 14, pp. 3237-3248.

Kulpa DA, Moran JV. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*, Vol. 13, pp. 655-660.

Kurose K, Hata K, Hattori M, Sakaki Y. (1995) RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. *Nucleic Acids Res*, Vol. 23, pp. 3704-3709.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan

Lynch HT, de la Chapelle A. (1999) Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet*, Vol. 36, pp. 801-818.

Maestre J, Tchénio T, Dhellin O, Heidmann T. (1995) mRNA retroposition in human cells: processed pseudogene formation. *EMBO J*, Vol. 14, pp. 6333-6338.

Malik HS, Burke WD, Eickbush TH. (1999) The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution*, Vol. 16, pp. 793-805.

Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, Vol. 424, pp. 99-103.

Martin SL, Bushman FD. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol*, Vol. 21, pp. 467-475.

Matakidou A, el Galta R, Webb EL, Rudd MF, Bridle H, Consortium G, Eisen T, Houlston RS. (2007) Genetic variation in the DNA repair genes is predictive of outcome in lung cancer. *Hum Mol Genet*, Vol. 16, pp. 2333-2340.

Mätlik K, Redik K, Speek M. (2006) L1 Antisense Promoter Drives Tissue-Specific Transcription of Human Genes. *J. Biomed. Biotech.*, Vol. 2006, pp. 1-17.

McCulloch SD, Gu L, Li GM (2003) Bi-directional processing of DNA loops by mismatch repair-dependent and -independent pathways in human cells. *J Biol Chem* **278:** 3891-3896

McMillan JP, Singer MF. (1993) Translation of the human LINE-1 element, L1Hs. *Proc Natl Acad Sci USA*, Vol. 90, pp. 11533-11537.

Meischl C, Boer M, Ahlin A, Roos D. (2000) A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Eur J Hum Genet*, Vol. 8, pp. 697-703.

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

Minakami R, Kurose K, Etoh K, Furuhata Y, Hattori M, Sakaki Y. (1992) Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. *Nucleic Acids Research*, Vol. 20, pp. 3139-3145.

Moran JV. (1999) Exon Shuffling by L1 Retrotransposition. *Science*, Vol. 283, pp. 1530-1534.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, Vol. 87, pp. 917-927.

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet*, Vol. 31, pp. 159-165.

Morse B, Rotherg PG, South VJ, Spandorfer JM, Astrin SM. (1988) Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature*, Vol. 333, pp. 87-90.

Muckenfuss H, Hamdorf M, Held U, Perkovic M, Löwer J, Cichutek K, Flory E, Schumann GG, Münk C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem*, Vol. 281, pp. 22161-22172.

Muriaux D, Darlix J-L. (2010) Properties and functions of the nucleocapsid protein in virus assembly. *rnabiology*, Vol. 7, pp. 744-753.

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

Niewiadomska AM, Tian C, Tan L, Wang T, Sarkis PTN, Yu X-F. (2007) Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *Journal of Virology*, Vol. 81, pp. 9577-9583.

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

Ostertag EM. (2001) Twin Priming: A Proposed Mechanism for the Creation of Inversions in L1 Retrotransposition. *Genome Research*, Vol. 11, pp. 2059-2065.

Palombo F, Iaccarino I, Nakajima E, Ikejima M, Shimada T, Jiricny J. (1996) hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr Biol*, Vol. 6, pp. 1181-1184.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD. (1994) Mutation of a mutL homolog in hereditary colon cancer. *Science*, Vol. 263, pp. 1625-1629.

Parsons R, Li GM, Longley MJ, Fang WH, Papadopoulos N, Jen J, de la Chapelle A, Kinzler KW, Vogelstein B, Modrich P. (1993) Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell*, Vol. 75, pp. 1227-1236.

Peltomaki P. (2003) Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer. *Journal of Clinical Oncology*, Vol. 21, pp. 1174-1179.

Peltomäki P. (2001) DNA mismatch repair and cancer. *Mutat Res*, Vol. 488, pp. 77-85.

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

Pickeral OK, Makałowski W, Boguski MS, Boeke JD. (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Research*, Vol. 10, pp. 411-415.

Piskareva O, Schmatchenko V. (2006) DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. *FEBS Lett*, Vol. 580, pp. 661-668.

Plotz G, Raedle J, Brieger A, Trojan J, Zeuzem S (2002) hMutSalpha forms an ATP-dependent complex with hMutLalpha and hMutLbeta on DNA. *Nucleic Acids Res* **30:** 711-718

Prolla TA, Baker SM, Harris AC, Tsao JL, Yao X, Bronner CE, Zheng B, Gordon M, Reneker J, Arnheim N, Shibata D, Bradley A, Liskay RM. (1998) Tumour susceptibility

and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat Genet*, Vol. 18, pp. 276-279.

Prolla TA, Pang Q, Alani E, Kolodner RD, Liskay RM (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. *Science* **265:** 1091-1093

Qiu J, Qian Y, Chen V, Guan MX, Shen B (1999) Human exonuclease 1 functionally complements its yeast homologues in DNA recombination, RNA primer removal, and mutation avoidance. *J Biol Chem* **274:** 17893-17900

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

Repanas K, Zingler N, Layer LE, Schumann GG, Perrakis A, Weichenrieder O. (2007) Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. *Nucleic Acids Res*, Vol. 35, pp. 4914-4926.

Roy-Engel AM, El-Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H, Oyeniran OO, Deininger PL (2005) Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet Genome Res* **110:** 365-371

Saydam N, Kanagaraj R, Dietschy T, Garcia PL, Peña-Diaz J, Shevelev I, Stagljar I, Janscak P. (2007) Physical and functional interactions between Werner syndrome helicase and mismatch-repair initiation factors. *Nucleic Acids Research*, Vol. 35, pp. 5706-5716.

Schmutte C, Marinescu RC, Sadoff MM, Guerrette S, Overhauser J, Fishel R (1998) Human exonuclease I interacts with the mismatch repair protein hMSH2. *Cancer Res* **58:** 4537-4542

Schmutte C, Sadoff MM, Shim KS, Acharya S, Fishel R (2001) The interaction of DNA mismatch repair proteins with human exonuclease I. *J Biol Chem* **276:** 33011-33018

Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, van Duijnhoven G, Kirschner R, Hemberger M, Bergen AA, Rosenberg T, Pinckers AJ, Fundele R, Rosenthal A, Cremers FP, Ropers HH, Berger W. (1998) Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet*, Vol. 19, pp. 327-332.

Schwerdtle T, Ebert F, Thuy C, Richter C, Mullenders LH, Hartwig A (2010) Genotoxicity of soluble and particulate cadmium compounds: impact on oxidative DNA damage and nucleotide excision repair. *Chem Res Toxicol* **23:** 432-442

Shahi A, Lee J-H, Kang Y, Lee SH, Hyun J-W, Chang I-Y, Jun J-Y, You HJ. (2011) Mismatch-repair protein MSH6 is associated with Ku70 and regulates DNA double-strand break repair. *Nucleic Acids Research*, Vol. 39, pp. 2130-2143.

Shcherbakova PV, Hall MC, Lewis MS, Bennett SE, Martin KJ, Bushel PR, Afshari CA, Kunkel TA. (2001) Inactivation of DNA mismatch repair by increased expression of yeast MLH1. *Mol Cell Biol*, Vol. 21, pp. 940-951.

Shen Y, Koh KD, Weiss B, Storici F. (2011) Mispaired rNMPs in DNA are mutagenic and are targets of mismatch repair and RNases H. *Nat Struct Mol Biol*, pp. 1-8.

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

Spampinato CP, Gomez RL, Galles C, Lario LD. (2009) From bacteria to plants: a compendium of mismatch repair assays. *Mutat Res*, Vol. 682, pp. 110-128.

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

Suzuki J, Yamaguchi K, Kajikawa M, Ichiyanagi K, Adachi N, Koyama H, Takeda S, Okada N. (2009) Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet*, Vol. 5, p. e1000461.

Swergold GD. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol*, Vol. 10, pp. 6718-6729.

Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. (2002) Human l1 retrotransposition is associated with genetic instability in vivo. *Cell*, Vol. 110, pp. 327-338.

Szak ST, Pickeral OK, Landsman D, Boeke JD. (2003) Identifying related L1 retrotransposons by analyzing 3' transduced sequences. *Genome Biol*, Vol. 4, p. R30.

Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. (2002) Molecular archeology of L1 insertions in the human genome. *Genome Biol*, Vol. 3, p. research0052.

Tchénio T, Casella JF, Heidmann T. (2000) Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Research*, Vol. 28, pp. 411-415.

Thayer RE, Singer MF, Fanning TG. (1993) Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene*, Vol. 133, pp. 273-277.

Tishkoff DX, Boerger AL, Bertrand P, Filosi N, Gaida GM, Kane MF, Kolodner RD (1997) Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc Natl Acad Sci U S A* **94:** 7487-7492

Toft NJ, Winton DJ, Kelly J, Howard LA, Dekker M, te Riele H, Arends MJ, Wyllie AH, Margison GP, Clarke AR (1999) Msh2 status modulates both apoptosis and mutation frequency in the murine small intestine. *Proc Natl Acad Sci U S A* **96:** 3911-3915

Vo AT, Zhu F, Wu X, Yuan F, Gao Y, Gu L, Li G-M, Lee T-H, Her C. (2005) hMRE11 deficiency leads to microsatellite instability and defective DNA mismatch repair. *EMBO Rep*, Vol. 6, pp. 438-444.

Wallace NA, Belancio VP, Deininger PL. (2008) L1 mobile element expression causes multiple types of toxicity. *Gene*, Vol. 419, pp. 75-81.

Wallace NA, Belancio VP, Faber Z, Deininger P. (2010) Feedback inhibition of L1 and alu retrotransposition through altered double strand break repair kinetics. *Mob DNA*, Vol. 1, p. 22.

Weber MJ (2006) Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet* **2:** e205

Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol*, Vol. 21, pp. 1429-1439.

Wieland M, Levin MK, Hingorani KS, Biro FN, Hingorani MM. (2009) Mechanism of Cadmium-Mediated Inhibition of Msh2-Msh6 Function in DNA Mismatch Repair. *Biochemistry*, Vol. 48, pp. 9492-9502.

Wissing S, Munoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-Perez JL, Moran JV, Greene WC. (2011) Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Genet*, pp. 1-11.

Woodcock DM, Lawler CB, Linsenmeyer ME, Doherty JP, Warren WD. (1997) Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J Biol Chem*, Vol. 272, pp. 7810-7816.

Wu J, Gu L, Wang H, Geacintov NE, Li GM (1999) Mismatch repair processing of carcinogen-DNA adducts triggers apoptosis. *Mol Cell Biol* **19:** 8292-8301

Xiong Y, Eickbush TH. (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Molecular Biology and Evolution*, Vol. 5, pp. 675-690.

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

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

Zeng X, Kinsella TJ. (2007) A novel role for DNA mismatch repair and the autophagic processing of chemotherapy drugs in human tumor cells. *Autophagy*, Vol. 3, pp. 368-370.

Zhang H, Richards B, Wilson T, Lloyd M, Cranston A, Thorburn A, Fishel R, Meuth M. (1999) Apoptosis induced by overexpression of hMSH2 or hMLH1. *Cancer Research*, Vol. 59, pp. 3021-3027.

**Biography**

Zachary J. Faber is the son of John and Peg Faber. He attended Adrian High School in Adrian, MN and completed his undergraduate degree at St. John's University in Collegeville, MN. After completing a BA in Biology at SJU he accepted an offer to enroll in the Molecular and Cellular Biology Graduate Program at Tulane University in the fall of 2004. A year later, on the day he was to join the lab of Dr. Prescott Deininger, Hurricane Katrina struck the Gulf Coast. The University of Pennsylvania graciously offered a position in their Cell and Molecular Biology group in the Biomedical Graduate Studies Program as a visiting scholar. During his time at UPenn, Zach was able to continue his course work and gain valuable experience of Dr. Nina Luning Prak. Upon his return to Tulane after one semester at UPenn, the Molecular and Cellular Biology program had been merged with other graduate programs to form a new, interdisciplinary program named Biomedical Sciences under the School of Medicine. He returned to the Deininger Lab and completed his PhD in the field of mobile elements, studying the cellular defense mechanisms to LINE-1 and Alu.