A Mechanistic Examination of APOBEC3-Mediated LINE-1 Inhibition

by

Sandra Rose Richardson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Human Genetics)
in The University of Michigan
2013

Doctoral Committee:

Professor John V. Moran, Chair
Assistant Professor Raymond C. Chan
Professor Thomas M. Glaser
Professor K. Sue O'Shea
Professor Diane M. Robins

With love and gratitude, I dedicate this thesis to my husband, Pete Tracey.

Table of Contents

List of Tables

List of Figures

Abstract

A Mechanistic Examination of APOBEC3-Mediated LINE-1 Inhibition

By

Sandra R. Richardson

Chair: John V. Moran

Protein-coding sequences account for ~3% of the human genome, and a typical gene resides permanently at a discrete chromosomal address. The human genome, however, is not simply a static catalogue of genes; in many ways, it is an ever-changing entity. One dynamic component of the human genome is transposable elements (TEs), or “jumping genes”. Long Interspersed Element-1 (LINE-1 or L1) is a TE whose sequences make up ~17% of human DNA. Although most L1s are inactive, a few retain the ability to mobilize by a process called retrotransposition. In a round of retrotransposition, an L1 in the genome is transcribed into RNA, and translated to produce the protein machinery essential for generating a new insertion. The L1-encoded proteins bind to the RNA from which they were translated to form an L1 ribonucleoprotein particle (RNP), enter the nucleus, and cleave the genomic DNA in a new location to

expose a free DNA end, which the L1 machinery uses to prime synthesis of a DNA copy of its associated RNA. This process is termed target-primed reverse transcription (TPRT). The new DNA copy of the L1 becomes integrated into the genome, giving rise to a novel insertion.

L1 is often regarded as a molecular parasite that can be damaging to the host. In fact, ~70 cases of human disease have been attributed to L1-mediated retrotransposition events, and it stands to reason that humans have evolved ways to curtail L1 mobility. The human APOBEC3 (A3) family of cytidine deaminases represents a component of innate immunity hypothesized to have evolved to restrict TE mobility. Interestingly, it has been suggested that the A3 family further evolved and expanded to combat retroviruses such as HIV, which are relative evolutionary "newcomers" that share similarities with TEs in the ways that they challenge the genome. The mechanisms by which A3 proteins combat exogenous and endogenous threats are the subject of ongoing study.

In this thesis, I undertake a mechanistic examination of A3-mediated L1 inhibition. I find that human A3 proteins can restrict retrotransposons from different species, which bear little sequence identity to human L1s. I therefore hypothesize that A3-mediated restriction targets conserved steps or intermediates in the retrotransposition pathway, rather than relying on sequence-dependent recognition of retroelements. Indeed, I elucidate a mechanism of L1 inhibition by APOBEC3A (A3A), in which A3A edits transiently exposed single-stranded DNA during L1 TPRT. In concert with the action of cellular repair factors which degrade deaminated TPRT intermediates, I conclude that this

editing is at least partially responsible for A3A-mediated L1 inhibition. This work represents the first mechanistic explanation for inhibition of an autonomous retroelement by an A3 cytidine deaminase, providing new insight into the dynamic interplay between endogenous retrotransposons and host genomes.

Chapter 1

Retrotransposon Control in the Mammalian Germline and Early Embryo

Abstract

Transposable elements (TEs), once disparaged as “junk DNA,” are now recognized as a biologically and evolutionarily significant component of mammalian genomes. The known active TEs in the mouse and human genomes are retrotransposons, which mobilize through an RNA intermediate by a process termed retrotransposition. Unlike retroviruses, which also mobilize through an RNA intermediate, retrotransposons are endogenous elements that generally lack the capacity for transmission between individuals, and consequently rely on a vertical mode of transmission. In order to generate heritable insertions, retrotransposons must mobilize in cells that will contribute to the next generation. Key opportunities for retrotransposon mobility arise during developmentally critical rounds of genome-wide epigenetic reprogramming in the early embryo, prior to germline specification, and in the primordial germline. Unchecked, heritable retrotransposition events are potentially deleterious to the host. Thus, mammalian genomes have evolved a battery of defense mechanisms against retrotransposition. This chapter provides a brief background on retroelement

biology followed by a discussion of evidence for retrotransposon activity during early mammalian embryo development and in the germline, as well as host defenses that may limit the generation of heritable retrotransposition events.

Mobile DNA in Modern Genomes

Genomes are not static catalogues of genes; indeed, they are ever-changing entities. Transposable elements (TEs), or "jumping genes", represent a dynamic component of genomes. While a typical cellular gene resides at a discrete locus, TEs are able to mobilize to new chromosomal loci [1]. This activity has led to a remarkable accumulation of TE sequences in mammalian genomes: at least 45% of human DNA and at least 37% of mouse DNA is derived from TEs [2, 3]. Most of these TE insertions have been rendered unable to mobilize through the accumulation of mutations over time [2]. However, a growing body of work has revealed that modern mammalian genomes contain intact elements with the potential to mobilize (For review, see [4]). The classification and characteristics of TEs presently active in human and mouse genomes is discussed below.

Retrotransposons are Active in Mammalian Genomes

Transposable elements that mobilize through reverse transcription of an RNA intermediate are termed retrotransposons [5]. Retrotransposons are further categorized based on whether they possess long terminal repeats (LTRs) as LTR retrotransposons or non-LTR retrotransposons (Figure 1.1) (Reviewed in [6, 7]).

LTR Retrotransposons

LTR retrotransposons replicate through the reverse transcription of an RNA intermediate [8-10]. Endogenous LTR retrotransposons and exogenous retroviruses are closely related, and share a long, intertwined evolutionary relationship [11]. It is unclear whether retroviruses originally arose from endogenous LTR retroelements, or if endogenous LTR retrotransposons originated from germline colonization by exogenous retroelements [5]. Indeed, both scenarios have likely occurred: some retroviruses are hypothesized to have evolved from gypsy-like LTR retrotransposons through the acquisition of genes and regulatory sequences which facilitate a viral lifestyle [11-13]. Paradoxically, exogenous retroviruses may have also given rise to a subset of endogenous LTR retrotransposons: the LTR retrotransposon superfamily of endogenous retroviruses (ERVs), are “domesticated” retroviruses which have become endogenized via germline colonization and the inactivation or loss of genes required for an extracellular infectious lifestyle [14].

LTR retrotransposon sequences in the human genome appear to be completely inactive [2]. Intriguingly, a few human endogenous retrovirus (HERV) insertions are polymorphic with respect to presence among individuals, suggesting that they retrotransposed relatively recently during human evolution [15-18]. In contrast, the mouse genome contains multiple active families of LTR retrotransposons (reviewed in [19]).

Endogenous retroviruses contain direct repeats (LTRs), which flank genes encoding structural and enzymatic functions necessary for retrotransposition. These include the GAG ORF, which encodes the structural proteins for virus-like particle (VLP) formation, and the POL ORF, which contains protease, reverse transcriptase, RNaseH, and integrase activities (Figure 1.1 A) (Reviewed in [11, 20, 21]). Some murine ERVs also contain envelope genes [22].

Mouse endogenous retroviruses are subdivided into three classes based on their similarity to genera of exogenous retroviruses (Figure 1.1 A). Class I contains elements which cluster with gamma- and epsilonretrovirus; active members include the type C murine leukemia virus (MuLV), murine retroviral-related sequences (MuRRS), GLN (named for its glutamine tRNA primer-binding site), and virus-like 30S (VL30) endogenous retroviruses [19].

Active members of Class II, which cluster with lentivirus, alpha-, beta-, and delta-retroviruses, include mouse mammary tumor virus (MMTV) as well as intracisternal A particle (IAP) and the mouse endogenous type D provirus (MusD) along with its internally deleted non-autonomous counterpart mouse early transposon (ETn) [23-26]. IAP and ETn/MusD elements are highly active and responsible for most of the endogenous retrovirus-induced germline mutations in mouse [19, 22].

Presently active Class III ERVs, distinguished by their relationship to spumaviruses, include the murine endogenous retrovirus (MuERV_L) and the internally deleted, non-autonomous mammalian apparent LTR retrotransposons (MaLR) [19].

LTR retrotransposon mobilization closely resembles that of retroviruses, except that the LTR retrotransposon life-cycle is strictly intracellular [27]. Mobilization begins with transcription of the element by RNA polymerase II from a promoter located within the 5' LTR, and polyadenylation is directed by a polyadenylation signal located in the 3' UTR [20]. The resultant mRNA is translated, giving rise to gag proteins, as well as the proteins containing reverse transcriptase and integrase activities [28, 29]. Gag proteins assemble to form the virus-like particle (VLP), which encapsulates the element mRNA, as well as the proteins required for generation of a new insertion [30]. Owing to their lack of a functional *env* gene, these viral particles remain intracellular and are not infectious. The element-encoded reverse transcriptase activity carries out reverse transcription of the mRNA in the cytoplasm, giving rise to a double-stranded cDNA (Reviewed in [31]). The element-encoded integrase activity facilitates integration of the cDNA into a new location in the host genome [10, 32].

Non-LTR Retrotransposons

Non-LTR retrotransposons are found throughout eukaryotes, and Long INterspersed Element 1 (LINE-1 or L1) non-LTR retrotransposons represent the only known active autonomous retroelements in humans [33-35]. LINE-1 sequences account for about 17% of human DNA (Figure 1.1 B) [2]. The vast majority of L1s in the human genome have been rendered inactive by 5' truncations, rearrangements, and the accumulation of mutations [2, 36-38]; however, ~80-100 L1s per individual retain the ability to mobilize [39, 40]. L1

sequences similarly comprise ~18% of the mouse genome [3]; however, in contrast to humans, mice are estimated to contain approximately 3,000 retrotransposition-competent L1s per individual [41].

A full-length human L1 is approximately 6 kb [33, 42], and begins with a 5'UTR harboring an internal promoter [43]; the human L1 5'UTR also contains an antisense promoter of unknown function (Figure 1.1 B) [44]. L1s encode two open reading frames, and end in a 3'UTR and a poly-A tail [33, 42]. L1s in the genome are typically flanked by variable-length target-site duplications (TSDs), which are structural hallmarks of the integration process [45, 46] (Figure 1.1 B).

LINE-1 retrotransposition (Figure 1.2) begins with transcription of a full-length element from its internal promoter located within the 5'UTR [43, 47-52]. The resulting bicistronic mRNA is translated, giving rise to two proteins, ORF1p and ORF2p. ORF1p is a ~40 kD protein that is essential for retrotransposition [34, 53, 54]. ORF1p possesses nucleic acid binding activity mediated by its central RNA recognition motif and carboxyl-terminal basic domain [34, 55-60]. Mouse and human ORF1p also contain an N-terminal coiled-coil domain, which facilitates ORF1p trimer formation [55, 59, 61]. Mouse ORF1p is reported to possess nucleic acid chaperone activity [62].

Human ORF2p is a ~150 kD protein that may be translated by an unconventional termination-reinitiation mechanism, which can occur independently of the ORF2 AUG initiation codon [63-66]. ORF2p contains endonuclease (EN) and reverse transcriptase (RT) activities, both of which are critical for retrotransposition [34, 67, 68]. ORF1p and ORF2p proteins bind to

their encoding mRNA in a phenomenon known as *cis*-preference, giving rise to the hypothesized L1 ribonucleoprotein particle (RNP) retrotransposition intermediate {Esnault, 2000 #36;Kulpa, 2005 #50;Kulpa, 2006 #57;Wei, 2001 #38;Hohjoh, 1996 #14;Martin, 1991 #51}. The RNP then enters the nucleus by a mechanism which is not completely understood, but which can occur independently of nuclear envelope breakdown during cell division [69].

Upon gaining access to the genome, the L1 EN activity nicks the genomic DNA, liberating a free 3' hydroxyl residue [67, 70, 71]. This 3' hydroxyl residue provides a primer from which the L1 RT activity can initiate reverse transcription of the L1 mRNA, giving rise to the first-strand L1 cDNA [67, 72, 73]. This process is termed target-primed reverse transcription (TPRT), based on models derived from work on the R2 element from *Bombyx mori* [74]. Second-strand DNA cleavage generally occurs some distance downstream of the initial single-strand endonucleolytic nick, facilitating second-strand DNA synthesis and leading to the generation of variable-length TSDs flanking the new L1 insertion [75]. The mechanistic details of second-strand DNA synthesis and integration are incompletely understood. Notably, *in vitro* studies on the R2 of *Bombyx mori* indicate that two molecules of the R2 protein, which encodes EN and RT activities, participate in the generation of new insertions. The first subunit, which binds upstream of the endonuclease cleavage site, is responsible for first-strand DNA cleavage and the initiation of reverse transcription of the R2 RNA template [76]. The second subunit, which binds downstream of the first, then carries out second-strand cleavage [76]; subsequently, the upstream subunit is

hypothesized to carry out second-strand DNA synthesis [76]. Furthermore, the R2 protein has been demonstrated *in vitro* to possess 5'-to-3' RNA displacement activity concurrent with second-strand DNA synthesis, suggesting that an RNA-DNA heteroduplex could persist until second-strand DNA synthesis, obviating the need for a cellular RNaseH activity to degrade the R2 RNA template [77]. As techniques and reagents become available to study L1 TPRT *in vitro*, it will be intriguing to discover whether L1 and R2 share similar mechanisms in the latter steps of TPRT and integration.

Although *cis*-preference is the prevailing rule in template selection by the L1 enzymatic machinery, the L1 encoded proteins also can mobilize non-autonomous retrotransposons *in trans*, such as the human short interspersed element (SINE) Alu and the mouse SINEs B1 and B2 [78, 79]. The human SINE-R/VNTR/Alu (SVA) composite element also is likely mobilized in *trans* by the L1-encoded machinery [80-84]. In addition, the L1-encoded proteins can occasionally mobilize cellular mRNAs, giving rise to processed pseudogenes [85, 86]. In sum, L1-mediated retrotransposition events are responsible for at least one third of human DNA (Figure 1.1 B) [2].

Transposable Elements and their Host Genomes

TEs share ancient and complex relationships with their host genomes, and while TE mobilization is typically neutral or detrimental to host fitness, to dismiss TEs as purely harmful or useless is an oversimplification. Indeed, in some cases, TEs can be viewed as inadvertently providing the raw material for host

genome evolution, their sequences becoming exploited by the host genome to serve a beneficial purpose. For example, the RAG1 and RAG2 endonucleases, vital components of VDJ recombination in the adaptive immune system, were likely derived from a DNA transposon [87-90]. Likewise, in *Drosophila*, the TEs Het-A, TART, and TAHRE function in place of a conventional telomerase to maintain chromosome length throughout rounds of cell division [91-95]. Furthermore, *Penelope*-like retroelements lacking endonuclease domains have recently been discovered to localize to telomeres in species representing four different eukaryotic kingdoms; these elements may represent an evolutionary link between retroelements and modern-day telomerases [96]. Recent work in mouse embryonic stem cells (mESCs) suggests that mouse endogenous retrovirus (MuERV) LTRs drive transcription of genes which regulate a totipotent state resembling the 2-cell stage embryo [97, 98]. Individual cells within mESC cultures cycle in and out of this totipotent state, which appears to be important for the sustainability of mESC cultures, as well as the ability of mESCs to undergo correct cell fate specification. Therefore, retroelement regulatory sequences may provide critical developmental functions in the early mammalian embryo [98].

In many cases, strategies for TE mobility may have evolved to be relatively innocuous to the host: certain TEs have specific target-site preferences that incur minimal risk to the host genome. For example, the non-LTR retrotransposon R2 of *Bombyx mori* encodes a sequence-specific endonuclease, directing its mobility to a unique sequence within the 28S rRNA gene [99].

Similarly, the yeast LTR retrotransposon *Ty1* and *Ty3* are targeted specifically to regions upstream of RNA Pol III transcription initiation sites, in a manner dependent on Pol III transcription [100-103]. In contrast, the yeast LTR retrotransposon *Ty5* is specifically targeted to silent chromatin [104-106], but can mobilize to other genomic locations under conditions of genomic stress [107]. In sum, retroelement targeting to specific integration sites may be beneficial to the TE and perhaps the host, as the propagation of these endogenous elements ultimately hinges on host fitness (for review, see [108]).

Contributions to genome evolution aside, unchecked TE mobilization represents a substantial threat to host fitness and genome stability. For example, approximately 96 cases of human disease have been associated with L1-mediated retrotransposition events [75, 109-111]. Furthermore, 10-12% of spontaneous mouse mutations are estimated to arise from non-LTR and LTR retrotransposon activity [22]. This considerable burden highlights the ongoing conflict between retrotransposons and their host genomes, in which retrotransposons strive to ensure their propagation to subsequent generations, and the host genome must defend itself from retrotransposition-mediated mutations.

Opportunities for Heritable Insertions in Early Development

Endogenous retrotransposons primarily rely on a vertical mode of transmission, as they generally lack the capacity for transmission between individuals. Therefore, in order to successfully propagate, retrotransposons must

mobilize in cells that have the potential to contribute to the next generation. In mammals, one of these critical developmental windows occurs in the pluripotent cells of the very early embryo, before the germ lineage is specified. If a retrotransposition event occurs at this stage, there is a chance that the cell in which it occurs could contribute to the germline, rendering the insertion heritable by subsequent generations (Figure 1.3) (Reviewed in [108, 112]). Alternately, *de novo* heritable events may occur within the germline, after it has diverged from the somatic lineage (Reviewed in [108, 112]). Thus, in principle, heritable insertions could occur at any point in germline development, from primordial germ cells in the developing embryo to mature haploid gametes.

Erasure and Establishment of DNA Methylation: "Windows of Opportunity" for Retrotransposon Mobilization

In most tissues and developmental stages, retrotransposons are kept transcriptionally quiescent by DNA methylation [113]. In fact, DNA methylation is proposed to have evolved to restrict TE activity [114]. However, the temporal windows where retrotransposons can ensure heritable insertions are also the developmental stages (the pre-implantation embryo and primordial germline) where the genome undergoes profound epigenetic reprogramming essential for proper development (Figure 1.4, discussed further below) [115]. This reprogramming includes the erasure and subsequent reestablishment of DNA methylation (reviewed in [116]). Such transient epigenetic relaxation may

release retrotransposons from transcriptional repression, enabling a permissive cellular environment for *de novo* retrotransposition events (Reviewed in [117]).

Methylation Erasure and Re-establishment in the Early Embryo

A round of genome-wide epigenetic reprogramming takes place in mammalian pre-implantation embryos (Figure 1.4 A). In mice, where this process has been extensively studied, the paternal genome in the zygote undergoes rapid demethylation shortly after fertilization, and before replicating its DNA [118-121]. From the one-cell stage until the morula/early blastocyst stage, during cleavage divisions, both the paternal and maternal genomes undergo passive demethylation, in which DNA methylation marks are not copied onto newly-synthesized DNA strands during rounds of cell division [122-128].

Retroelements in the mouse genome are differentially demethylated during reprogramming in the early embryo: IAP elements (as well as imprinted genes) remain heavily methylated during this period, while L1 elements undergo extensive demethylation [122, 124]. The maternally-inherited short form of the maintenance DNA methyltransferase, DNMT1o, has been demonstrated to regulate IAP methylation in the early embryo [129]; however, it remains unknown how IAP sequences, but not L1 sequences, are specifically targeted for maintenance of DNA methylation.

Methylation is re-established around the time of implantation [125] by the *de novo* DNA methyltransferases Dnmt3a and Dnmt3b [127]. During subsequent rounds of cell division, the maintenance DNA methyltransferase Dnmt1 is

responsible for copying methylation marks from the template strand to the newly-synthesized daughter strand [130].

Methylation Erasure and Re-establishment in the Primordial Germline

In mice, a second round of genome-wide reprogramming, including retroelement demethylation, X-reactivation, and imprint erasure, occurs in the primordial germline of the developing fetus (Figure 1.4 B). Mouse germline development begins with the appearance of the primordial germ cells (PGCs), which arise from epiblast cells. The PGCs arise in the posterior epiblast at 7.25 days post-coitum (dpc) [131] (Figure 1.4 B). PGCs proliferate and migrate to the genital ridge at 8.5 dpc [132]. Migrating PGCs show random X-inactivation, imprinting, and high levels of DNA methylation; however, following entry into the genital ridge, they undergo genome-wide demethylation which is completed by 13-14 dpc [116, 125, 133-136].

Retroelement sequences are differentially demethylated during primordial germline epigenetic reprogramming: IAP sequences partially retain DNA methylation, while LINE-1 sequences are more extensively demethylated [124, 134]. Following the completion of this round of demethylation, male germ cells enter mitotic arrest, and female germ cells enter meiotic prophase. Establishment of *de novo* methylation in the male germline begins at 13.5 dpc and is completed around birth, before re-entry of male germ cells into mitosis [116, 125, 133-136].

DNA methylation of TEs in primordial germ cells is carried out by the *de novo* DNA methyltransferases Dnmt3a and Dnmt3b [127]. In addition, the non-catalytic paralog DNA methyltransferase 3-like (Dnmt3L) is critical for Dnmt3A-mediated methylation of L1 and IAP elements [137, 138]. Dnmt3L has been demonstrated to stabilize the active conformation of Dnmt3A, allowing more efficient transfer of methyl groups onto target DNA sequences [139, 140]. Analysis of fetal prospermatagonia revealed that in mice, SINE-B1 repeats are methylated by DNMT3a, while IAP and L1 elements are methylated by both Dnmt3a and Dnmt3b [137].

Intriguingly, Dnmt3L-deficient mice manifest a meiotic phenotype--failure of synapsis at the zygotene/pachytene transition—after the actual window of Dnmt3L expression has elapsed [138]. This meiotic defect results in sterility with a complete lack of germ cells in adult males [141]. Furthermore, LINE-1 and IAP elements in Dnmt3L-deficient germ cells are highly demethylated relative to wild-type [137, 138]. This lack of methylation of L1 and IAP repeats in Dnmt3L-deficient testes is accompanied by a high level of retroelement transcriptional reactivation. Aberrant transposon upregulation and activity has been suggested to contribute to meiotic catastrophe [138]. However, the mechanistic link between loss of transposon methylation, transcriptional reactivation and meiotic catastrophe in Dnmt3L-deficient male germ cells is unclear and requires further investigation. Likewise, much remains to be elucidated regarding the mechanistic details of Dnmt3L targeting to TE sequences.

Developmental timing and location of heritable L1 retrotransposition events

Potential windows for TE mobilization occur in the developing germline and early embryo; however, it is currently unclear where and when the majority of heritable retrotransposition events actually occur. In this section, human case studies, experimental animal models, and cell culture systems which have provided insight about the developmental timing of heritable L1 retrotransposition events are presented.

Evidence for L1 retrotransposition in the germline

A 2002 study by Brouha *et al.* [142] examined the case of a male patient afflicted with chronic granulomatous disease (CGD) caused by mutation of the X-linked gene *CYBB*, which encodes the gp91phox subunit of cytochrome b_{558}. Characterization of the patient's *CYBB* gene revealed an exonic L1 insertion, not present in the patient's mother, that consisted of ~1720 bp of L1 sequence and a 280 bp 3' transduction. The 3' transduced sequence allowed identification of the likely precursor L1 as LRE3, located on chromosome 2q24.1 within a region of imperfect "GAAA" repeats. Notably, the patient's mother was homozygous for the precursor LRE3 element on chromosome 2; however, one of her LRE3 alleles was followed by 28 GAAA repeats, and the other was followed by 29 GAAA repeats. The patient was heterozygous for the chromosome 2 LRE3 precursor element. Furthermore, his chromosome 2 LRE3 allele was followed by 28 GAAA repeats, yet the disease-causing LRE3 insertion into *CYBB* on his X-

chromosome contained 29 GAAA repeats in the 3' transduced sequence. Thus, the insertion within *CYBB* must have come from an *LRE3* allele not inherited by the patient. The explanation favored by the authors is that RNA from the mother's *LRE3* allele, flanked by 29 GAAA repeats, retrotransposed in the female germline into *CYBB* on an X-chromosome prior to the onset of meiosis II. The *de novo* insertion then segregated away from the donor LRE3 allele during meiosis II [142]. Although this model adequately explains the data, subsequent studies have provided possible reinterpretations regarding the developmental timing of the LRE3 retrotransposition event (see below).

Evidence from a transgenic mouse model suggests that engineered L1 retrotransposition events can occur in the male germline [143]. In a study by Ostertag *et al*., a retrotransposition-competent L1 transgene was engineered to contain an EGFP reporter under control of an acrosin promoter and tagged with an acrosin signal peptide, allowing EGFP to be expressed and localize to the acrosome in sperm [143, 144]. One line of transgenic mice produced faint green sperm, indicating successful retrotransposition in the male germline. In addition, PCR assays indicated that approximately 1/100 developing spermatids contained a retrotransposition event [143]. When male mice harboring the L1 transgene were bred to wild-type females, 2/135 resultant offspring (F_2 generation) contained engineered L1 retrotransposition events, one of which was inherited by the subsequent (F_3) generation independently of the transgene [143]. This result is consistent with L1 mobilization in the male germline prior to the onset of meiosis II, allowing the *de novo* L1 insertion to segregate away from the

chromosome harboring the L1 transgene [143, 144]. However, similar to the Brouha *et al.* study [142], more recent data provide alternative explanations for the timing of this retrotransposition event (see below).

In contrast to the findings of Ostertag *et al.,* a recent effort to uncover evidence of endogenous, *de novo* L1 insertions in sperm arrived at a much lower estimate for the rate of L1 mobilization in the mammalian male germline [145]. Freeman *et al.* developed a technique termed L1 hybridization enrichment and examined human sperm DNA for *de novo* retrotransposition events at genomic loci previously demonstrated to accommodate L1 insertions in humans [145]. Although this technique is capable of single-molecule amplification of insertion-bearing target sites, no *de novo* insertions were detected among large quantities of sperm DNA. This result suggests that L1 insertion in the male germline is extremely rare, with a rate lower than one insertion per 400 haploid genomes [145]. Thus, the frequency with which L1 retrotransposition events occur in the male mammalian germline remains an open question.

Recent experimental evidence suggests that heritable retrotransposition events may take place in the female germline. Georgiou et al. [146] performed RT-PCR analysis using RNA obtained from human denucleated germinal vesicle (GV) oocytes, and detected expression of L1, HERV-K10, and SVA retrotransposon transcripts. To determine whether retroelement expression was accompanied by the capacity to accommodate retrotransposition events, GV oocytes were microinjected with a plasmid expressing a retrotransposition-competent L1 tagged with an EGFP indicator cassette, or a retrotransposition-

defective mutant L1 as a control. EGFP positive oocytes were detected in the RC-L1 injected population, but not in the negative control oocytes, suggesting that these cells are capable of harboring *de novo* retrotransposition events. The rate of retrotransposition reported in GV oocytes is remarkably high compared to what is reported for human HeLa and other cell types: of 30 GVs microinjected with an RC-L1 construct, 26 were EGFP positive, indicating that 86.6% of GVs receiving the marked L1 construct harbored retrotransposition events [146]. Thus, human oocytes may represent a permissive developmental stage for the generation of heritable retrotransposon insertions.

Evidence for L1 retrotransposition in the early embryo

Evidence for L1 retrotransposition in early human development *in vivo* has been uncovered through the study of a male patient afflicted with X-linked choroideremia [147, 148]. The patient harbored an L1 insertion into the *CHM* gene, resulting in aberrant splicing of the *CHM* transcript [148]. PCR analysis of lymphocyte DNA from the patient's mother revealed that she was a somatic mosaic with respect to this L1 insertion. Analysis of the patient and his two sisters showed that all three shared the same haplotype in the *CHM* region, but only the patient and one of his sisters bore the L1 insertion [147]. Thus, the mother was necessarily both a germline and somatic mosaic with respect to this L1 insertion. This striking result indicates that the retrotransposition event must have taken place very early in the mother's embryonic development, before the segregation of the germline and the soma.

Evidence that L1 retrotransposition can take place during embryogenesis has also been provided by studies in transgenic model organisms. In a study by Kano *et al.*, mice and rats harboring a marked retrotransposition-competent L1 transgene were occasionally observed to produce offspring with *de novo* L1 insertions derived from the transgene without transmitting the transgene itself [149]. A possible explanation for this result is that the retrotransposition event took place in the parental germline before the completion of meiosis I, followed by segregation of the transgene away from the new insertion during meiosis II, as proposed by Ostertag *et al.* [143]. If this was the case, the F_2 offspring inheriting the *de novo* insertion would be expected to harbor it in every cell, and would subsequently pass it on to half of their offspring (F_3). Unexpectedly, none of the insertion-bearing F_2 mice transmitted an insertion to the F_3 generation, indicating that the insertion was not present in their germ lineage. Therefore, the most parsimonious explanation is that transcription of the marked L1 transgene took place during gamete development in the original transgenic mouse. Instead of undergoing retrotransposition in the germline, it is proposed that the L1 RNA was carried over in the gamete through fertilization, perhaps protected by formation of the L1 ribonucleoprotein particle, and was subsequently able to carry out TPRT and generate a *de novo* insertion during embryogenesis in the offspring. Quantitative PCR analyses for retrotransposition events in spermatogenic cells and pre-implantation embryos corroborated these findings; evidence of retrotransposition of marked elements was much more prevalent in pre-implantation embryos than spermatogenic cells [149].

It is interesting to note that in the Ostertag *et al*. study, described above, inheritance of an engineered L1 insertion independently of the L1 transgene was interpreted as evidence for L1 retrotransposition in the germ line prior to meiosis II [143]. In addition, inheritance of a *de* novo L1 insertion independently of the precursor element by a human patient was interpreted by Brouha *et al.* [142] as evidence of L1 retrotransposition in the female germline prior to meiosis II. However, as elucidated by the Kano *et al.* study, both results are also consistent with carry-over of an L1 RNP in a gamete, and retrotransposition post-fertilization in the early embryo [149]. Nevertheless, detection by Ostertag *et al.* of sperm harboring engineered L1 retrotransposition events, as evidenced by EGFP expression, demonstrates in principle that L1 retrotransposition can occur in the male germline [143].

Cell culture models provide additional support for the hypothesis that L1 retrotransposition can occur in the early embryo. Human embryonic stem cells (hESCs), isolated from the inner cell mass (ICM) of human embryos at the blastocyst stage, represent a relevant model for studying retrotransposition during this early embryonic development [150]. Supporting the notion that early human embryos may accommodate L1 retrotransposition, L1 ribonucleoprotein particles are highly expressed in hESCs, and a significant proportion of expressed L1Hs elements in cultured hESCs are retrotransposition-competent [151, 152]. Furthermore, it has been demonstrated that engineered L1 constructs tagged with a retrotransposition indicator cassette can mobilize in

cultured hESCs, albeit at low levels when compared to L1 mobilization in transformed cell types such as HeLa cells [34, 151, 153].

To identify potentially active retroelements in hESCs, Macia *et al.* [152] examined the cohort of transposable elements expressed in hESCs. By cloning and sequencing expressed Alu elements, and testing a sample of these elements in a cultured-cell retrotransposition assay, they determined that hESCs express a wide variety of Alu elements from different subfamilies.

To identify potentially expressed L1 elements in the genome, Macia *et al.* [152] took advantage of the L1 antisense promoter of genomic L1 elements [44]. Transcripts originating from genomic L1 antisense promoters were cloned in a 3' RACE strategy, and mapped to the genome to identify potentially transcriptionally active genomic L1s. Notably, this strategy is based on the assumption that expression from an L1 antisense promoter indicates that the L1 sense promoter is active as well [152]. Their results reveal that ~30% of human-specific L1 elements expressed in hESCs correspond to known retrotransposition-competent L1s. Furthermore, L1 antisense promoter activity from genomic loci lacking a previously-annotated L1 insertion in the human genome working draft sequence was predictive of polymorphic insertions present in hESC lines. The robust expression of endogenous retrotransposition-competent L1s in hESCs, combined with the ability of hESCs to accommodate the retrotransposition of tagged transfected elements [151], support the idea that heritable retrotransposition events can occur in the early embryo.

The case for L1 mobilization in pluripotent cells is strengthened by recent work in induced pluripotent stem cells (iPSCs)[154]. These cells, which are generated by the introduction of defined transcription factors (Oct3/4, Sox2, c-Myc Klf4 [155], or Oct3/4, Sox2, NANOG, Lin28 [156]) into cultured somatic cells such as fibroblasts, share a similar transcriptional profile and pluripotent characteristics with embryonic stem cells [155-157]. L1 mRNA and L1 ORF1p expression were observed at very low levels in parental fibroblasts, and increased to levels comparable to those observed in hESCs, as verified by Northern blotting, upon reprogramming of those fibroblasts into iPSCs. Furthermore, transcriptional upregulation of L1 elements upon reprogramming correlated with a decrease in L1 promoter methylation, particularly for known “hot” L1s [154]. These results suggest that pluripotent iPSCs, resembling the state found in the very early embryo, may be permissive for L1 mRNA and ORF1p expression, which could lead to endogenous retrotransposition events. Induced pluripotent stem cells also accommodated higher levels of retrotransposition than parental fibroblasts, when measured in an assay employing a marked L1 retrotransposition indicator provided as a transfected episomal plasmid [154].

Somatic retrotransposition

As discussed above, methylation patterns on retroelements are established during early development, and are maintained throughout somatic tissues. Recent studies, however, have uncovered evidence of retrotransposition in certain somatic cell types. Rat neuronal precursor cells overexpress

endogenous L1 transcripts and accommodate retrotransposition of a transfected engineered L1 equipped with a reporter cassette [158]. Furthermore, transgenic mice harboring a marked L1 transgene under the control of the native L1 5'UTR promoter can accommodate engineered L1 insertions consistent with L1 retrotransposition events occurring during embryonic and adult neurogenesis [158]. Similarly, human fetal brain stem cells and neural progenitor cells derived from human embryonic stem cells can accommodate retrotransposition of an engineered reporter L1 [159, 160]. Notably, fetal brain samples were found to be hypomethylated when compared to matched skin samples, suggesting that L1 expression may be released from methylation-dependent suppression in neuronal tissues [159]. Finally, quantitative multiplex PCR analysis revealed that L1 copy number may be higher in brain than in other somatic tissues [159]. However, observed increases in L1 copy number without the characterization of *bona fide* novel L1 insertions do not conclusively demonstrate *de novo* retrotransposition in brain. General genome instability resulting in duplication or deletion of L1 sequences in different tissues could also, in principle, explain such discrepancies in L1 copy number.

Recently, compelling evidence for L1 activity in the human brain was uncovered in a high-throughput search for novel L1, Alu, and SVA insertions in genomic DNA derived from brain tissues from three human individuals [161]. By a method termed retrotransposon capture sequencing (RC-seq), potentially novel (relative to known catalogued polymorphisms) L1 insertions were identified in DNA samples from brain and blood of the donors; reads found in both were

classified as germline insertions, while those present in brain only were regarded as potential somatic events. Only a minority of L1 insertions were designated as germline, suggesting that the majority of events observed represent genuine somatic insertions. The 5' ends of several insertions were confirmed by PCR, and 3' end validation was possible for three potential somatic insertions. Among these three insertions, one L1 and one Alu insertion were flanked by target-site duplications (TSDs) indicative of retrotransposition by target-primed reverse transcription. Thus, this study provides support for increased retrotransposon copy number in neuronal tissues resulting from *bona fide, de novo* retrotransposition events.

In addition to mobilization in normal neuronal cell types, a growing number of studies have reported *de novo* retrotransposon insertions in several different tumor types, including lung cancer [162], colorectal cancer [163, 164], and hepatocellular carcinoma [165]. Retroelement activity in tumors is correlated with decreased methylation on retroelement promoters [162, 165], suggesting that loss of methylation contributes to deregulated retrotransposition during tumorigenesis. It remains to be determined whether most retroelement insertions found in tumors represent drivers of tumorigenesis, or are merely passenger mutations occurring in a deregulated genomic environment (Reviewed in [108]).

Mechanisms for Restricting Heritable Retrotransposition Events

In this section, host mechanisms of defense against retroelement mobility are presented. These mechanisms include a small RNA-mediated silencing

pathway which functions in the male primordial germline, histone modifications in early embryonic cell types, and a cohort of cellular factors which have been demonstrated to restrict L1 mobilization in cultured cells.

PIWI/piRNA in the Mammalian Male Germline

Accumulating evidence has led to the proposal of an evolutionarily conserved adaptive defense mechanism against transposons in the mouse male germline. This pathway involves murine members of the PIWI clade of Argonaute proteins and their associated 26-31 nucleotide small RNAs, termed piRNAs (PIWI-interacting RNAs), which are proposed to guide PIWI proteins to silence mobile elements. Silencing is proposed to take place both through PIWI-mediated slicer cleavage of transposon transcripts, and piRNA-guided epigenetic silencing of elements in the genome.

The details of the PIWI-piRNA pathway in mammals are not completely understood, and prevailing models are largely built upon work in *Drosophila* [166, 167]. A pathway involving PIWI proteins and piRNAs, known as the “ping-pong” model, is active in *Drosophila* germ cells. In this pathway, piRNAs are derived from single-stranded RNAs transcribed from regions containing a high density of diverse transposon-derived sequences, termed piRNA clusters. Notably, the *Flamenco* locus had long been recognized as a master regulator of the endogenous retrovirus *gypsy* [168-171]), and only recently was identified as a piRNA cluster directing PIWI-mediated silencing of expressed *gypsy* elements in *Drosophila* ovarian somatic cells [167, 172].

Transposon transcripts derived from germline piRNA clusters are processed into individual piRNAs by an unknown mechanism. These genome-encoded piRNAs are termed primary piRNAs, and are usually antisense with respect to the transposon sequence [167]. Primary piRNAs have a bias for uracil at their 5' end (1U) [166, 167]. The Piwi protein Aubergine binds primary piRNAs, and the piRNA guides its associated Aubergine protein to a transposon mRNA target. Aubergine-mediated slicer cleavage and 3' end processing of this target gives rise to a new piRNA, termed a secondary piRNA, which is sense-oriented with respect to the element. Secondary piRNAs have a bias for adenine at the 10^{th} position from their 5' end (10A), a hallmark of primary piRNA-guided PIWI-mediated cleavage [166, 167] Such secondary piRNAs bind to the PIWI protein AGO3 and direct AGO3-mediated cleavage of antisense transposon RNAs, presumably derived from piRNA clusters [173]. This cycle gives rise to new antisense piRNAs, which can direct destruction of sense transposon RNAs, simultaneously combating existing transposon transcripts and fueling subsequent rounds of the cycle [166, 167].

In addition to post-transcriptional regulation, *Drosophila* PIWI proteins may have a role in epigenetic modification. Aubergine and PIWI are effectors of position-effect variegation (PEV), the phenomenon wherein a euchromatic gene becomes silenced when placed in proximity to a heterochromatic region of the genome (Reviewed in [174]). Loss of Aubergine or PIWI function relieves PEV-mediated silencing of certain genes apparently through the loss of heterochromatin protein 1a (HP1a) from the region [175]. HP1a and PIWI

physically interact, and PIWI localizes to the nucleus and associates with chromatin, strengthening the evidence for its role as an effector of epigenetic modifications [176].

The existence of a ping-pong pathway in mammals is currently under investigation. In the mouse male primordial germline, two PIWI clade proteins, MIWI2 and MILI, are expressed during the temporal window when DNA methylation is erased and reestablished [177, 178] (Figure 1.3). Deficiency in MIWI2 or MILI results in meiotic arrest at the pachytene stage of meiosis I, and correlates with an aberrant increase in L1 and IAP transcripts, as well as a failure to establish methylation on retroelements in the genome [179-182]. In addition, piRNAs expressed during this time, termed “pre-pachytene piRNAs”, bear signatures suggestive of a ping-pong type amplification mechanism, and are enriched for transposon-derived sequences [178, 180, 182]. Unlike *Drosophila*, where piRNAs are derived from transposon-dense piRNA clusters, transposon-derived mammalian pre-pachytene piRNAs appear to be derived from individual transposons scattered throughout the genome [178]. The potential participation of MIWI2 and MILI in a piRNA amplification pathway, particularly the roles of MIWI2 and MILI endonuclease activities [183], are currently areas of intense study. Indeed, MILI and MIWI2 are hypothesized to degrade transposon transcripts, fuelling a ping-pong amplification cycle, while also somehow targeting reestablishment of methylation to transposon sequences in a piRNA-directed manner [178]. However, evidence for a mechanistic link between PIWI/piRNA complexes and transposon methylation remains to be identified.

Histone modifications and retrotransposon control in the early embryo

Much of what is known regarding transposon control in the pre-implantation embryo comes from experiments in cultured pluripotent cell types. Indeed, evidence for a retrotransposon-specific silencing mechanism in pluripotent cells has been gained through experiments in human embryonic carcinoma (EC) cells. Human EC cells have a similar transcriptional profile to hESCs, and have been regarded as a relevant model for early human development [184]. Furthermore, like hESCs, EC cells express L1 transcripts and L1 ORF1 protein [55, 151, 185, 186]. L1 retrotransposition events from a marked indicator cassette have been demonstrated to occur in EC cells, but undergo rapid silencing of the L1-delivered reporter gene either during or immediately after retrotransposition [185]. Treatment with histone deacetylase inhibitors reactivated expression of the reporter gene; silencing and reactivation of the reporter gene were reversible with removal and addition of HDAC inhibitor [185]. This silencing appears to specifically affect reporter genes delivered by target-primed reverse transcription (TPRT): reporter cassettes delivered by LINE elements from mouse and zebrafish, which carry out TPRT, were efficiently silenced, while similar reporter genes delivered by human immunodeficiency virus (HIV) or Moloney murine leukemia retrovirus (MMLV), which undergo different mechanisms of insertion, were less efficiently silenced and exhibited only modest reactivation [185]. Importantly, retrotransposition events occurring in differentiating EC cells were not subject to efficient reporter gene silencing;

however, induced differentiation of cells already harboring silenced integration events was not sufficient to reactivate these silenced events [185]. Thus, nascent L1 integration events in pluripotent EC cells are hypothesized to be recognized by an unknown mechanism and efficiently silenced, while differentiating cells do not employ such a silencing mechanism. However, silencing of L1 retrotransposition events occurring during the pluripotent state appears to be maintained during differentiation [185]. Future studies will no doubt elucidate the nature of this epigenetic mechanism.

In mouse ES cells (mESCs), a variety of chromatin modifying enzymes and epigenetic marks have been demonstrated to repress repetitive sequences. A recent report demonstrated that the histone lysine methyltransferase SETDB1 (also known as ESET) is essential for silencing a subset of endogenous retroviruses in mESCs [187]. Upon conditional SETDB1 knockout, MusD, IAP, and endogenous MLV sequences were transcriptionally upregulated. Transcriptional upregulation was coordinated with loss of H3K9 trimethylation and H4K20 trimethylation on MusD, IAP, and MLV sequences.

The SETDB1 interacting partner KAP1 (also known as TRIM28) recruits SETDB1 as well as the NuRD histone deacetylase complex to effect transcriptional silencing of target genes [188]. The recruitment of KAP1 to target genes is mediated by Kruppel-associated box domain-zinc finger proteins (KRAB-ZFPs), which can recognize target genes in a sequence-specific manner [189, 190]. Indeed, loss of KAP1 results in substantial upregulation of endogenous retroviruses, particularly IAP elements [191]. SETDB1 removal

does not affect KAP1 recruitment to target genes, while KAP1 depletion causes increased ERV expression and a decrease in SETDB1 and H3K9 trimethylation at ERV sequences [187]. Thus, KAP1 apparently acts upstream of SETDB1 in ERV silencing in mESCs.

The role of DNA methylation in concert with histone lysine methylation in silencing ERVs in mESCs is not completely understood. Rowe et al. [191] found that treating mESCs with 5-azacytidine (5-aza), a drug which causes passive DNA demethylation, resulted in upregulation of IAP transcription that was synergistic with KAP1 knockout. By comparison, a microarray study by Karimi et al. [192], employed Dnmt1/Dnmt3a/Dnmt3b triple knock-out (DNMT TKO) mESCs and SETDB1-depleted mESCs, and demonstrated that DNA methylation and SETDB1/H3K9 trimethylation regulate predominantly non-overlapping sets of genes. A synergistic transcriptional upregulation upon Dnmt1 and SETDB1 double knock-down was only observed for the young IAPE-z subfamily of endogenous retroviruses. This result is consistent with earlier reports suggesting that DNMT TKO cell lines generally do not de-repress endogenous retroviruses [187, 193, 194]. Thus, while methylation may play a role in the regulation of specific, young ERV elements, the control of ERVs in mESCs appears to occur predominantly through histone modification. In differentiated cells, however, histone modifications are less important for ERV silencing, while DNA methylation takes over as the vital repressive mark [187, 195, 196].

Several additional chromatin modifying factors have been implicated in ERV silencing in mESCs. A recent microarray study has implicated HDAC1

(histone deacetylase 1) in endogenous retrovirus silencing in mESCs, as well as identifying retrotransposons subject to regulation by DNA methylation, SETDB1-mediated histone modification, and polycomb repressive complex (PRC) [197]. Indeed, simultaneous knockout of polycomb-repressive complexes PRC1 and PRC2 results in de-repression of IAP and MLV in mESCs [198]. The PRC-interacting factor RYBP has also been implicated in retroelement silencing; RYBP-deficient mESCs derepress MuERV but not MusD or IAP sequences [199]. Likewise, loss of the lysine-specific demethylase KDM1A/LSD1 results in >10-fold upregulation of MuERV_L ERVs, and marked increase in MuERV_L Gag protein in mESCs and blastocysts. Although KDM1A was not directly demonstrated to occupy ERV promoters, upregulation of MuERV_Ls was correlated with hypomethylation of H3K4, hypoacetylation of H3K27, and increased H3K9 dimethylation. Notably, DNA methylation at MuERV_L elements was unperturbed [200].

Cellular Inhibitors of L1 Retrotransposition

Mov10

Mov10 is a putative RNA helicase originally identified as an inhibitor of Moloney murine leukemia virus infection in mice [201, 202], and has subsequently been demonstrated to affect the infectivity, positively or negatively, of several RNA viruses, leading to its investigation as a regulator of endogenous retrotransposon activity [203]. Mov10 was independently identified as a possible regulator of retrotransposition due to its association with the L1 RNP as

determined by mass spectrometry analysis [204]. Indeed, several studies have demonstrated that Mov10 overexpression restricts the mobilization of L1, Alu, SVA, and IAP elements in a cell culture assay [203-205], and that knock-down of endogenous Mov10 enhances L1, Alu, and IAP retrotransposition efficiency [203, 204]. The mechanism by which MOV10 restricts L1 retrotransposition is currently unknown; however, Mov10-mediated L1 inhibition requires intact helicase motifs [204]. Furthermore, other ATP-dependent RNA helicases were demonstrated to have no effect on L1 retrotransposition, indicating that inhibition of retrotransposition is not a general feature of ATP-dependent RNA helicases [204]. Mov10 overexpression decreases L1 RNA and protein levels [204]; however, the significance of this observation in terms of a mechanism of inhibition remains to be determined.

TrexI

TrexI is a highly abundant 3'-to-5' DNA exonuclease [206-208]. TrexI mutations in humans cause Acardi-Goutieres syndrome (AGS) [209], a severe genetic encephalopathy which mimics the symptoms of *in utero* viral infection [210]. In mice, TrexI deficiency results in inflammatory myocarditis, leading to cardiomyopathy, circulatory failure, and premature death [211]. In a 2008 study [212], Stetson *et al.* implicated TrexI as a negative regulator of a cell-intrinsic defense pathway targeting cytosolic single-stranded DNA, termed the interferon-stimulatory DNA (ISD) response [213-216]. In the absence of TrexI, self-derived DNA substrates accumulate in cells and activate the ISD response. Thus, TrexI

is likely responsible for metabolizing endogenous DNA substrates, thereby preventing auto-immunity. Indeed, cytoplasmic DNA accumulates in the heart cells of *TrexI*-deficient mice compared to wild-type mice. Intriguingly, accumulated DNA in *TrexI*-deficient heart cells was enriched for retrotransposon sequences, suggesting that retroelement reverse transcripts may be metabolized by TrexI. Consistent with this hypothesis, expression of TrexI in cultured cell assays inhibited both IAP and L1 retrotransposition, and mutations in TrexI found in AGS patients abolished TrexI-mediated retrotransposition inhibition. Thus, TrexI may represent a cellular defense mechanism against retrotransposition, degrading retroelement reverse transcripts to prevent their recognition by the ISD pathway [212].

Intriguingly, mutations in any of the three subunits of human RNaseH2 can also cause AGS [210]. It is therefore tempting to speculate that RNaseH2 cooperates with TrexI to metabolize retrotransposition intermediates. For example, RNaseH2 could degrade the retroelement RNA template following first-strand cDNA synthesis, rendering the retroelement cDNA vulnerable to degradation by TrexI, which acts on single-stranded DNA [208]. However, how the cytoplasmic TrexI can access L1 cDNAs generated by TPRT in the nucleus needs to be addressed. The vulnerability of single-stranded DNA in L1 retrotransposition intermediates in the face of cellular restriction factors is emphasized by data presented in Chapter 3 of this thesis.

APOBEC3 Cytidine Deaminases

The human APOBEC3 (A3) proteins (A3A, A3B, A3C, A3DE, A3F, A3G, A3H) constitute a family of seven cytidine deaminases which reside in a head-to-tail arrangement within a gene cluster on chromosome 22 [217, 218]. The A3 family is named for its homology to APOBEC1 (apolipoprotein B mRNA-editing catalytic polypeptide 1), an RNA-editing enzyme which precisely edits one base of the ApoB mRNA to create a stop codon, facilitating generation of two ApoB isoforms, ApoB-48 and ApoB-100, which have distinct roles in cholesterol metabolism [219, 220]. The ApoB-100 is required in the liver for the production of very low density lipoprotein (VLDL), and ApoB-48 functions in the small intestine and is essential for dietary fat absorbtion (Reviewed in [221]). The A3 proteins are also related to activation-induced deaminase (AID), a single-stranded DNA editing enzyme which participates in class-switch recombination and somatic hypermutation required for immunoglobulin diversification in the adaptive immune system [222, 223].

The human APOBEC3 proteins each contain either one (A3A, A3C, A3H) or two (A3B, A3DE, A3F, A3G) cytidine deaminase (CDA) active sites [217, 218, 224]. The A3 CDA active site consists of the conserved motif H-X-E-X_{23-28}-P-C-X_{2-4}-C, where X represents any amino acid [225]. The two cysteine residues are proposed to coordinate a Zn^{2+} ion, while the glutamic acid residue is involved in proton shuttling important for the deaminase reaction [226-229]. The A3 proteins preferentially act on single-stranded DNA [230-235]. Interestingly, for double deaminase domain proteins A3B, A3G, and A3F, (not determined for A3DE), only the C-terminal deaminase domain is catalytically active [236, 237].

A3G (originally known as CEM15) was identified as an intracellular factor conferring resistance to *vif*-deficient HIV-1 infection [2, 36, 238]. At the time of viral production, A3G is incorporated into budding HIV viral particles, and is delivered to target cells upon infection. In the cytoplasm of the target cell, A3G can act upon the minus-strand HIV cDNA during reverse transcription, deaminating cytidine (dC) to uridine (dU) [233-235, 239]. Deaminated HIV genomes may be subject to degradation by the base excision repair machinery [240]; however, it has been demonstrated that the activity of uracil DNA glycosylase (UNG) is not required for the antiviral activity of A3G [241-243]. Alternatively, hyperedited HIV genomes may give rise to a mutated progeny virus, rendering it inactive and curtailing infectivity. The HIV-encoded protein virion infectivity factor (Vif) appears to have evoloved as a counter-measure against A3G. In virus-producing cells, Vif targets A3G for proteasomal degradation, excluding A3G from budding viral particles (reviewed in [244]).

Intriguingly, A3G may also elicit a deaminase-independent mechanism of HIV restriction, as deaminase deficient A3G mutants retain antiretroviral activity [245-248]. Bishop *et al.* reported a defect in the accumulation of full-length HIV reverse transcripts in the presence of A3G, and proposed a model in which A3G acts as a physical impediment to HIV RT procession on the HIV RNA [249]. Other studies have reported A3G-mediated deaminase-independent defects in primer annealing and strand transfer during HIV reverse transcription, and HIV proviral integration [250, 251]. Recently, Bélanger *et al.* described mutations in the N-terminal (non-catalytic) domain of A3G which abolish RNA binding while

retaining deaminase activity [252]. These A3G mutant proteins extensively edit HIV cDNAs, but do not restrict the generation of HIV reverse transcripts or proviral integration [252]. Thus, although the relative contribution of deaminase-dependent and independent mechanisms to A3G-mediated HIV inhibition remains the subject of investigation and controversy, ongoing studies are beginning to shed light on this difficult question [235, 239, 246, 248, 249, 251-255].

In addition to APOBEC3G, A3 family members A3B, A3DE, A3F and A3H have been demonstrated to restrict *vif*-deficient HIV infection [224, 231, 256-259]. Members of the A3 family have also been demonstrated to restrict endogenous retroelement mobilization in cell culture assays, with each member displaying activity against a specific repertoire of retroelements (Summarized in Table 1.1) [230, 237, 260-271].

The data presented in Chapters 2, 3, and 4 of this thesis build upon previous studies of A3-mediated L1 inhibition in cultured cells. Among the numerous studies that have examined the impact of A3 proteins on L1 mobilization by co-transfection experiments in a cultured-cell retrotransposition assay [34, 153], all reports find that A3A and A3B potently restrict L1 retrotransposition to ~20% of control levels [230, 237, 260-262, 264]. Interestingly, A3A and A3B can enter the nucleus—A3A by diffusion and A3B via a nuclear localization signal [237]--and therefore presumably can access L1 TPRT [230, 237]. A3B-mediated L1 inhibition clearly occurs independently of A3B cytidine deaminase activity [237, 262]. For A3A, the question of deaminase-

independence is more complicated: deaminase-domain mutants of A3A abolish its ability to restrict retrotransposition [230, 237]; however, due to the small size of the A3A protein (199 amino acids, ~26 kD), concerns have been raised as to whether mutations in the deaminase active site disrupt protein folding and therefore indirectly abolish potential deaminase-independent activities [237]. The potential for deaminase-dependent and -independent modes of A3A-mediated L1 inhibition is addressed in Chapters 3 and 4 of this thesis.

A3C, which can enter the nucleus by diffusion [237], has also been demonstrated to inhibit L1 retrotransposition, with some studies reporting strong inhibition similar to A3A and A3B [230, 264], while others report modest inhibition, to ~50% of control levels [237, 260, 261]. Whether A3C deaminase-deficient mutants can restrict L1 retrotransposition has not been reported. Stable variants of A3H, which also can enter the nucleus by diffusion, restrict L1 retrotransposition to ~20% of control levels [267, 271]. A deaminase domain mutant of A3H was demonstrated to restrict L1 retrotransposition to ~35% of control levels [267].

A3DE, A3F, and A3G localize predominantly to the cytoplasm [237, 260]. A3DE was found to have no effect on L1 retrotransposition in one study [266], and a modest effect on L1 retrotransposition (inhibition to ~50% of control levels) in another [260]. There is a distinct dichotomy in the literature about whether A3F restricts L1 retrotransposition: our lab [237] consistently observes no inhibition of L1 by A3F; in fact, we often observe that A3F expression slightly enhances L1 retrotransposition. However, other labs [230, 260-262, 264] report

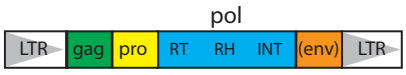
significant A3F-mediated inhibition (to 20-30% of control levels), independent of A3F deaminase activity [262]. Most reports agree that A3G, which is also cytoplasmic, does not potently restrict L1 retrotransposition [230, 237, 261, 262, 264, 270, 272]. However, one study reports robust inhibition of L1 retrotransposition [260]. Overall, differences in the reported abilities of A3 proteins to restrict L1 retrotransposition likely stem from subtle differences in experimental approach, the types of vectors from which A3 proteins are expressed, and the cell lines in which retrotransposition assays were carried out.

Most assays reporting inhibition of L1 by APOBEC3 proteins were performed by overexpression of APOBEC3 proteins and engineered L1 retrotransposition indicator plasmids in transformed cultured cells [230, 237, 260-262, 264, 266, 270-272]. Several APOBEC3 proteins (A3B, A3C, A3DE, A3F, A3G), however, are expressed endogenously in cultured human embryonic stem cells (hESCs) [237, 273]. To determine whether endogenous APOBEC3 expression in hESCs restricts retrotransposition, Wissing et al. [273] carried out exquisitely-controlled retrotransposition assays in the presence of shRNA-mediated knockdown of APOBEC3 proteins. While knockdown of A3C, A3DE, A3F, and A3G had no effect on retrotransposition levels, suppression of A3B resulted in a ~2-3.7 fold increase in the retrotransposition of a marked L1 element. Therefore, endogenous A3B may restrict L1 activity in pluripotent cells, suggesting that APOBEC3-mediated inhibition of LINE-1 may represent an early-embryonic defense against retroelement mobility.

Despite numerous studies reporting A3-mediated L1 inhibition in cultured cells, little is known about the molecular mechanism by which A3 proteins restrict L1 retrotransposition. In this thesis, I undertake a mechanistic examination of APOBEC3-mediated L1 inhibition. Chapter 2 entails a survey of A3-mediated inhibition of a variety of autonomous and non-autonomous retroelements from various species. The data presented in Chapter 2 point to a sequence-independent mode of L1 inhibition by A3A and A3B that relies on inhibition of conserved L1-encoded activities, such as endonuclease cleavage and reverse transcription, or on recognition of the L1 TPRT intermediate. In Chapter 3, I elucidate a deaminase-dependent mechanism of A3A-mediated L1 inhibition, which occurs by deamination of single-stranded DNA transiently exposed during TPRT, followed by the action of cellular repair factors to degrade deaminated TPRT intermediates. In Chapter 4, I describe an A3A mutant, A3A_F75L, which lacks deaminase activity *in vitro* yet retains the ability to restrict L1 retrotransposition. Although this mutant initially suggested a deaminase-independent mode of A3A-mediated L1 inhibition, subsequent investigation revealed that A3A_F75L likely retains deaminase activity *in vivo.* Finally, in Chapter 5, I discuss the results presented in Chapters 2-4, articulate remaining questions, and provide suggestions for future experimental directions.

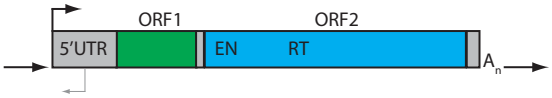
Acknowledgements: I would like to acknowledge members of the Moran lab for helpful discussion, especially Dr. Huira Kopera for critical readings of drafts of this chapter.

A. LTR Retrotransposons



Species:	% of Genome:	Active Members:
Human:	**8.55%**	none known
Mouse:	**9.87%**	
Class I	0.68%	MuLV, MuRRS, GLN, VL30
Class II	3.14%	MusD/ETn, IAP, MMTV
Class III	5.40%	MaLR, MuERV

B. Non-LTR Retrotransposons



Species:	% of Genome:	Active Members:
Human	**33.56%**	
LINEs	20.42%	LINE-1 (16.98%)
SINEs	13.41%	Alu (10.60%)
Mouse	**27.42%**	
LINEs	19.20%	LINE-1 (18.78%)
SINEs	8.22%	B1 (2.66%), B2 (2.39%), ID (0.25%)

Figure 1.1: Active Retrotransposons in the human and mouse genomes

A. Schematic of a typical LTR retrotransposon/endogenous retrovirus (ERV). White boxes with gray triangles represent the long terminal repeats (LTRs) flanking the element. Colored rectangles represent open reading frames for gag, protease (pro), polymerase (pol). The pol gene contains reverse transcriptase (RT), RnaseH (RH) and integrase (INT) activities. Some, but not all, ERVs contain an envelope (env) gene. Below, types and prevalence of ERVs in the human and mouse genomes are indicated, with a list of currently-active elements of each type [2, 3]. Reviewed in [108].

B. Schematic of a human LINE-1 autonomous non-LTR retrotransposon. The 5'UTR contains an internal promoter (black arrow); human L1 5'UTRs have an antisense promoter of unknown function (gray arrow). Open reading frames are depicted as colored rectangles. ORF1 (green) contains nucleic acid binding and chaperone activity. ORF2 harbors endonuclease (EN) and reverse transcriptase (RT) activities. LINE-1 elements end in a 3'UTR (gray rectangle) and a poly-A tail (A_n). Elements residing in the genome are typically flanked by variable-length target-site duplications (TSDs), depicted as black arrows. Below, prevalence of autonomous (LINEs) and non-autonomous (SINEs) non-LTR retrotransposons in the human and mouse genomes are indicated [2, 3]. Reviewed in [108].

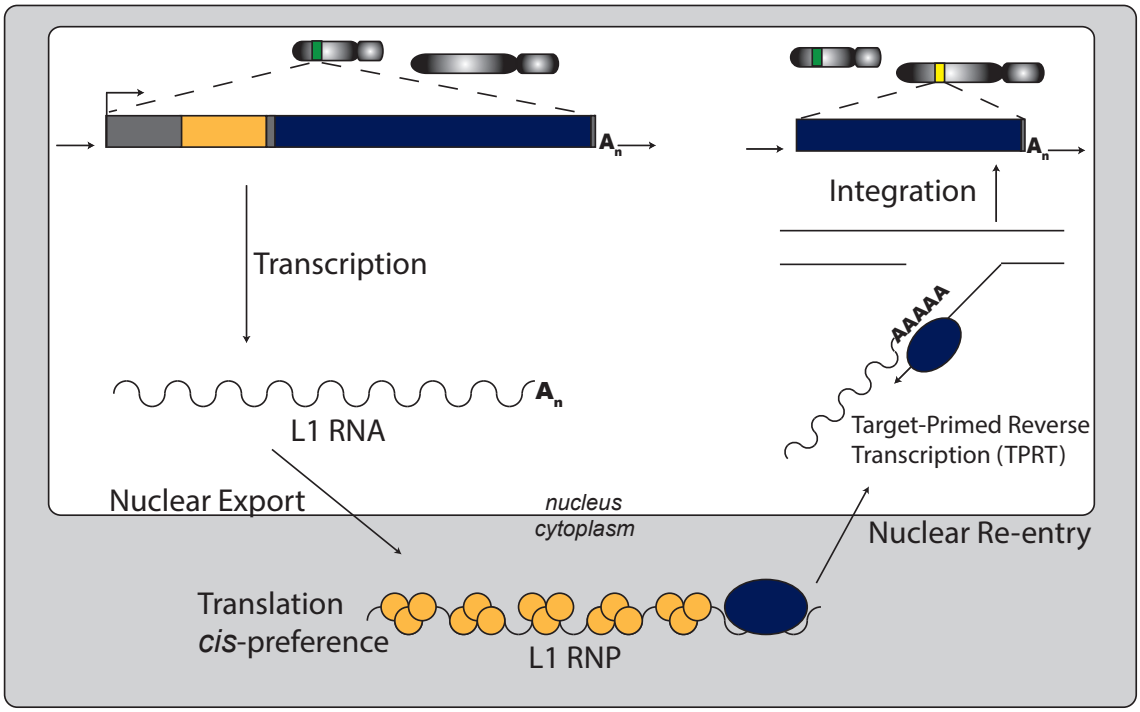


Figure 1.2: The L1 retrotransposition mechanism

A round of L1 retrotransposition begins with transcription of a full-length retrotransposition-competent L1 from its internal promoter. The L1 RNA (wavy line) is exported to the cytoplasm and translated by an unconventional termination-reinitiation mechanism. The L1-encoded proteins, ORF1p (gold circles) and ORF2p (blue circle) bind back to the RNA from which they were translated, to generate the L1 ribonucleoprotein particle (RNP), a hypothesized retrotransposition intermediate. The L1 RNP re-enters the nucleus, where the ORF2p-encoded endonuclease (EN) activity nicks genomic DNA in a new location. EN-mediated cleavage exposes a free 3'-hydroxyl residue, from which the L1 reverse transcriptase (RT) activity initiates synthesis of the first-strand L1 cDNA on the L1 RNA template. Second-strand cleavage often takes place downstream of first-strand cleavage, ultimately giving rise to variable-length target site duplications (TSDs)(black arrows) flanking the *de novo* L1 insertion upon integration. Reviewed in [274].

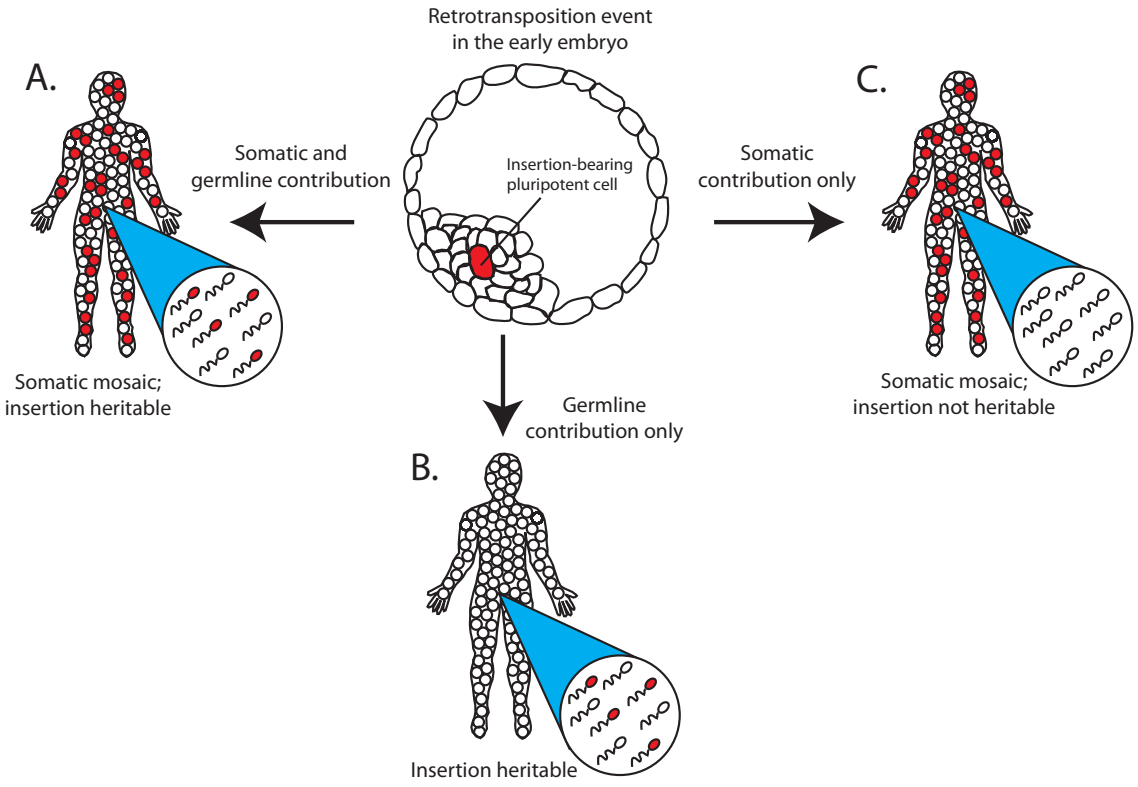


Figure 1.3: Hypothetical consequences of retrotransposition in pluripotent cells of the early embryo

A. As demonstrated by Van den Hurk et al. [147], cells harboring a de novo retrotransposition event could contribute both to the soma and germline, resulting in an individual with somatic as well as germline mosaicism. This event would be transmissible to the next generation.

B. Conceivably, cells harboring the insertion could contribute solely to the germline, giving rise to germline mosaicism and therefore rendering the insertion heritable.

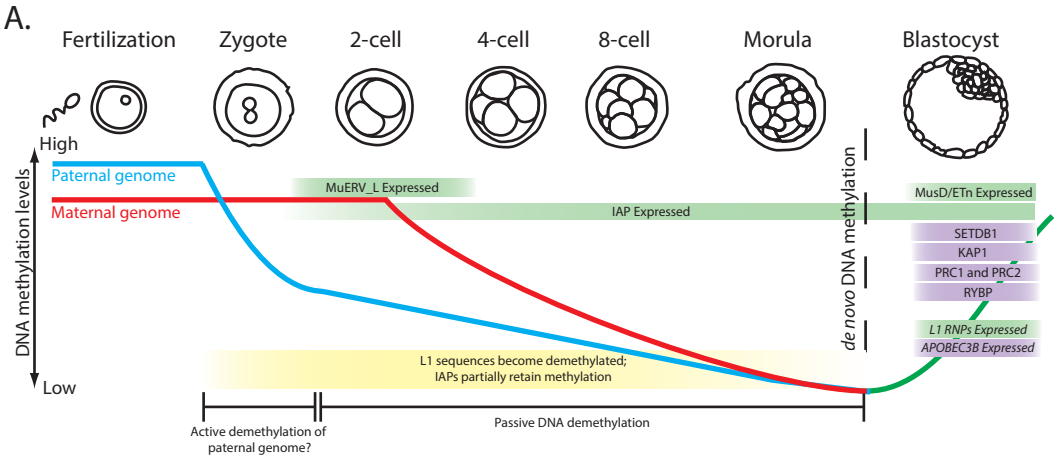
C. Insertion-bearing cells could contribute to the somatic lineage but not the germline, resulting in somatic mosaicism. Such an event would not be transmissible to the next generation. Red and white shaded circles in the human figures and sperm represent insertion-bearing and non insertion-bearing cells in the soma, and germline, respectively.

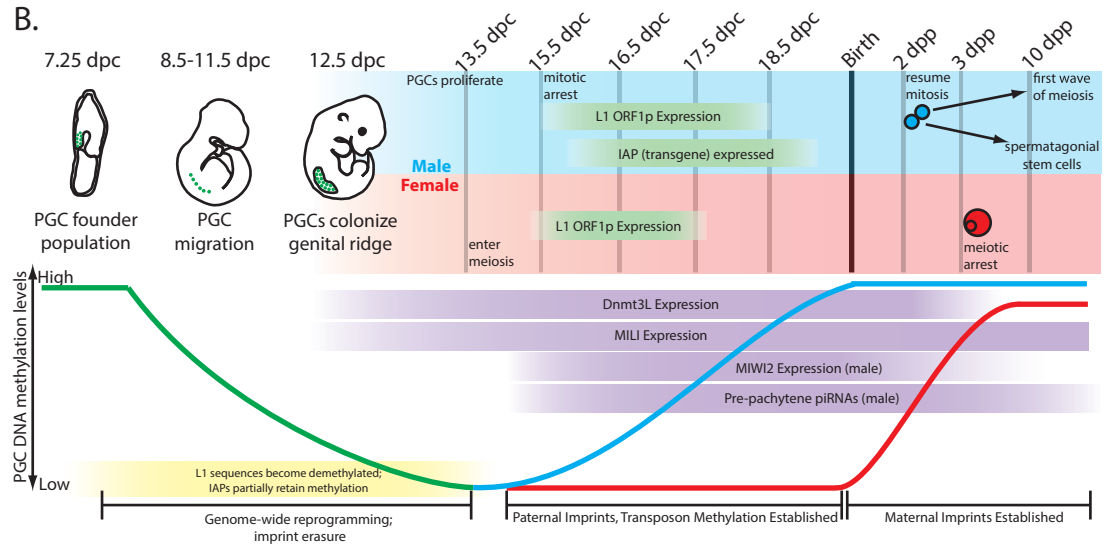
Figure 1.4: Retroelement Expression and Control During Mammalian Early Embryonic Development

A. Pre-implantation embryonic development. Above, stages of embryonic development from fertilization through implantation are represented. Below, the graph represents the relative methylation levels of the paternal (blue), maternal (red), and embryonic (green) genomes. Green bars indicate reported temporal windows of retroelement expression [24, 97, 151, 275-279]; violet bars indicate reported anti-retroelement host defense mechanisms discussed in the text [187, 188, 191, 197-199, 237, 273]. Notes on the methylation status of retroelements are highlighted in yellow [122, 124]. Italics indicate observations made in human cells.

B. Mouse primordial germline development. Above, progress of the primordial germ cells [PGCs] is represented, from their appearance at 7.25 dpc, through migration to the genital ridge ending at 11.5-12.5 dpc. Gender determination occurs at this time. Blue shading: Male PGCs proliferate until 15.5 dpc, when they undergo mitotic arrest until 2 days post-partum (dpp). At this point, mitosis resumes, and some male PGCs give rise to spermatagonial stem cells which will sustain adult meioses, while others contribute to the first wave of meiosis. Below; pink shading: Female PGCs enter meiosis prenatally at 13.5 dpc, then undergo meiotic arrest at ~3 dpp. Meiosis resumes in adult mice, but is not completed until after fertilization. Below: the graph represents relative methylation levels in PGC (green), then male (blue) and female (red) germ cell genomes. Temporal windows of retroelement (green) [280-282] and anti-retroelement factor (violet) [138, 177, 178, 180, 182] expression are indicated. Notes on the methylation status of retroelements are highlighted in yellow [124, 134].

This figure is similar to figures supplied in numerous reviews on DNA methylation during early development (*i.e.* [115]). Here, it has been re-drawn, and notes on retroelement expression, retroelement methylation, and the expression of anti-retroelement defense mechanisms have been added.


A.
Fertilization
Zygote
2-cell
4-cell
8-cell
Morula
Blastocyst
High
Low
DNA methylation levels
Paternal genome
Maternal genome
MuERV_L Expressed
IAP Expressed
de novo DNA methylation
MusD/ETn Expressed
SETDB1
KAP1
PRC1 and PRC2
RYBP
L1 RNPs Expressed
APOBEC3B Expressed
L1 sequences become demethylated;
IAPs partially retain methylation
Active demethylation of
paternal genome?
Passive DNA demethylation


B.
7.25 dpc
8.5-11.5 dpc
12.5 dpc
13.5 dpc
15.5 dpc
16.5 dpc
17.5 dpc
18.5 dpc
Birth
2 dpp
3 dpp
10 dpp
PGCs proliferate
mitotic
arrest
resume
mitosis
first wave
of meiosis
L1 ORF1p Expression
IAP (transgene) expressed
spermatagonial
stem cells
Male
Female
L1 ORF1p Expression
enter
meiosis
meiotic
arrest
PGC founder
population
PGC
migration
PGCs colonize
genital ridge
High
Low
PGC DNA methylation levels
Dnmt3L Expression
MILI Expression
MIWI2 Expression (male)
Pre-pachytene piRNAs (male)
L1 sequences become demethylated;
IAPs partially retain methylation
Genome-wide reprogramming;
imprint erasure
Paternal Imprints, Transposon Methylation Established
Maternal Imprints Established

	IAP	MusD	L1	Alu
A3A nuc/cyt	Yes: [230, 237, 263]	Yes [230]	Yes: [230, 237, 260, 261, 264]	Yes: [237]
A3B nuc/cyt	[230, 237, 263]	Yes: [230] No: [265]	Yes: [230, 237, 260-262, 264]	Yes: [237]
A3C nuc/cyt	Yes: [230] Modest: [263]	Yes: [230, 265]	Yes: [230, 260, 264] Modest: [237, 261]	Modest: [237]
A3DE cyt	n/d	Yes: [266]	Modest: [260] No: [266]	Yes: [266, 267]
A3F cyt	Yes: [230] Modest: [263]	Yes: [230, 265]	Yes: [230, 260-262, 264, 283] No: [237]	No: [268]
A3G cyt	Yes: [230, 263, 272] Modest: [265]	Yes:[265, 272] No: [230]	No: [230, 237, 262, 264, 270, 272] Modest: [283] Yes: [260, 261]	Yes: [268, 269, 283]
A3H nuc/cyt	n/d	n/d	Depends on the variant. DelN15 or G105-bearing variants (unstable): No: [264] Modest: [260] N15 intact and R105 variants (stable): Yes: [267, 271]	N15 intact and R105 variants (stable): Yes: [267]

Table 1.1: Reported APOBEC3-mediated inhibition of human and mouse endogenous retroelements

"nuc" = nuclear localization. "cyt" = cytoplasmic localization

References

1. Mc, C.B., *The origin and behavior of mutable loci in maize.* Proceedings of the National Academy of Sciences of the United States of America, 1950. **36**(6): p. 344-55.
2. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
3. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.
4. Beck, C.R., et al., *LINE-1 elements in structural variation and disease.* Annu Rev Genomics Hum Genet, 2011. **12**: p. 187-215.
5. Temin, H.M., *Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts.* Molecular biology and evolution, 1985. **2**(6): p. 455-68.
6. Boeke, J.D. and K.B. Chapman, *Retrotransposition mechanisms.* Curr Opin Cell Biol, 1991. **3**(3): p. 502-7.
7. Weiner, A.M., P.L. Deininger, and A. Efstratiadis, *Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information.* Annual review of biochemistry, 1986. **55**: p. 631-61.
8. Garfinkel, D.J., J.D. Boeke, and G.R. Fink, *Ty element transposition: reverse transcriptase and virus-like particles.* Cell, 1985. **42**(2): p. 507-17.
9. Boeke, J.D., et al., *Ty elements transpose through an RNA intermediate.* Cell, 1985. **40**(3): p. 491-500.
10. Eichinger, D.J. and J.D. Boeke, *The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition.* Cell, 1988. **54**(7): p. 955-66.
11. Eickbush, T.H. and H.S. Malik, *Origins and Evolution of Retrotransposons*, in *Mobile DNA II*, C.R. Craig N, Gellert M, Lambowitz A., Editor 2002, ASM Press: Washington, D.C. p. 1111-1144.
12. Frankel, A.D. and J.A. Young, *HIV-1: fifteen proteins and an RNA.* Annu Rev Biochem, 1998. **67**: p. 1-25.
13. Seelamgari, A., et al., *Role of viral regulatory and accessory proteins in HIV-1 replication.* Front Biosci, 2004. **9**: p. 2388-413.
14. Capy, P., *Classification and nomenclature of retrotransposable elements.* Cytogenet Genome Res, 2005. **110**(1-4): p. 457-61.
15. Belshaw, R., et al., *Long-term reinfection of the human genome by endogenous retroviruses.* Proc Natl Acad Sci U S A, 2004. **101**(14): p. 4894-9.
16. Belshaw, R., et al., *Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity.* J Virol, 2005. **79**(19): p. 12507-14.
17. Subramanian, R.P., et al., *Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses.* Retrovirology, 2011. **8**: p. 90.

18. Moyes, D., D.J. Griffiths, and P.J. Venables, *Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease.* Trends Genet, 2007. **23**(7): p. 326-33.
19. Stocking, C. and C.A. Kozak, *Murine endogenous retroviruses.* Cell Mol Life Sci, 2008. **65**(21): p. 3383-98.
20. Boeke, J.D. and J.P. Stoye, *Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements.* 1997.
21. Wicker, T., et al., *A unified classification system for eukaryotic transposable elements.* Nat Rev Genet, 2007. **8**(12): p. 973-82.
22. Maksakova, I.A., et al., *Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line.* PLoS Genet, 2006. **2**(1): p. e2.
23. Mager, D.L. and J.D. Freeman, *Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences.* Journal of virology, 2000. **74**(16): p. 7221-9.
24. Brulet, P., et al., *Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse.* Proceedings of the National Academy of Sciences of the United States of America, 1983. **80**(18): p. 5641-5.
25. Kaghad, M., L. Maillet, and P. Brulet, *Retroviral characteristics of the long terminal repeat of murine E.Tn sequences.* The EMBO journal, 1985. **4**(11): p. 2911-5.
26. Sonigo, P., et al., *Nucleotide sequence and evolution of ETn elements.* Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(11): p. 3768-71.
27. Levin, H.L., *Newly Identified Retrotransposons of the Ty3/gypsy Class in Fungi, Plants, and Vertebrates*, in *Mobile DNA II*, R.C. Nancy L. Craig, Martin Gellert, Alan M. Lambowitz, Editor 2002, ASM Press: Washington DC. p. 684-701.
28. Garfinkel, D.J., et al., *Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1.* Journal of virology, 1991. **65**(9): p. 4573-81.
29. Kirchner, J. and S. Sandmeyer, *Proteolytic processing of Ty3 proteins is required for transposition.* Journal of virology, 1993. **67**(1): p. 19-28.
30. Adams, S.E., et al., *The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins.* Cell, 1987. **49**(1): p. 111-9.
31. Telesnitsky, A. and S.P. Goff, *Reverse Transcriptase and the Generation of Retroviral DNA*, in *Retroviruses*, J.M. Coffin, S.H. Hughes, and H.E. Varmus, Editors. 1997: Cold Spring Harbor (NY).
32. Eichinger, D.J. and J.D. Boeke, *A specific terminal structure is required for Ty1 transposition.* Genes Dev, 1990. **4**(3): p. 324-30.
33. Dombroski, B.A., et al., *Isolation of an active human transposable element.* Science, 1991. **254**(5039): p. 1805-8.
34. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
35. Kazazian, H.H., Jr., et al., *Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man.* Nature, 1988. **332**(6160): p. 164-6.

36. Graur, D., Y. Shuali, and W.H. Li, *Deletions in processed pseudogenes accumulate faster in rodents than in humans.* J Mol Evol, 1989. **28**(4): p. 279-85.
37. Ostertag, E.M. and H.H. Kazazian, Jr., *Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition.* Genome Res, 2001. **11**(12): p. 2059-65.
38. Grimaldi, G., J. Skowronski, and M.F. Singer, *Defining the beginning and end of KpnI family segments.* The EMBO journal, 1984. **3**(8): p. 1753-9.
39. Brouha, B., et al., *Hot L1s account for the bulk of retrotransposition in the human population.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5280-5.
40. Sassaman, D.M., et al., *Many human L1 elements are capable of retrotransposition.* Nature genetics, 1997. **16**(1): p. 37-43.
41. Goodier, J.L., et al., *A novel active L1 retrotransposon subfamily in the mouse.* Genome Res, 2001. **11**(10): p. 1677-85.
42. Scott, A.F., et al., *Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence.* Genomics, 1987. **1**(2): p. 113-25.
43. Swergold, G.D., *Identification, characterization, and cell specificity of a human LINE-1 promoter.* Mol Cell Biol, 1990. **10**(12): p. 6718-29.
44. Speek, M., *Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes.* Mol Cell Biol, 2001. **21**(6): p. 1973-85.
45. Shapiro, J.A., *Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements.* Proceedings of the National Academy of Sciences of the United States of America, 1979. **76**(4): p. 1933-7.
46. Grimaldi, G., J. Skowronski, and M.F. Singer, *Defining the beginning and end of KpnI family segments.* EMBO J, 1984. **3**(8): p. 1753-9.
47. Athanikar, J.N., R.M. Badge, and J.V. Moran, *A YY1-binding site is required for accurate human LINE-1 transcription initiation.* Nucleic Acids Res, 2004. **32**(13): p. 3846-55.
48. Becker, K.G., et al., *Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element.* Hum Mol Genet, 1993. **2**(10): p. 1697-702.
49. Kuwabara, T., et al., *Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis.* Nat Neurosci, 2009. **12**(9): p. 1097-105.
50. Minakami, R., et al., *Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element.* Nucleic Acids Res, 1992. **20**(12): p. 3139-45.
51. Tchenio, T., J.F. Casella, and T. Heidmann, *Members of the SRY family regulate the human LINE retrotransposons.* Nucleic Acids Res, 2000. **28**(2): p. 411-5.
52. Yang, N., et al., *An important role for RUNX3 in human L1 transcription and retrotransposition.* Nucleic Acids Res, 2003. **31**(16): p. 4929-40.
53. Holmes, S.E., M.F. Singer, and G.D. Swergold, *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, 1992. **267**(28): p. 19765-8.

54. Leibold, D.M., et al., *Translation of LINE-1 DNA elements in vitro and in human cells.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(18): p. 6990-4.
55. Hohjoh, H. and M.F. Singer, *Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA.* EMBO J, 1996. **15**(3): p. 630-9.
56. Hohjoh, H. and M.F. Singer, *Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon.* EMBO J, 1997. **16**(19): p. 6034-43.
57. Januszyk, K., et al., *Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1.* J Biol Chem, 2007. **282**(34): p. 24893-904.
58. Basame, S., et al., *Spatial Assembly and RNA Binding Stoichiometry of a LINE-1 Protein Essential for Retrotransposition.* J Mol Biol, 2006.
59. Khazina, E. and O. Weichenrieder, *Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 731-6.
60. Kolosha, V.O. and S.L. Martin, *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 U S A, 1997. **94**(19): p. 10155-60.
61. Martin, S.L., et al., *Trimeric structure for an essential protein in L1 retrotransposition.* Proc Natl Acad Sci U S A, 2003. **100**(24): p. 13815-20.
62. Martin, S.L. and F.D. Bushman, *Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon.* Mol Cell Biol, 2001. **21**(2): p. 467-75.
63. Alisch, R.S., et al., *Unconventional translation of mammalian LINE-1 retrotransposons.* Genes Dev, 2006. **20**(2): p. 210-24.
64. Dmitriev, S.E., et al., *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, 2007. **27**(13): p. 4685-97.
65. McMillan, J.P. and M.F. Singer, *Translation of the human LINE-1 element, L1Hs.* Proc Natl Acad Sci U S A, 1993. **90**(24): p. 11533-7.
66. Ergun, S., et al., *Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues.* J Biol Chem, 2004. **279**(26): p. 27753-63.
67. Feng, Q., et al., *Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.* Cell, 1996. **87**(5): p. 905-16.
68. Mathias, S.L., et al., *Reverse transcriptase encoded by a human transposable element.* Science, 1991. **254**(5039): p. 1808-10.
69. Kubo, S., et al., *L1 retrotransposition in nondividing and primary human somatic cells.* Proc Natl Acad Sci U S A, 2006. **103**(21): p. 8036-41.
70. Cost, G.J. and J.D. Boeke, *Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure.* Biochemistry, 1998. **37**(51): p. 18081-93.
71. Cost, G.J., et al., *Target DNA chromatinization modulates nicking by L1 endonuclease.* Nucleic Acids Res, 2001. **29**(2): p. 573-7.

72. Cost, G.J., et al., *Human L1 element target-primed reverse transcription in vitro.* EMBO J, 2002. **21**(21): p. 5899-910.
73. Kulpa, D.A. and J.V. Moran, *Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles.* Nat Struct Mol Biol, 2006. **13**(7): p. 655-60.
74. Luan, D.D., et al., *Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.* Cell, 1993. **72**(4): p. 595-605.
75. Goodier, J.L. and H.H. Kazazian, Jr., *Retrotransposons revisited: the restraint and rehabilitation of parasites.* Cell, 2008. **135**(1): p. 23-35.
76. Christensen, S.M. and T.H. Eickbush, *R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA.* Mol Cell Biol, 2005. **25**(15): p. 6617-28.
77. Kurzynska-Kokorniak, A., et al., *DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon.* J Mol Biol, 2007. **374**(2): p. 322-33.
78. Dewannieux, M., C. Esnault, and T. Heidmann, *LINE-mediated retrotransposition of marked Alu sequences.* Nat Genet, 2003. **35**(1): p. 41-8.
79. Dewannieux, M. and T. Heidmann, *L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells.* J Mol Biol, 2005. **349**(2): p. 241-7.
80. Damert, A., et al., *5′-Transducing SVA retrotransposon groups spread efficiently throughout the human genome.* Genome Res, 2009. **19**(11): p. 1992-2008.
81. Hancks, D.C., et al., *Exon-trapping mediated by the human retrotransposon SVA.* Genome Res, 2009. **19**(11): p. 1983-91.
82. Hancks, D.C., et al., *Retrotransposition of marked SVA elements by human L1s in cultured cells.* Hum Mol Genet, 2011. **20**(17): p. 3386-400.
83. Ostertag, E.M., et al., *SVA elements are nonautonomous retrotransposons that cause disease in humans.* Am J Hum Genet, 2003. **73**(6): p. 1444-51.
84. Bennett, E.A., et al., *Natural genetic variation caused by transposable elements in humans.* Genetics, 2004. **168**(2): p. 933-51.
85. Esnault, C., J. Maestre, and T. Heidmann, *Human LINE retrotransposons generate processed pseudogenes.* Nat Genet, 2000. **24**(4): p. 363-7.
86. Wei, W., et al., *Human L1 retrotransposition: cis preference versus trans complementation.* Mol Cell Biol, 2001. **21**(4): p. 1429-39.
87. Agrawal, A., Q.M. Eastman, and D.G. Schatz, *Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.* Nature, 1998. **394**(6695): p. 744-51.
88. Hencken, C.G., X. Li, and N.L. Craig, *Functional characterization of an active Rag-like transposase.* Nature structural & molecular biology, 2012. **19**(8): p. 834-6.
89. Kapitonov, V.V. and J. Jurka, *RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons.* PLoS Biol, 2005. **3**(6): p. e181.
90. Zhou, L., et al., *Transposition of hAT elements links transposable elements and V(D)J recombination.* Nature, 2004. **432**(7020): p. 995-1001.

91. Levis, R.W., et al., *Transposons in place of telomeric repeats at a Drosophila telomere.* Cell, 1993. **75**(6): p. 1083-93.
92. Pardue, M.L. and P.G. DeBaryshe, *Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres.* Annu Rev Genet, 2003. **37**: p. 485-511.
93. Biessmann, H., et al., *HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in Drosophila melanogaster.* Mol Cell Biol, 1992. **12**(9): p. 3910-8.
94. Abad, J.P., et al., *TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres.* Mol Biol Evol, 2004. **21**(9): p. 1620-4.
95. Sheen, F.M. and R.W. Levis, *Transposition of the LINE-like retrotransposon TART to Drosophila chromosome termini.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12510-4.
96. Gladyshev, E.A. and I.R. Arkhipova, *Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes.* Proc Natl Acad Sci U S A, 2007. **104**(22): p. 9352-7.
97. Peaston, A.E., et al., *Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos.* Dev Cell, 2004. **7**(4): p. 597-606.
98. Macfarlan, T.S., et al., *Embryonic stem cell potency fluctuates with endogenous retrovirus activity.* Nature, 2012. **487**(7405): p. 57-63.
99. Burke, W.D., C.C. Calalang, and T.H. Eickbush, *The site-specific ribosomal insertion element type II of Bombyx mori (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme.* Molecular and cellular biology, 1987. **7**(6): p. 2221-30.
100. Devine, S.E. and J.D. Boeke, *Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III.* Genes Dev, 1996. **10**(5): p. 620-33.
101. Chalker, D.L. and S.B. Sandmeyer, *Ty3 integrates within the region of RNA polymerase III transcription initiation.* Genes & development, 1992. **6**(1): p. 117-28.
102. Chalker, D.L. and S.B. Sandmeyer, *Sites of RNA polymerase III transcription initiation and Ty3 integration at the U6 gene are positioned by the TATA box.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(11): p. 4927-31.
103. Kirchner, J., C.M. Connolly, and S.B. Sandmeyer, *Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element.* Science, 1995. **267**(5203): p. 1488-91.
104. Zou, S. and D.F. Voytas, *Silent chromatin determines target preference of the Saccharomyces retrotransposon Ty5.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(14): p. 7412-6.
105. Zou, S., D.A. Wright, and D.F. Voytas, *The Saccharomyces Ty5 retrotransposon family is associated with origins of DNA replication at the telomeres and the silent mating locus HMR.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(3): p. 920-4.

106. Zou, S., et al., *The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci.* Genes & development, 1996. **10**(5): p. 634-45.
107. Dai, J., et al., *Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin.* Molecular cell, 2007. **27**(2): p. 289-99.
108. Levin, H.L. and J.V. Moran, *Dynamic interactions between transposable elements and their hosts.* Nat Rev Genet, 2011. **12**(9): p. 615-27.
109. Hancks, D.C. and H.H. Kazazian, Jr., *Active human retrotransposons: variation and disease.* Current opinion in genetics & development, 2012. **22**(3): p. 191-203.
110. Cordaux, R. and M.A. Batzer, *The impact of retrotransposons on human genome evolution.* Nat Rev Genet, 2009. **10**(10): p. 691-703.
111. Belancio, V.P., D.J. Hedges, and P. Deininger, *Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health.* Genome Res, 2008. **18**(3): p. 343-58.
112. Beck, C.R., et al., *LINE-1 retrotransposition activity in human genomes.* Cell, 2010. **141**(7): p. 1159-70.
113. Bestor, T.H. and D. Bourc'his, *Transposon silencing and imprint establishment in mammalian germ cells.* Cold Spring Harb Symp Quant Biol, 2004. **69**: p. 381-7.
114. Yoder, J.A., C.P. Walsh, and T.H. Bestor, *Cytosine methylation and the ecology of intragenomic parasites.* Trends Genet, 1997. **13**(8): p. 335-40.
115. Saitou, M., S. Kagiwada, and K. Kurimoto, *Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells.* Development, 2012. **139**(1): p. 15-31.
116. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-93.
117. Castaneda, J., P. Genzor, and A. Bortvin, *piRNAs, transposon silencing, and germline genome integrity.* Mutat Res, 2011. **714**(1-2): p. 95-104.
118. Oswald, J., et al., *Active demethylation of the paternal genome in the mouse zygote.* Curr Biol, 2000. **10**(8): p. 475-8.
119. Mayer, W., et al., *Demethylation of the zygotic paternal genome.* Nature, 2000. **403**(6769): p. 501-2.
120. Santos, F., et al., *Dynamic reprogramming of DNA methylation in the early mouse embryo.* Dev Biol, 2002. **241**(1): p. 172-82.
121. Wossidlo, M., et al., *Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes.* EMBO J, 2010. **29**(11): p. 1877-88.
122. Howlett, S.K. and W. Reik, *Methylation levels of maternal and paternal genomes during preimplantation development.* Development, 1991. **113**(1): p. 119-27.
123. Kafri, T., et al., *Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line.* Genes Dev, 1992. **6**(5): p. 705-14.
124. Lane, N., et al., *Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse.* Genesis, 2003. **35**(2): p. 88-93.

125. Monk, M., M. Boubelik, and S. Lehnert, *Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development.* Development, 1987. **99**(3): p. 371-82.
126. Oda, M., et al., *DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner.* Genes Dev, 2006. **20**(24): p. 3382-94.
127. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
128. Rougier, N., et al., *Chromosome methylation patterns during mammalian preimplantation development.* Genes Dev, 1998. **12**(14): p. 2108-13.
129. Gaudet, F., et al., *Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing.* Mol Cell Biol, 2004. **24**(4): p. 1640-8.
130. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
131. Ginsburg, M., M.H. Snow, and A. McLaren, *Primordial germ cells in the mouse embryo during gastrulation.* Development, 1990. **110**(2): p. 521-8.
132. Chiquoine, A.D., *The identification, origin, and migration of the primordial germ cells in the mouse embryo.* Anat Rec, 1954. **118**(2): p. 135-46.
133. Szabo, P.E., et al., *Allele-specific expression of imprinted genes in mouse migratory primordial germ cells.* Mech Dev, 2002. **115**(1-2): p. 157-60.
134. Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells.* Mech Dev, 2002. **117**(1-2): p. 15-23.
135. Lee, J., et al., *Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells.* Development, 2002. **129**(8): p. 1807-17.
136. Yamazaki, Y., et al., *Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12207-12.
137. Kato, Y., et al., *Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse.* Hum Mol Genet, 2007. **16**(19): p. 2272-80.
138. Bourc'his, D. and T.H. Bestor, *Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.* Nature, 2004. **431**(7004): p. 96-9.
139. Chedin, F., M.R. Lieber, and C.L. Hsieh, *The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16916-21.
140. Jia, D., et al., *Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation.* Nature, 2007. **449**(7159): p. 248-51.
141. Bourc'his, D., et al., *Dnmt3L and the establishment of maternal genomic imprints.* Science, 2001. **294**(5551): p. 2536-9.
142. Brouha, B., et al., *Evidence consistent with human L1 retrotransposition in maternal meiosis I.* Am J Hum Genet, 2002. **71**(2): p. 327-36.

143. Ostertag, E.M., et al., *A mouse model of human L1 retrotransposition.* Nat Genet, 2002. **32**(4): p. 655-60.
144. Athanikar, J.N., T.A. Morrish, and J.V. Moran, *Of man in mice.* Nature genetics, 2002. **32**(4): p. 562-3.
145. Freeman, P., et al., *L1 hybridization enrichment: a method for directly accessing de novo L1 insertions in the human germline.* Hum Mutat, 2011. **32**(8): p. 978-88.
146. Georgiou, I., et al., *Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes.* Hum Mol Genet, 2009. **18**(7): p. 1221-8.
147. van den Hurk, J.A., et al., *L1 retrotransposition can occur early in human embryonic development.* Hum Mol Genet, 2007. **16**(13): p. 1587-92.
148. van den Hurk, J.A., et al., *Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon.* Hum Genet, 2003. **113**(3): p. 268-75.
149. Kano, H., et al., *L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism.* Genes Dev, 2009. **23**(11): p. 1303-12.
150. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
151. Garcia-Perez, J.L., et al., *LINE-1 retrotransposition in human embryonic stem cells.* Hum Mol Genet, 2007. **16**(13): p. 1569-77.
152. Macia, A., et al., *Epigenetic control of retrotransposon expression in human embryonic stem cells.* Mol Cell Biol, 2011. **31**(2): p. 300-16.
153. Wei, W., et al., *A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events.* Anal Biochem, 2000. **284**(2): p. 435-8.
154. Wissing, S., et al., *Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility.* Hum Mol Genet, 2011.
155. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
156. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.
157. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
158. Muotri, A.R., et al., *Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition.* Nature, 2005. **435**(7044): p. 903-10.
159. Coufal, N.G., et al., *L1 retrotransposition in human neural progenitor cells.* Nature, 2009. **460**(7259): p. 1127-31.
160. Muotri, A.R., et al., *L1 retrotransposition in neurons is modulated by MeCP2.* Nature, 2010. **468**(7322): p. 443-6.
161. Baillie, J.K., et al., *Somatic retrotransposition alters the genetic landscape of the human brain.* Nature, 2011.
162. Iskow, R.C., et al., *Natural mutagenesis of human genomes by endogenous retrotransposons.* Cell, 2010. **141**(7): p. 1253-61.

163. Solyom, S., et al., *Extensive somatic L1 retrotransposition in colorectal tumors.* Genome research, 2012. **22**(12): p. 2328-38.
164. Miki, Y., et al., *Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer.* Cancer research, 1992. **52**(3): p. 643-5.
165. Shukla, R., et al., *Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma.* Cell, 2013. **153**(1): p. 101-11.
166. Gunawardane, L.S., et al., *A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila.* Science, 2007. **315**(5818): p. 1587-90.
167. Brennecke, J., et al., *Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila.* Cell, 2007. **128**(6): p. 1089-103.
168. Sarot, E., et al., *Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene.* Genetics, 2004. **166**(3): p. 1313-21.
169. Robert, V., et al., *Characterization of the flamenco region of the Drosophila melanogaster genome.* Genetics, 2001. **158**(2): p. 701-13.
170. Prud'homme, N., et al., *Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster.* Genetics, 1995. **139**(2): p. 697-711.
171. Pelisson, A., et al., *Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene.* The EMBO journal, 1994. **13**(18): p. 4401-11.
172. Malone, C.D., et al., *Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary.* Cell, 2009. **137**(3): p. 522-35.
173. Li, C., et al., *Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies.* Cell, 2009. **137**(3): p. 509-21.
174. Girton, J.R. and K.M. Johansen, *Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila.* Advances in genetics, 2008. **61**: p. 1-43.
175. Pal-Bhadra, M., et al., *Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery.* Science, 2004. **303**(5658): p. 669-72.
176. Brower-Toland, B., et al., *Drosophila PIWI associates with chromatin and interacts directly with HP1a.* Genes Dev, 2007. **21**(18): p. 2300-11.
177. Kuramochi-Miyagawa, S., et al., *Two mouse piwi-related genes: miwi and mili.* Mech Dev, 2001. **108**(1-2): p. 121-33.
178. Aravin, A.A., et al., *A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice.* Mol Cell, 2008. **31**(6): p. 785-99.
179. Kuramochi-Miyagawa, S., et al., *Mili, a mammalian member of piwi family gene, is essential for spermatogenesis.* Development, 2004. **131**(4): p. 839-49.
180. Kuramochi-Miyagawa, S., et al., *DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.* Genes Dev, 2008. **22**(7): p. 908-17.
181. Carmell, M.A., et al., *MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline.* Dev Cell, 2007. **12**(4): p. 503-14.
182. Aravin, A.A., et al., *Developmentally regulated piRNA clusters implicate MILI in transposon control.* Science, 2007. **316**(5825): p. 744-7.

183. De Fazio, S., et al., *The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements.* Nature, 2011.
184. Sperger, J.M., et al., *Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13350-5.
185. Garcia-Perez, J.L., et al., *Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells.* Nature, 2010. **466**(7307): p. 769-73.
186. Martin, S.L. and D. Branciforte, *Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells.* Mol Cell Biol, 1993. **13**(9): p. 5383-92.
187. Matsui, T., et al., *Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET.* Nature, 2010. **464**(7290): p. 927-31.
188. Sripathy, S.P., J. Stevens, and D.C. Schultz, *The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression.* Mol Cell Biol, 2006. **26**(22): p. 8623-38.
189. Friedman, J.R., et al., *KAP-1, a novel corepressor for the highly conserved KRAB repression domain.* Genes Dev, 1996. **10**(16): p. 2067-78.
190. Emerson, R.O. and J.H. Thomas, *Adaptive evolution in zinc finger transcription factors.* PLoS genetics, 2009. **5**(1): p. e1000325.
191. Rowe, H.M., et al., *KAP1 controls endogenous retroviruses in embryonic stem cells.* Nature, 2010. **463**(7278): p. 237-40.
192. Karimi, M.M., et al., *DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs.* Cell Stem Cell, 2011. **8**(6): p. 676-87.
193. Hutnick, L.K., et al., *Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism.* J Biol Chem, 2010. **285**(27): p. 21082-91.
194. Tsumura, A., et al., *Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b.* Genes Cells, 2006. **11**(7): p. 805-14.
195. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.
196. Walsh, C.P. and T.H. Bestor, *Cytosine methylation and mammalian development.* Genes Dev, 1999. **13**(1): p. 26-34.
197. Reichmann, J., et al., *Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells.* PLoS Comput Biol, 2012. **8**(4): p. e1002486.
198. Leeb, M., et al., *Polycomb complexes act redundantly to repress genomic repeats and genes.* Genes Dev, 2010. **24**(3): p. 265-76.
199. Hisada, K., et al., *RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells.* Mol Cell Biol, 2012. **32**(6): p. 1139-49.
200. Macfarlan, T.S., et al., *Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A.* Genes Dev, 2011. **25**(6): p. 594-607.

201. Mooslehner, K., et al., *Structure and expression of a gene encoding a putative GTP-binding protein identified by provirus integration in a transgenic mouse strain.* Molecular and cellular biology, 1991. **11**(2): p. 886-93.
202. Schnieke, A., et al., *Endogenous Moloney leukemia virus in nonviremic Mov substrains of mice carries defects in the proviral genome.* Journal of virology, 1983. **45**(2): p. 505-13.
203. Arjan-Odedra, S., et al., *Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses.* Retrovirology, 2012. **9**: p. 53.
204. Goodier, J.L., L.E. Cheung, and H.H. Kazazian, Jr., *MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells.* PLoS genetics, 2012. **8**(10): p. e1002941.
205. Li, X., et al., *The MOV10 helicase inhibits LINE-1 mobility.* The Journal of biological chemistry, 2013.
206. Hoss, M., et al., *A human DNA editing enzyme homologous to the Escherichia coli DnaQ/MutD protein.* The EMBO journal, 1999. **18**(13): p. 3868-75.
207. Lindahl, T., J.A. Gally, and G.M. Edelman, *Properties of deoxyribonuclease 3 from mammalian tissues.* The Journal of biological chemistry, 1969. **244**(18): p. 5014-9.
208. Mazur, D.J. and F.W. Perrino, *Identification and expression of the TREX1 and TREX2 cDNA sequences encoding mammalian 3'-->5' exonucleases.* The Journal of biological chemistry, 1999. **274**(28): p. 19655-60.
209. Crow, Y.J., et al., *Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus.* Nature genetics, 2006. **38**(8): p. 917-20.
210. Rice, G., et al., *Clinical and molecular phenotype of Aicardi-Goutieres syndrome.* American journal of human genetics, 2007. **81**(4): p. 713-25.
211. Morita, M., et al., *Gene-targeted mice lacking the Trex1 (DNase III) 3'-->5' DNA exonuclease develop inflammatory myocarditis.* Molecular and cellular biology, 2004. **24**(15): p. 6719-27.
212. Stetson, D.B., et al., *Trex1 prevents cell-intrinsic initiation of autoimmunity.* Cell, 2008. **134**(4): p. 587-98.
213. Ishii, K.J., et al., *A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA.* Nature immunology, 2006. **7**(1): p. 40-8.
214. Martin, D.A. and K.B. Elkon, *Intracellular mammalian DNA stimulates myeloid dendritic cells to produce type I interferons predominantly through a toll-like receptor 9-independent pathway.* Arthritis and rheumatism, 2006. **54**(3): p. 951-62.
215. Okabe, Y., et al., *Toll-like receptor-independent gene induction program activated by mammalian DNA escaped from apoptotic DNA degradation.* The Journal of experimental medicine, 2005. **202**(10): p. 1333-9.
216. Stetson, D.B. and R. Medzhitov, *Recognition of cytosolic DNA activates an IRF3-dependent innate immune response.* Immunity, 2006. **24**(1): p. 93-103.
217. Jarmuz, A., et al., *An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.* Genomics, 2002. **79**(3): p. 285-96.

218. OhAinle, M., et al., *Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H.* J Virol, 2006. **80**(8): p. 3853-62.
219. Teng, B., C.F. Burant, and N.O. Davidson, *Molecular cloning of an apolipoprotein B messenger RNA editing protein.* Science, 1993. **260**(5115): p. 1816-9.
220. Navaratnam, N., et al., *The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase.* The Journal of biological chemistry, 1993. **268**(28): p. 20709-12.
221. Chan, L., et al., *Apobec-1 and apolipoprotein B mRNA editing.* Biochimica et biophysica acta, 1997. **1345**(1): p. 11-26.
222. Muramatsu, M., et al., *Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells.* J Biol Chem, 1999. **274**(26): p. 18470-6.
223. Muramatsu, M., et al., *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.* Cell, 2000. **102**(5): p. 553-63.
224. Dang, Y., et al., *Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family.* J Virol, 2006. **80**(21): p. 10522-33.
225. Chiu, Y.L. and W.C. Greene, *The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements.* Annu Rev Immunol, 2008. **26**: p. 317-53.
226. Betts, L., et al., *Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex.* Journal of molecular biology, 1994. **235**(2): p. 635-56.
227. Navaratnam, N., et al., *Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site.* Cell, 1995. **81**(2): p. 187-95.
228. Chen, K.M., et al., *Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G.* Nature, 2008. **452**(7183): p. 116-9.
229. Prochnow, C., et al., *The APOBEC-2 crystal structure and functional implications for the deaminase AID.* Nature, 2007. **445**(7126): p. 447-51.
230. Chen, H., et al., *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.* Curr Biol, 2006. **16**(5): p. 480-5.
231. Bishop, K.N., et al., *Cytidine deamination of retroviral DNA by diverse APOBEC proteins.* Curr Biol, 2004. **14**(15): p. 1392-6.
232. Chelico, L., et al., *APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA.* Nat Struct Mol Biol, 2006. **13**(5): p. 392-9.
233. Harris, R.S., et al., *DNA deamination mediates innate immunity to retroviral infection.* Cell, 2003. **113**(6): p. 803-9.
234. Lecossier, D., et al., *Hypermutation of HIV-1 DNA in the absence of the Vif protein.* Science, 2003. **300**(5622): p. 1112.
235. Mangeat, B., et al., *Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.* Nature, 2003. **424**(6944): p. 99-103.

236. Hache, G., M.T. Liddament, and R.S. Harris, *The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain.* J Biol Chem, 2005. **280**(12): p. 10920-4.
237. Bogerd, H.P., et al., *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8780-5.
238. Sheehy, A.M., et al., *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.* Nature, 2002. **418**(6898): p. 646-50.
239. Zhang, H., et al., *The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.* Nature, 2003. **424**(6944): p. 94-8.
240. Yang, B., et al., *Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA.* J Biol Chem, 2007. **282**(16): p. 11667-75.
241. Mbisa, J.L., et al., *Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration.* J Virol, 2007. **81**(13): p. 7099-110.
242. Kaiser, S.M. and M. Emerman, *Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G.* J Virol, 2006. **80**(2): p. 875-82.
243. Schumacher, A.J., et al., *The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction.* J Virol, 2008. **82**(6): p. 2652-60.
244. Chiu, Y.L. and W.C. Greene, *APOBEC3G: an intracellular centurion.* Philos Trans R Soc Lond B Biol Sci, 2009. **364**(1517): p. 689-703.
245. Shindo, K., et al., *The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity.* J Biol Chem, 2003. **278**(45): p. 44412-6.
246. Newman, E.N., et al., *Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity.* Curr Biol, 2005. **15**(2): p. 166-70.
247. Bishop, K.N., R.K. Holmes, and M.H. Malim, *Antiviral potency of APOBEC proteins does not correlate with cytidine deamination.* J Virol, 2006. **80**(17): p. 8450-8.
248. Iwatani, Y., et al., *Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G.* Nucleic Acids Res, 2007. **35**(21): p. 7096-108.
249. Bishop, K.N., et al., *APOBEC3G inhibits elongation of HIV-1 reverse transcripts.* PLoS Pathog, 2008. **4**(12): p. e1000231.
250. Luo, K., et al., *Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation.* J Virol, 2007. **81**(13): p. 7238-48.
251. Li, X.Y., et al., *APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription.* J Biol Chem, 2007. **282**(44): p. 32065-74.
252. Belanger, K., et al., *Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses.* Nucleic acids research, 2013.
253. Anderson, J.L. and T.J. Hope, *APOBEC3G restricts early HIV-1 replication in the cytoplasm of target cells.* Virology, 2008. **375**(1): p. 1-12.

254. Browne, E.P., C. Allers, and N.R. Landau, *Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent.* Virology, 2009. **387**(2): p. 313-21.
255. Miyagi, E., et al., *Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1.* J Virol, 2007. **81**(24): p. 13346-53.
256. Liddament, M.T., et al., *APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo.* Curr Biol, 2004. **14**(15): p. 1385-91.
257. Wiegand, H.L., et al., *A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins.* EMBO J, 2004. **23**(12): p. 2451-8.
258. Doehle, B.P., A. Schafer, and B.R. Cullen, *Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif.* Virology, 2005. **339**(2): p. 281-8.
259. Dang, Y., et al., *Human cytidine deaminase APOBEC3H restricts HIV-1 replication.* J Biol Chem, 2008. **283**(17): p. 11606-14.
260. Kinomoto, M., et al., *All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.* Nucleic Acids Res, 2007. **35**(9): p. 2955-64.
261. Niewiadomska, A.M., et al., *Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association.* J Virol, 2007. **81**(17): p. 9577-83.
262. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.* J Biol Chem, 2006. **281**(25): p. 16837-41.
263. Bogerd, H.P., et al., *APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.* Nucleic Acids Res, 2006. **34**(1): p. 89-95.
264. Muckenfuss, H., et al., *APOBEC3 proteins inhibit human LINE-1 retrotransposition.* J Biol Chem, 2006. **281**(31): p. 22161-72.
265. Esnault, C., et al., *Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses.* Nucleic Acids Res, 2006. **34**(5): p. 1522-31.
266. Duggal, N.K., H.S. Malik, and M. Emerman, *Positive selection of Apobec3DE in chimpanzees has driven breadth in anti-viral activity.* J Virol, 2011.
267. Tan, L., et al., *Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1.* FASEB J, 2009. **23**(1): p. 279-87.
268. Hulme, A.E., et al., *Selective inhibition of Alu retrotransposition by APOBEC3G.* Gene, 2007. **390**(1-2): p. 199-205.
269. Chiu, Y.L., et al., *High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15588-93.
270. Turelli, P., S. Vianin, and D. Trono, *The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay.* J Biol Chem, 2004. **279**(42): p. 43371-3.
271. OhAinle, M., et al., *Antiretroelement activity of APOBEC3H was lost twice in recent human evolution.* Cell Host Microbe, 2008. **4**(3): p. 249-59.

272. Esnault, C., et al., *APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses.* Nature, 2005. **433**(7024): p. 430-3.
273. Wissing, S., et al., *Endogenous APOBEC3B Restricts LINE-1 Retrotransposition in Transformed Cells and Human Embryonic Stem Cells.* J Biol Chem, 2011. **286**(42): p. 36427-37.
274. Beck, C.R., et al., *LINE-1 Elements in Structural Variation and Disease.* Annu Rev Genomics Hum Genet, 2010.
275. Calarco, P.G., *Intracisternal A particles in preimplantation embryos of feral mice (Mus musculus).* Intervirology, 1979. **11**(6): p. 321-5.
276. Poznanski, A.A. and P.G. Calarco, *The expression of intracisternal A particle genes in the preimplantation mouse embryo.* Dev Biol, 1991. **143**(2): p. 271-81.
277. Maksakova, I.A. and D.L. Mager, *Transcriptional regulation of early transposon elements, an active family of mouse long terminal repeat retrotransposons.* J Virol, 2005. **79**(22): p. 13865-74.
278. Baust, C., et al., *Structure and expression of mobile ETnII retroelements and their coding-competent MusD relatives in the mouse.* J Virol, 2003. **77**(21): p. 11448-58.
279. Brulet, P., H. Condamine, and F. Jacob, *Spatial distribution of transcripts of the long repeated ETn sequence during early mouse embryogenesis.* Proc Natl Acad Sci U S A, 1985. **82**(7): p. 2054-8.
280. Branciforte, D. and S.L. Martin, *Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition.* Mol Cell Biol, 1994. **14**(4): p. 2584-92.
281. Trelogan, S.A. and S.L. Martin, *Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis.* Proc Natl Acad Sci U S A, 1995. **92**(5): p. 1520-4.
282. Dupressoir, A. and T. Heidmann, *Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice.* Mol Cell Biol, 1996. **16**(8): p. 4495-503.
283. Khatua, A.K., et al., *Exosomes packaging APOBEC3G confer human immunodeficiency virus resistance to recipient cells.* J Virol, 2009. **83**(2): p. 512-21.

Chapter 2

Inhibition of Autonomous and Nonautonomous Retrotransposons by Human APOBEC3 Proteins

I designed and carried out the cultured cell retrotransposition assays in this chapter, with input from Dr. Jose Garcia-Perez on the experimental design of the assays shown in Figure 2.3 and Figure 2.4, and technical assistance from Randy Planegger on the assay depicted in Figure 2.5.

Abstract

In this chapter, I employ LINE and SINE elements from different species to gain insight regarding how various human APOBEC3 proteins restrict retrotransposition. I find that APOBEC3A and APOBEC3B inhibit LINE retrotransposition in a sequence-independent manner, suggesting that they inhibit a conserved step in the retrotransposition pathway. I also find that unlike human and mouse LINE-1 elements, a zebrafish LINE-2 element that does not encode an ORF1 protein is restricted by APOBEC3G, and that this inhibition is

deaminase-independent. A3G also restricts retrotransposition of the non-autonomous SINE Alu [1, 2]; here I find that A3G also restricts mobilization of the mouse SINEs B1 and B2. Thus, I observe a correlation between lack of ORF1p coding potential and A3G-mediated retrotransposition restriction.

Introduction

The seven human APOBEC3 (A3) proteins belong to a vertebrate-specific family of nucleic acid mutators, the AID/APOBEC proteins, which catalyze the deamination of cytidine to uridine in single-stranded DNA (Figure 2.1 B) or RNA (Reviewed in [3]). Apolipoprotein B mRNA editing catalytic subunit 1 (APOBEC1) is an RNA-editing enzyme which acts specifically on the apolipoprotein B (apoB) mRNA, allowing the generation of two ApoB isoforms with distinct roles in cholesterol metabolism [4, 5]. Although APOBEC1 was the first AID/APOBEC family member to be functionally characterized, it is actually a relative newcomer evolutionarily, and is restricted to the mammalian lineage [6]. Indeed, the DNA-editing enzyme activation induced deaminase (AID), which in humans is critical for class-switch recombination and immunoglobulin gene diversification in the adaptive immune system [7-9], and APOBEC2, whose biological function in humans remains to be elucidated [10], are the ancestral AID/APOBEC family members. AID and APOBEC2 have homologues in bony fish, and AID has homologues in cartiliginous fish [6]. Furthermore, the sea lamprey, a jawless vertebrate, encodes an AID homologue implicated in a unique form of adaptive immunity based on variable lymphocyte receptors [11]. An

ancestor of this lamprey deaminase is hypothesized to be the evolutionary predecessor of the AID/APOBEC family [3, 11, 12].

Like APOBEC1, the APOBEC3 (A3) cytidine deaminases are specific to the mammalian lineage [6]. The A3 family has experienced a dramatic evolutionary expansion and diversification; for example, mice have a single A3 gene, while humans encode seven A3s: APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3DE (A3DE), APOBEC3F (A3F), APOBEC3G (A3G), and APOBEC3H (A3H) (Figure 2.1) [13-15](for a comprehensive review see [16]). The A3 genes reside in a cluster on chromosome 22q13.2, where they are arranged in a head-to-tail configuration [6, 13-15]. Although initially described as “orphan” C-to-U editing enzymes [13], various A3s have since been demonstrated to restrict the mobilization or infectivity of a spectrum of endogenous retroelements and exogenous viruses (Reviewed in [16]). Indeed, the remarkable expansion of the A3 gene family is hypothesized to result from selective pressure imposed by genetic conflict with endogenous retroelements, with antiretroviral activity representing a relatively recent evolutionary “repurposing” of certain A3 factors [17].

As discussed in Chapter 1, the autonomous retrotransposon L1 and the non-autonomous SINE Alu are the only presently active elements in the human genome [18]; for review see [19]. Previous studies have demonstrated that members of the APOBEC3 family, most potently A3A and A3B, restrict L1 mobilization in a cell culture assay [20-25]. Notably, A3A and A3B can enter the nucleus, where L1 target-primed reverse transcription (TPRT) and integration

take place [20]. Interestingly, there is some conflict in the literature about whether the predominantly cytoplasmic A3F restricts L1: one study [20] reports no inhibition and perhaps a slight increase in L1 retrotransposition in the presence of A3F, while others [21-25] report significant inhibition. Similarly, most reports agree that A3G, which also localizes to the cytoplasm, does not potently restrict L1 retrotransposition [20-22, 24-27]. However, one study reports robust inhibition of L1 retrotransposition [23], though in this case A3G was expressed from an SV40 ori-containing vector in T-antigen transformed cells, likely leading to massive overexpression of the A3G protein. In contrast, A3G potently restricts Alu retrotransposition in cultured cells [1, 2]. The mechanism of inhibition is proposed to involve sequestration of Alu RNA to high molecular mass (HMM) A3G complexes, preventing Alu RNA from accessing the L1 retrotransposition machinery [1].

In this chapter, we assess the ability of human A3 proteins to restrict the mobility of non-human LINE and SINE retroelements. The data provide insight into the importance of retroelement nucleotide sequence and retroelement structure for A3-mediated inhibition.

Results

The L1 Retrotransposition Assay and Controls for A3 Cytotoxicity

To quantify L1 retrotransposition, we employed a cell culture retrotransposition assay [28, 29]. In this assay, we use constructs containing a retrotransposition-competent L1 element tagged in its 3'UTR with a neomycin

phosphotransferase reporter gene in the opposite transcriptional orientation of the L1 element. This reporter gene, dubbed *mneol,* is equipped with a heterologous promoter and polyadenylation signal. The neomycin phosphotransferase coding sequence is interrupted by an intron with splice donor and splice accepter sites in the same transcriptional orientation as the L1 element (Figure 2.2 A, above). This arrangement ensures that a functional copy of the neomycin phosphotransferase gene is only generated upon transcription of the L1 from its 5'UTR (or a heterologous CMV promoter), splicing of the L1 mRNA, translation of the L1-encoded proteins, and successful retrotransposition into a new genomic location (Figure 2.2 A, below). Selection with G418 for 12-14 days results in the generation of G418-resistant foci, which are quantified to determine L1 retrotransposition efficiency. Notably, we employ two different HeLa cell lines in the retrotransposition assays described in this chapter. HeLa-JVM accommodates L1 retrotransposition, but not Alu *trans*-mobilization. In contrast, HeLa-HA accommodates both L1 and Alu retrotransposition. These cell lines have been reported previously [2], and the molecular explanation for their differential ability to accommodate Alu retrotransposition is currently under investigation. HeLa-JVM is used in figures 2.2, 2.5, and 2.6, and HeLa-HA is used in Figure 2.4.

To determine the impact of A3 proteins on L1 retrotransposition, we co-transfect cells with L1 indicator plasmids, as described above, and A3 expression vectors [2, 20]. Co-transfection with a vector expressing rat β-arrestin, which does not affect L1 retrotransposition [2, 20], serves as a positive control.

Following selection, G418-resistant colonies are quantified, and percent retrotransposition in the presence of A3 expression is determined relative to β-arrestin control co-transfections.

We have observed, and other groups have reported, “off-target” effects on cell viability caused by ectopic A3 expression [30, 31]. In the case of A3A, these effects may arise from a DNA damage response and cell-cycle arrest [32]. To ensure that the diminishment in G418-resistant foci we observe in the presence of A3 expression is not solely due to non L1-specific effects on cell viability, we developed a stringent control in which a circular or linear neomycin phosphotransferase expression vector (hereafter referred to as “linear NEO” or “circular NEO” controls), is co-transfected with the A3 expression vector side-by-side with every retrotransposition assay performed. Results from the NEO control assays are used to normalize data from the L1 retrotransposition assay. The details of this experimental control are laid out graphically in Figure 3.2.

Inhibition of Autonomous LINE elements by APOBEC3 Proteins

We first asked whether autonomous LINE elements of divergent nucleotide sequences are inhibited by human A3 proteins. We employed autonomous LINE retrotransposons from mouse (TG_f21, L1SM) and zebrafish (Zfl2-2), which bear little nucleotide sequence identity to human L1.3 (Figure 2.3 A and B) in a cultured cell retrotransposition assay [28, 29](Figure 2.2 A). TG_f21 is a natural active mouse LINE-1 element of the G_F subfamily [33]. L1SM is a synthetic mouse LINE-1 element based on the L1spa T_F subfamily element [34],

engineered to increase GC richness while preserving the amino acid sequence of the L1-encoded proteins [35]. Zfl2-2 is a zebrafish element of the LINE-2 clade, which encodes only one ORF containing endonuclease and reverse transcriptase activities (orthologous to human L1 ORF2) [36]. The Zfl2-2 reporter construct is unique in that the *mneoI* cassette is cloned upstream of, and thus does not interrupt, the Zfl2-2 3'UTR. This modification is critical, as Zfl2-2 is a stringent type LINE and the integrity of the 3'UTR sequence is essential for recognition of the transcript by the Zfl2-2-encoded enzymatic machinery [36].

Consistent with previous studies [20-24], A3A potently inhibited human L1.3, to ~17% of β-arr control levels, and A3B efficiently inhibited human L1.3 retrotransposition, to ~17% of control levels (Figure 2.3 C). Furthermore, A3A inhibited L1SM, (~45% of control) TGf21 (~21% of control) and Zfl2-2 (~35% of control). A3B inhibited L1SM, TGf21 and Zfl2-2 retrotransposition, to ~48%, ~23%, and ~39% of control levels, respectively (Figure 2.3 C). We hypothesize that the ability of L1SM to partially evade A3-mediated inhibition may stem from its ability to generate higher levels of transcripts and RNPs [35], and possibly generate more insertions per cell (Sean Ferris, previous rotation student in the Moran lab, unpublished preliminary data)(See Chapter 5 for further discussion).

A3G did not efficiently inhibit human L1.3 retrotransposition, in agreement with previous studies [20-22, 25, 27, 37]. Similarly, retrotransposition of the mouse elements L1SM and TGf21 was not strongly impacted by A3G expression (Figure 2.3 C). Strikingly, however, the zebrafish LINE-2 element, Zfl2-2, was potently inhibited by A3G, to ~29% of control levels (Figure 2.3 C).

Inhibition of Nonautonomous Retrotransposons by APOBEC3G

Zfl2-2 does not encode a protein analogous to the LINE-1 ORF1p [36]. Similarly, the nonautonomous retrotransposon Alu, which is potently restricted by A3G [2], does not have protein-coding potential, and requires only L1 ORF2p for mobilization in *trans* [38, 39]. Thus, we observe a correlation between ORF1p coding potential and resistance to A3G-mediated retrotransposition inhibition. We therefore asked whether other ORF1p-less non-autonomous retroelements, specifically the mouse SINEs B1 and B2, are likewise susceptible to A3G-mediated inhibition. The mouse SINE B1, like human Alu, is derived from the 7SL gene [40]; however, its structure is monomeric instead of dimeric [41]. Unlike Alu and B1, the mouse SINE B2 is derived from a tRNA gene [42].

To assay the mobilization of non-autonomous retroelements by the L1-encoded proteins, we employed a specialized *trans* retrotransposition assay (Figure 2.2 B)[38, 43, 44]. In this assay, a reporter-less L1 “driver” construct, expressing either a full-length L1 element or L1 ORF2p only, is co-transfected with a “reporter” construct consisting of the non-autonomous retroelement tagged with a *mneoI* reporter cassette. For reporter constructs in which the element is a Pol III-derived transcript (Alu and mouse SINEs), a special reporter cassette, termed neo^{TET} is used, in which the *mneoI* gene is interrupted by a group I autocatalytic intron, as Pol III transcripts are not processed by the spliceosome [38, 43, 45]. We also employed L1.3/ORF1*mneoI*, which is a reporter for ORF1p-mediated processed pseudogene formation and consists of L1.3 ORF1

followed by the *mneoI* indicator cassette [44]. This construct mimics the SINE HAL1, recently identified in mammalian genomes and characterized by Dr. Arian Smit [46, 47]. In this experiment, L1.3/ORF1*mneoI* served as an internal control, as it must undergo mobilization in *trans* but retains ORF1p coding potential.

We carried out *trans* retrotransposition assays consisting of Alu, B1, B2 and ORF1*mneoI* reporter constructs, co-transfected with the 5'UTR/ORF2/noNeo "driver" construct (Figure 2.4 A), and APOBEC3 or β-arr control plasmids. Consistent with previous studies [2], Alu retrotransposition was efficiently inhibited by A3A, A3B, and A3G, to ~5%, ~9%, and ~13% of control, respectively (Figure 2.4 B). Furthermore, the mouse SINEs B1 and B2 were also inhibited by A3A (~5% and ~17% of control), A3B (~6% and ~14% of control), and A3G (~5% and ~9% of control) (Figure 2.4 B). Indeed, L1.3/ORF1*mneoI* mobilization was only modestly reduced by A3G expression (~55% of control), but to a lesser extent than Alu, B1, and B2. In sum, these data strengthen the correlation between ORF1p coding potential and resistance to A3G-mediated retrotransposition inhibition.

Deaminase-Independent Inhibition of Zfl2-2 Retrotransposition by APOBEC3G

Previous studies indicate that A3G-mediated Alu inhibition occurs independently of A3G deaminase activity, as an A3G catalytic mutant (A3G_E259Q, or A3Gm) efficiently inhibits Alu retrotransposition [2]. To determine whether inhibition of Zfl2-2 retrotransposition by A3G is dependent

upon A3G cytidine deaminase activity, we employed A3Gm in the cell culture retrotransposition assay. A3Gm inhibited Zfl2-2 retrotransposition to 33.8% of β-arr control levels, similar to wild-type A3G (32.8% of control levels) (Figure 2.5).

In an orthogonal approach to the question of deaminase-dependence, we also asked whether A3G-mediated inhibition of Zfl2-2 could be alleviated by expression of a small protein inhibitor of the DNA repair factor uracil DNA glycosylase (UNG). Uracils in DNA are recognized as damage by UNG, which removes the uracil base [48]. The resulting abasic site is subsequently cleaved by apurinic/apyrimidinic endonuclease (APE)[49]. We reasoned that if A3G inhibits Zfl2-2 retrotransposition by a mechanism that entails deamination of cytidine to uracil in DNA, and degradation by the downstream DNA repair response to deamination, inhibition of UNG by uracil glycosylase inhibitor (UGI) would alleviate A3G-mediated L1 inhibition. We found that UGI expression has no effect on Zfl2-2 inhibition by A3G or A3Gm (Figure 2.5 A), while UGI expression alleviated A3A-mediated inhibition of Zfl2-2 retrotransposition (from 6.3% to 20.0% of control). Notably, A3A-mediated inhibition of human L1.3 retrotransposition is alleviated by UGI expression (Figure 2.5 B and Figure 3.17); this result is expended upon and discussed in depth in Chapter 3. Taken together, the data suggest that A3G-mediated inhibition of Zfl2-2 retrotransposition takes place by a deaminase-independent mechanism.

APOBEC3F Does Not Inhibit L1 Retrotransposition in HeLa-JVM cells

Previous work from our lab has shown that APOBEC3F (A3F) does not inhibit L1 retrotransposition in HeLa-JVM cells [2, 20]. However, several other publications report significant inhibition of L1 by A3F [21-25]. Specifically, the lab of Dr. Reuben Harris reports that A3F efficiently inhibits L1 retrotransposition in a deaminase-independent manner [25]. We hypothesized that this discrepancy could by due to differences in nucleotide or amino acid sequence between A3F expression plasmids used by different labs. We therefore obtained the A3F expression vector used by the Harris lab (A3F_Harris, or A3FH), and compared the nucleotide sequence to the plasmid obtained from Dr. Brian Cullen (A3F_Cullen, or A3FC) used in our assays. The A3F construct from Dr. Harris' lab contains three nucleotide changes relative to the construct from Dr. Cullen's lab: t322g, resulting in a serine to alanine change (S108A), t429c which is silent with respect to amino acid sequence, and a691g, which causes an isoleucine to valine change (I231V) (Figure 2.6 A). In addition, we obtained an A3F expression plasmid from Dr. Michael Malim's lab (A3F_Malim, or A3FM). Although based on the "chain of custody" this construct should be identical to the one received from Dr. Cullen's lab, sequencing revealed one nucleotide change (a814g; T272A) in A3FM relative to A3FC.

To assess whether the nucleotide or resultant amino acid changes present in the A3F construct from the Harris lab impart the ability to inhibit L1 retrotransposition, we carried out retrotransposition assays using the A3FC, A3FH, and A3FM constructs. In HeLa-JVM cells, none of the A3F constructs tested effected significant inhibition of L1 retrotransposition (Figure 2.6 B). Thus,

we conclude that sequence changes in the Harris A3F construct relative to the Cullen A3F or Malim A3F constructs do not, by themselves, impart the ability to inhibit L1 retrotransposition.

Discussion

A3A and A3B inhibit autonomous and non-autonomous retrotransposons

In this Chapter, we observe that A3A and A3B inhibit human, mouse, and zebrafish LINE elements with divergent nucleotide sequences (Figure 2.3, A and B). These data suggest that A3A and A3B do not recognize a conserved sequence motif in the human L1 element, which might be expected as the human APOBEC3 locus evolved and expanded with exposure to human, but not mouse or zebrafish, LINE element retrotransposition [3]. Instead, we hypothesize that A3A and A3B, which can enter the nucleus [20] act in some manner at target-primed reverse transcription (TPRT) (Figure 2.7 A)(See Chapter 3). Indeed, TPRT is required to generate insertions by all of the elements tested, either directly (autonomous LINEs) or in *trans* (non-autonomous SINEs and L1.3/ORF1*mneol*). In Chapter 3, we pursue this hypothesis and undertake a detailed mechanistic examination of A3A-mediated inhibition of L1 retrotransposition.

ORF1p may be protective against A3G-mediated retrotransposition inhibition

The prevailing model for A3G-mediated Alu inhibition is that A3G sequesters Alu RNA to high molecular mass (HMM) cytoplasmic complexes,

thereby restricting access to the L1 retrotransposition machinery [1]. Intriguingly, we found that an autonomous zebrafish LINE-2 element (Zfl2-2), which encodes only one ORF analogous to human and mouse L1 ORF2p [36], is also restricted by A3G, in a deaminase-independent manner (Figure 2.3, Figure 2.5). Likewise, the mouse SINEs B1 and B2 are also potently inhibited by A3G, revealing a correlation between ORF1p coding capacity and escape from A3G-mediated inhibition. We therefore speculate that ORF1p protects L1 RNPs from sequestration by A3G in HMM complexes, while Alu, B1, B2, and Zfl2-2 are susceptible to A3G-mediated restriction (Figure 2.7 B).

Why are reports on A3F-mediated L1 inhibition inconsistent with one another?

We find that nucleotide and resulting minor amino acid changes between A3F expression vectors from different labs [2, 20-25] do not account for the conflicting reports of the ability of A3F to inhibit L1 retrotransposition (Figure 2.6). We hypothesize that A3F-mediated L1 inhibition may depend on the presence of an unidentified cellular factor that may associate with A3F or the L1 RNP. Notably, we do not observe A3F-mediated L1 inhibition in HeLa-HA cells or HeLa-JVM cells. Should a set of cell lines that do and do not allow A3F-mediated L1 inhibition be identified, it would be interesting to compare the proteins and RNAs that associate with A3F (and the L1 RNP) in each cell line, and identify candidate cellular factors that may facilitate or impede A3F-mediated L1 restriction. However, such a factor may not directly associate with A3F or the

L1-encoded proteins, which would make its identification a complicated undertaking.

In sum, we conclude that different APOBEC3 proteins likely effect retroelement restriction by different mechanisms. A3 factors that can access the nucleus are hypothesized to act during TPRT. In contrast, cytoplasmic A3s, specifically A3G, may carry out inhibition via a cytoplasmic sequestration mechanism.

Materials and Methods

Plasmids

All plasmids were grown in DH5α (F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1) competent *E. coli* (Invitrogen; Carlsbad, CA. Prepared in house as described in [50]). Plasmids were prepared using the Qiagen Plasmid Midi Kit (QIAGEN; Hilden, Germany) according to the manufacturer's protocol.

APOBEC3 expression constructs:

The pK_βarr, A3A, A3B, A3F_Cullen, A3G, and A3Gm expression plasmids have been described previously [51]. pK_A3F_Malim was received from Dr. Michael Malim. pK_A3F_Harris has been described previously [25].

LINE expression constructs:

pJM101/L1.3 has been described previously [52, 53]. It consists of the pCEP4 backbone (Invitrogen/Life Technologies; Carlsbad, CA) containing a full-length copy of the L1.3 element with the *mneoI* indicator cassette in the 3'UTR.

pDK101 has been described previously [54], and consists of JM101/L1.3 modified by PCR mutagenesis to contain the T7 *gene 10* epitope tag at the C-terminus of ORF1p.

pCEP4/L1SM has been described previously [35], and consists of a synthetic mouse L1 sharing the same amino acid sequence as L1spa [34], but with 24% of its nucleotide sequence replaced for optimal GC-richness. It contains the *mneoI* indicator cassette in the 3'UTR, and is cloned in the pCEP4 backbone.

pCEP4/TGf21 has been described previously [33] and consists of a natural mouse element with the *mneoI* indicator cassette in the 3' UTR, cloned in the pCEP4 backbone.

pCEP4/Zfl2-2 has been described previously [36] and consists of a zebrafish LINE-2 element cloned in the pCEP4 backbone. The Zfl2-2 3'UTR is cloned 3' of the *mneoI* cassette.

L1.3/ORF1*mneoI* has been described previously [44] and consists of the pCEP4 backbone containing the L1.3 5'UTR, L1.3 ORF1, and the *mneoI* indicator cassette.

KUB-101-LRE3-sv+ consists of the pBSKS-II backbone (Stratagene/Agilent Technologies; Santa Clara, CA) containing a full-length LRE3 element driven by a mouse UBC promoter, with the *mneoI* indicator cassette in the 3'UTR, and the SV40 late polyadenylation signal.

SINE Expression Constructs

pAluNF1-NEO^{III} has been described previously [38] and contains an Alu element isolated from a de novo insertion into the NF1 gene [55] downstream of the 7SL enhancer upstream sequence. The *neo*^{TET} reporter gene [56], driven by the SV40 promoter, is situated between Alu element sequence and the Alu polyA tract. The 7SL transcription terminator follows the Alu polyA tract.

***Neo*^{TET}-marked B1 and B2 elements** have been described previously [43], and are similar to pAluNF1-NEO^{III}, except they contain active murine B1 or B2 elements in place of Alu.

LGCX vector has been described previously [57] and is a variant of the LNCX retroviral vector [58] in which the neomycin phosphotransferase gene has been replaced by GFP.

LGCX/UGI has been described previously [57] and consists of the LGCX vector containing a uracil glycosylase inhibitor (UGI) gene codon-optimized for expression in human cells (hUGI).

pU6i NEO is a pBSKS-based plasmid with the neomycin phosphotransferase (NEO) gene from pEGFPN1 (Clontech) introduced into the backbone. The multi-cloning site contains the U6 promoter. To generate linearized plasmid for control transfections, pU6i NEO was digested with *Bgl*II (New England Biolabs; Ipswitch, Massachussets), which does not disrupt NEO gene expression, and run on an agarose gel to confirm linearization. The restriction digest reactions were purified using the Qiagen gel extraction kit.

Cell Culture

HeLa cells: Cells were grown at 37°C in the presence of 7% CO_2 and 100% humidity, using Dulbecco's modified Eagle medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine calf serum (FBS) (Invitrogen) and 1X penicillin/streptomycin/glutamine (Invitrogen).

LINE and SINE Retrotransposition Assays

Retrotransposition assays in HeLa-JVM and HeLa-HA cells were carried out as previously described [28, 29]. Cells were plated at an appropriate density in 6-well dishes (BD Biosciences; San Jose, California), T-75 flasks (BD Biosciences) or 10cm dishes (BD Biosciences or Corning; Corning, New York) to obtain quantifiable colonies for the retroelement expression construct used (see figure legends). Eighteen hours later, transfections were carried out using the FuGene 6 transfection reagent (Roche; Penzberg, Germany) and Opti-MEM (Invitrogen), according to manufacturer's protocol (3 µl FuGene and 97 µl Opti-MEM per µg of DNA transfected). Cell culture media was replaced the following day. Cells were subjected to selection with 400 µg/ml G418 (Invitrogen) 72 hours post-transfection. Selection was carried out for 12-14 days, replacing the selection media every other day. Colonies were washed with 1X phosphate buffered saline (PBS) (Gibco), fixed with 37% paraformaldehyde/8% glutaraldehyde, and stained with 0.1% crystal violet solution.

Acknowledgements

I would like to thank members of the Moran lab, especially John Moldovan, for in-depth discussion about mechanisms of L1 restriction by cellular factors. Dr. Jose Garcia-Perez helped with the initial experimental design for Figure 2.3 and Figure 2.4. Randy Planegger helped plan and provided technical assistance with the retrotransposition assay shown in Figure 2.5.

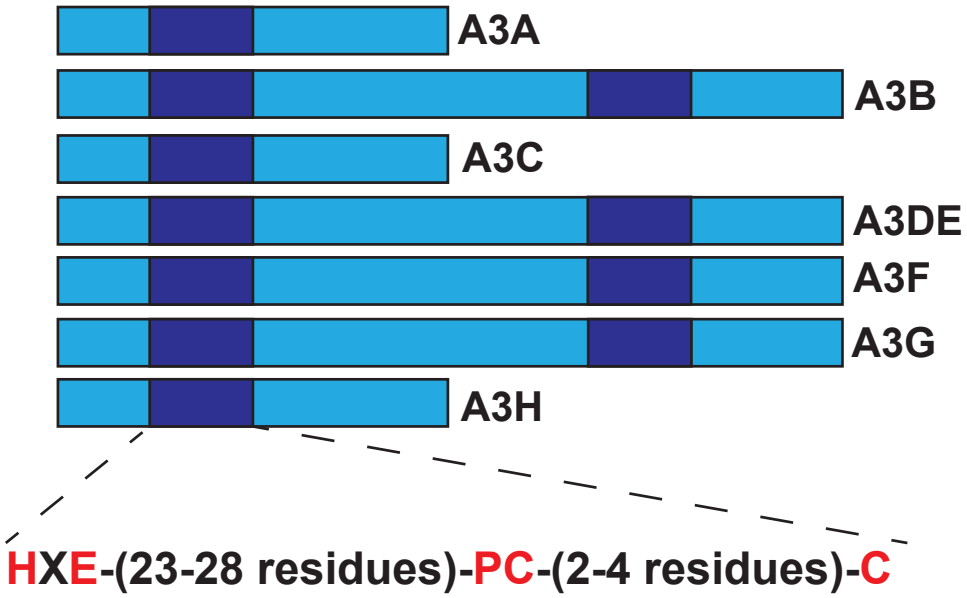
Figure 2.1: The APOBEC3 Cytidine Deaminases

A. The human APOBEC3 (A3) family consists of seven cytidine deaminases. Each A3 protein harbors one (A3A, A3C, A3H) or two (A3B, A3DE, A3F, A3G) deaminase active sites (dark blue rectangles). Below, the consensus A3 active site is shown, with zinc-coordinating residues critical for enzymatic function highlighted in red [13-15].

B. The cytidine deamination reaction. Cytidine deaminases convert cytidine to uridine through the consumption of a water molecule (H_2O) and the release of ammonia (NH_3).

Figure 2.1 is similar to figures supplied in several review articles [16, 59, 60], but was redrawn here.

A.


A3A
A3B
A3C
A3DE
A3F
A3G
A3H
HXE-(23-28 residues)-PC-(2-4 residues)-C

B.

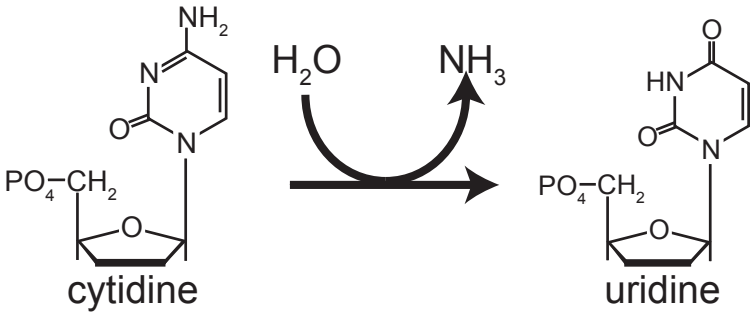
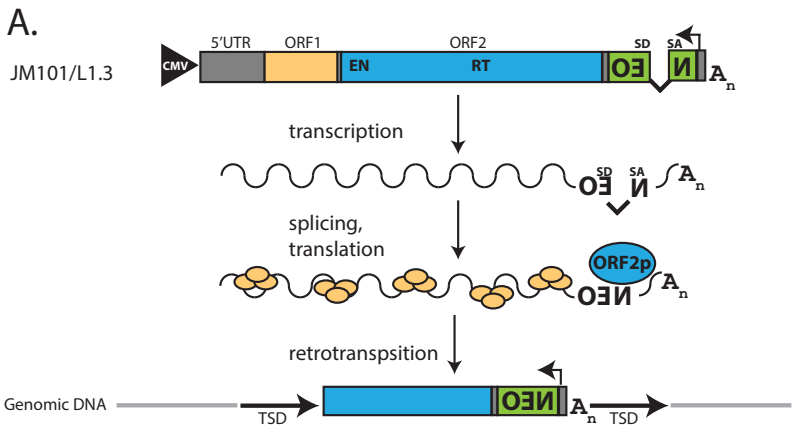

NH2
N
O
N
PO4–CH2
O
cytidine
H2O
NH3
O
HN
O
N
PO4–CH2
O
uridine

Figure 2.2: Cultured-cell assays for *cis* and *trans* retrotransposition

A. The L1 retrotransposition assay. A retrotransposition-competent L1 (in this case, L1.3) is tagged in its 3'UTR with a neomycin phosphotransferase reporter gene (shown in green) under control of a heterologous promoter (black arrow), in the opposite transcriptional orientation of the L1 element [28, 61]. The *neo* gene is interrupted by an intron in the same transcriptional orientation as the L1 element [28, 61]. This cassette is referred to as *mneoI* [28]. This arrangement ensures that a functional *neo* gene can only be expressed if the L1 is transcribed from its 5'UTR (or a heterologous CMV promoter), and the mRNA is spliced, translated, and the resulting L1 RNP undergoes a round of retrotransposition into a new genomic location [28]. The integrated *neo* reporter gene allows the generation of G418-resistant foci [28, 29].

B. Representative assays for trans *retrotransposition*. In *trans* retrotransposition assays, a plasmid expressing an untagged full-length retrotransposition-competent L1 or ORF2p only (5'UTR/ORF2/NoNeo [62], center) is co-transfected with a tagged reporter construct. On the left, L1.3/ORF1*mneoI* contains the L1 5'UTR, ORF1, and the *mneoI* reporter cassette described above [44]. The ORF2p from 5'UTR/ORF2/NoNeo mobilizes the spliced L1.3/ORF1*mneoI* reporter mRNA in trans, resulting in G418-resistant foci. On the right is an example assay for nonautonomous retrotransposon *trans* mobilization. The pAluNF1-NEO^{III} [38] contains an active Alu element under control of the 7SL enhancer upstream sequence. The two Alu monomers are separated by a short internal poly-A tract (A_n). The Alu sequence is followed by the *neo*^{TET} reporter cassette (green, interrupted by a self-splicing intron, in red). G418-resistant colonies arise upon *trans* retrotransposition of the Alu RNA by L1 ORF2p.


A.
JM101/L1.3
5'UTR
ORF1
ORF2
EN
RT
SD
SA
CMV
transcription
splicing,
translation
ORF2p
retrotranspsition
Genomic DNA
TSD
TSD

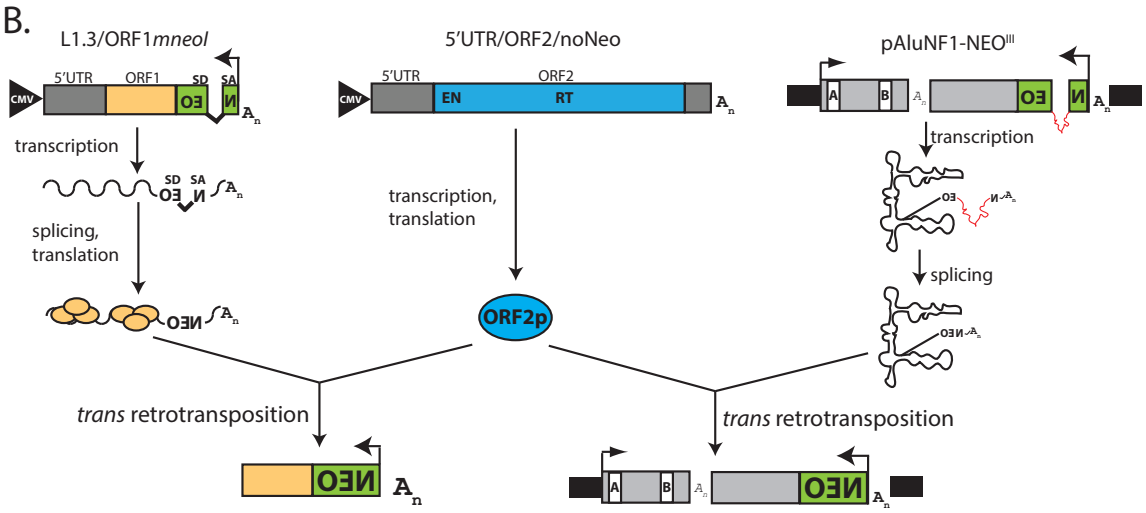

B.
L1.3/ORF1mneol
5'UTR
ORF1
CMV
transcription
splicing,
translation
5'UTR/ORF2/noNeo
ORF2
EN
RT
transcription,
translation
ORF2p
pAluNF1-NEO^III
A
B
transcription
splicing
trans retrotransposition
trans retrotransposition

Figure 2.3: Inhibition of LINE Elements by Human APOBEC3 Proteins

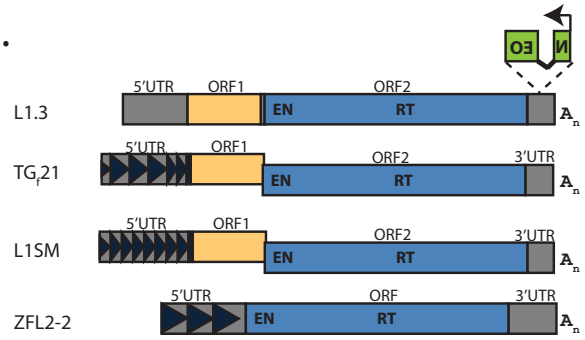
A. Human, mouse, and zebrafish LINE elements. The *mneoI* indicator cassette [28, 61] (green) is shown within the 3'UTR of human L1.3. Dark blue triangles represent the monomers in the 5'UTR of the mouse L1s (TG_f21 and L1SM) [33, 35], and repetitive units in the 5'UTR of the zebrafish LINE-2 element (Zfl2-2) [36]. L1.3, TG_f21 and L1SM each encode ORF1 and ORF2 [33, 35]; Zfl2-2 contains only one ORF which contains endonuclease (EN) and reverse transcriptase (RT) activities, analogous to ORF2 of the human and mouse elements [36].

B. Percent nucleotide identity of mouse and zebrafish elements to human L1.3.

C. Human APOBEC3 proteins inhibit diverse retroelements. HeLa-JVM cells were plated in six-well dishes, at appropriate density for the retroelement used (L1.3, $2x10^{3;}$ L1SM, $2x10^3$; TGf21, $2x10^4$; Zfl2-2, $4x10^4$ cells/well). Wells were transfected (in triplicate per condition) with 0.5 µg L1 reporter plasmid and 0.5 µg APOBEC3 reporter plasmid.

Human (L1.3), mouse (L1SM and TGf21), and zebrafish (Zfl2-2) retroelements are indicated on the x-axis. The y-axis indicates percent retrotransposition relative to control (β-arrestin). Colored bars indicate the A3 expression plasmid used (blue=A3A, red=A3B, green=A3G, purple=β-arrestin). These data represent the average of triplicate wells; error bars indicate percent standard deviation. Data are normalized to linear NEO control co-transfections, performed as described in Figure 3.3. A representative assay is shown; this experiment was performed three times representing biological replicates.

A.



B.

% Nucleotide Identity with L1.3

	ORF1	ORF2
TGf21	28.2	59.3
L1SM	32.6	51.7
Zfl2-2	n/a	25.8

C.

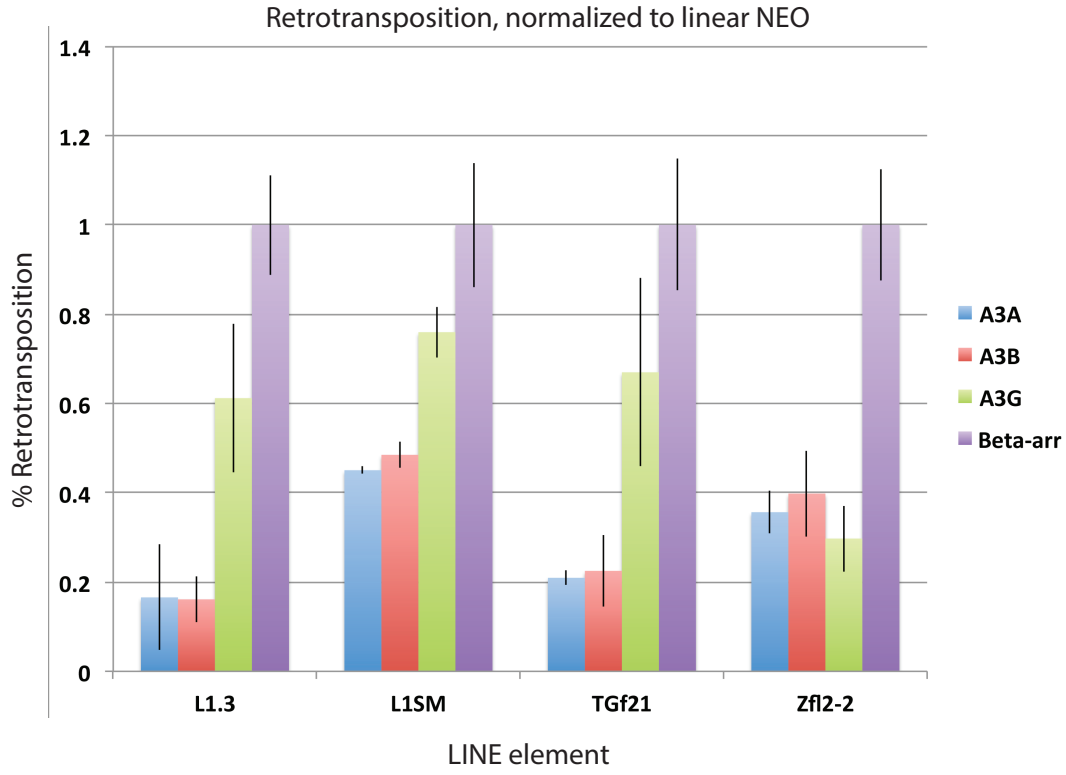


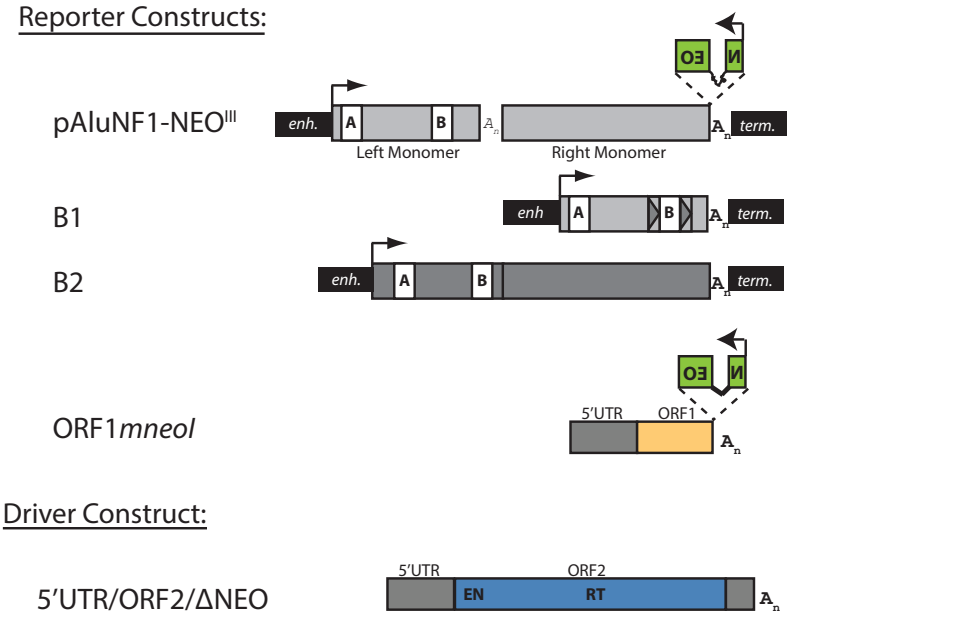
Figure 2.4: Inhibition of SINE Retrotransposition by APOBEC3 Proteins

A. SINE elements from human and mouse. Reporter constructs: The human SINE *Alu* (pAluNF1-NEO^{III}) [38] and the mouse SINE B1 and B2 reporter constructs [43] are equipped with a specialized *mneoI* reporter gene interrupted by an autocatalytic intron (*neo^{tet};* green rectangles, shown for *Alu* only). The element and reporter gene are flanked by the 7SL enhancer sequence (*enh*) and the 7SL transcription terminator (*term)* [38, 43]. The left and right monomers of the Alu element are separated by a short internal poly-A tract (A_n). The L1.3/ORF1*mneoI* [44] reporter construct consists of the L1.3 5'UTR (grey rectangle) and ORF1 (gold rectangle), and bears the *mneoI* cassette (green rectangles). Driver construct: 5'UTR/ORF2/Δneo consists of the L1.3 5'UTR, ORF2, and 3'UTR, and does not contain a reporter cassette.

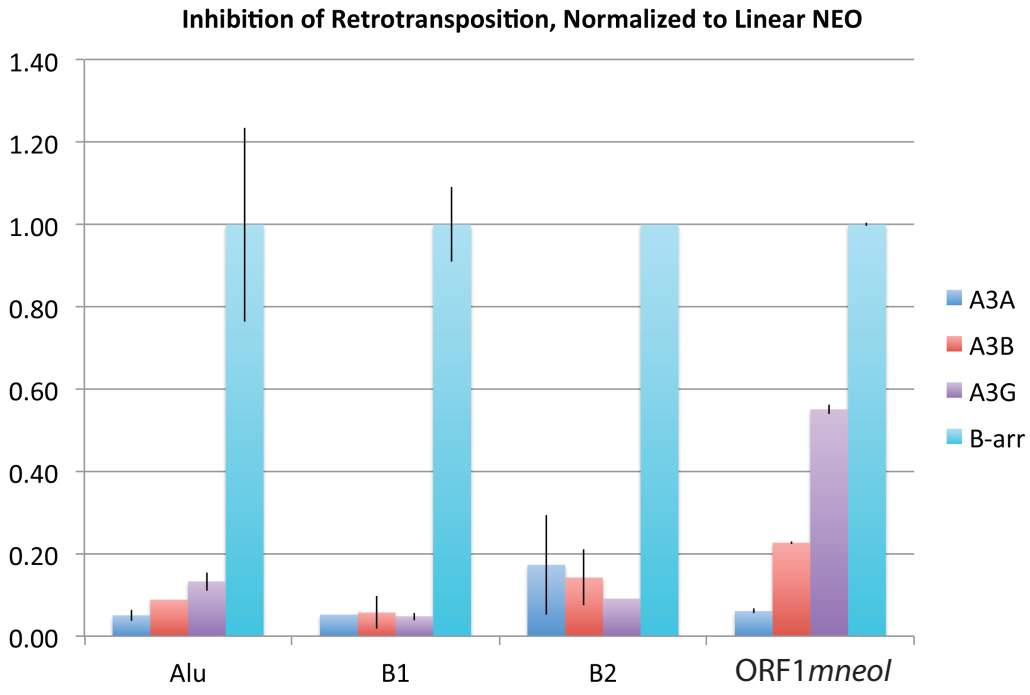
B. Human APOBEC3 proteins inhibit nonautonomous retrotransposons (SINEs). Approximately 5x10^5 HeLa-HA cells were plated per T-75 flask. Each flask was transfected with a total of 4 µg of plasmid DNA: 2 µg of the reporter plasmid (Alu, B1, B2, or ORF1mneoI), 1µg of "driver" plasmid (5'UTR/ORF2/noNeo) [62], and 1 µg of APOBEC3 plasmid. Flasks were transfected in duplicate for each condition.

The reporter element is indicated along the x-axis (human Alu, mouse B1 and B1, and the internal control ORF1*mneoI*); y-axis indicates percent retrotransposition relative to control (β-arr). Colored bars correspond to the co-transfected APOBEC3 plasmid (blue = A3A, red = A3B, purple = A3G, turquoise = β-arr). Values are the average of duplicate co-transfections per condition; error bars indicate percent standard deviation. Data were normalized to linear NEO control transfections, carried out as described in Figure 3.3. A representative assay is shown; this assay was performed twice representing biological replicates.

A.


Reporter Constructs:
pAluNF1-NEO
enh.
A
B
Left Monomer
Right Monomer
term.
B1
enh
A
B
term.
B2
enh.
A
B
term.
ORF1mneol
5'UTR
ORF1
Driver Construct:
5'UTR/ORF2/ΔNEO
5'UTR
ORF2
EN
RT

B.


Inhibition of Retrotransposition, Normalized to Linear NEO
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Alu
B1
B2
ORF1mneol
A3A
A3B
A3G
B-arr

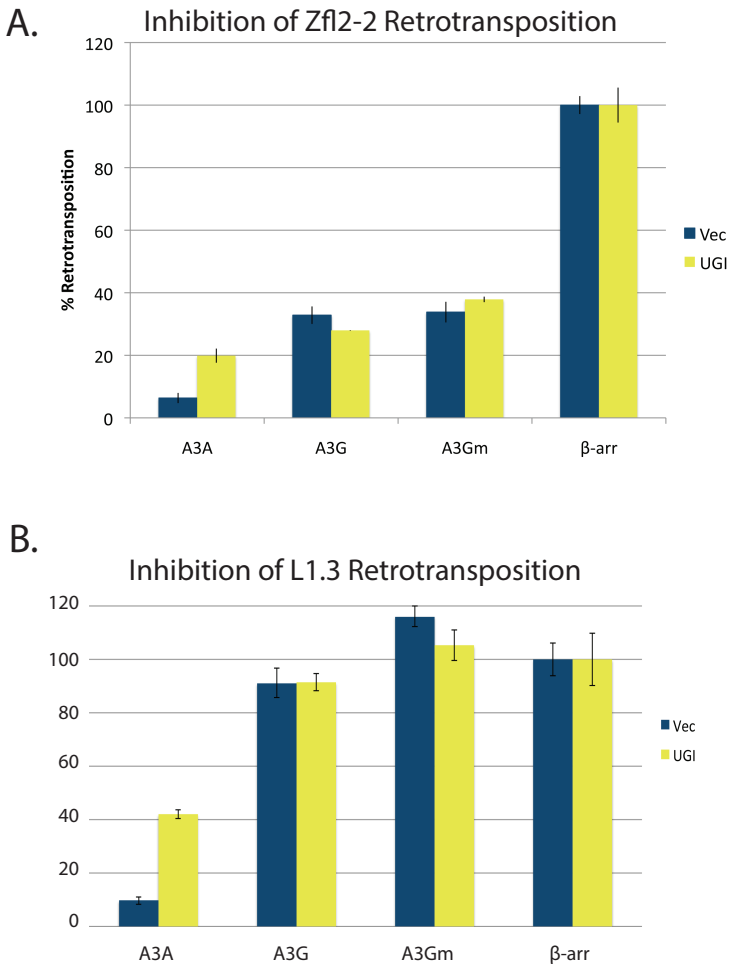


Figure 2.5: Deaminase-Independent Inhibition of Zfl2-2 Retrotransposition by APOBEC3G

The data in A and B are from the same experiment, but are depicted separately for clarity. HeLa-JVM cells were plated in 10 cm dishes at appropriate density for the retroelement used ($4x10^5$ for Zfl2-2, $1x10^5$ for L1.3, $2x10^5$ for circular NEO). Each dish was transfected with a total of 3 µg plasmid DNA: 1 µg Zfl2-2, pJM101/L1.3, or circular NEO, 1 µg APOBEC3 plasmid (or β-arr control), and 1 µg pLGCX_UGI or empty vector control).

The APOBEC3 (or β-arr control) plasmid used is indicated on the x-axes. A3Gm is a deaminase-deficient mutant of APOBEC3G [2]. The y-axis indicates percent retrotransposition relative to the appropriate β-arr control co-transfection. Navy blue bars represent LGCX empty vector control, yellow bars represent LGCX_UGI. Error bars represent percent standard deviation between duplicate transfections. Data were normalized to circular NEO control co-transfections carried out as described in Figure 3.3. A representative assay is shown; this assay was performed twice representing biological replicates.

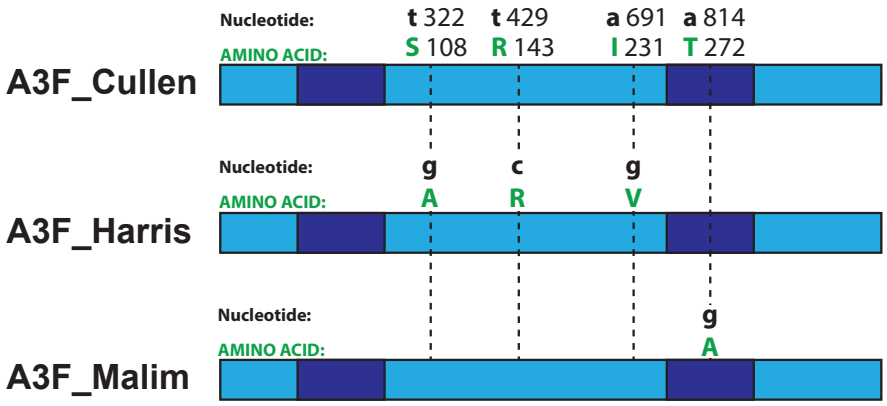
Figure 2.6: APOBEC3F does not inhibit L1 in HeLa-JVM Cells

A. APOBEC3F construct sequences from the Cullen (A3FC) [20], Harris (A3FH) [25], and Malim (A3FM) [63] labs. The A3F coding sequence is depicted as a light blue rectangle; dark blue rectangles indicate cytidine deaminase active sites. Positions of nucleotide (black) and amino acid (green) residues where A3FH and A3FM differ from A3FC are indicated.

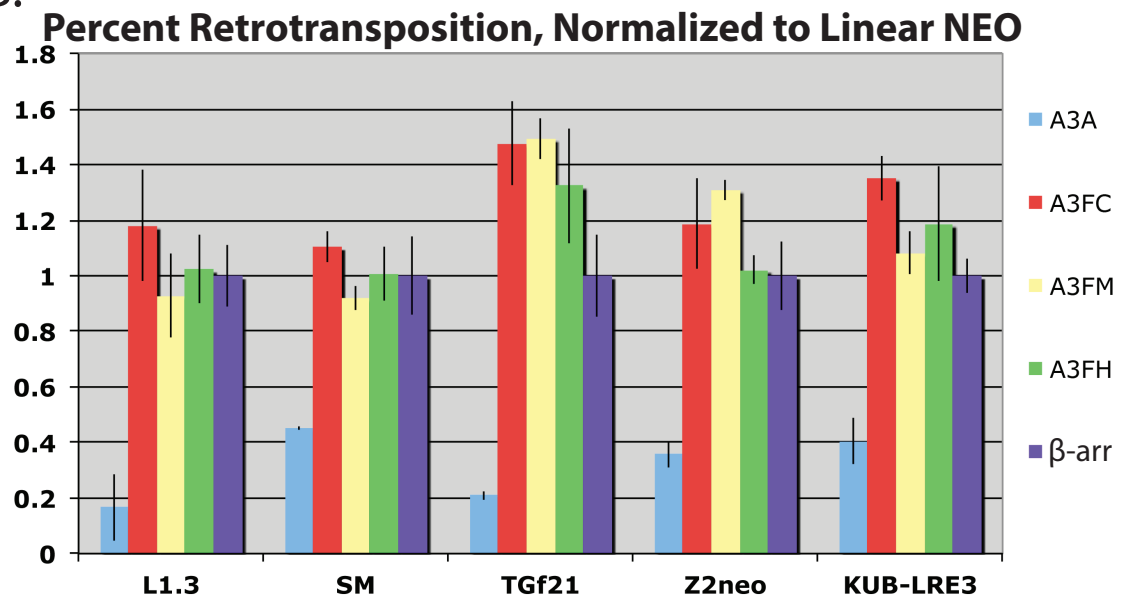
B. APOBEC3F constructs from the Cullen lab, Malim lab, and Harris lab do not inhibit L1 retrotransposition in HeLa-JVM cells. Retrotransposition assays: Approximately $2x10^3$ HeLa-JVM cells were plated per well of a 6-well dish. Each well was transfected with a total of 1 µg plasmid DNA comprised of 0.5 µg L1 reporter plasmid (or linear NEO control plasmid) and 0.5 µg APOBEC3F (or control β-arr) plasmid. Transfections were performed in triplicate per experimental condition.

The LINE element is indicated on the x-axis. Percent retrotransposition relative to β-arr control is indicated on the y-axis. Blue bars represent A3A, red bars represent A3F-Cullen, yellow bars represent A3F-Malim, green bars represent A3F-Harris, and violet bars represent β-arr control. Error bars represent percent standard deviation among triplicate co-transfections. Data are normalized to linear NEO control assays carried out as depicted in Figure 3.3. A representative assay is shown, this experiment was performed three times representing biological replicates.

A.


Nucleotide:
t 322
t 429
a 691
a 814
AMINO ACID:
S 108
R 143
I 231
T 272
A3F_Cullen
Nucleotide:
g
c
g
AMINO ACID:
A
R
V
A3F_Harris
Nucleotide:
g
AMINO ACID:
A
A3F_Malim

B.


Percent Retrotransposition, Normalized to Linear NEO
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
L1.3
SM
TGf21
Z2neo
KUB-LRE3
A3A
A3FC
A3FM
A3FH
β-arr

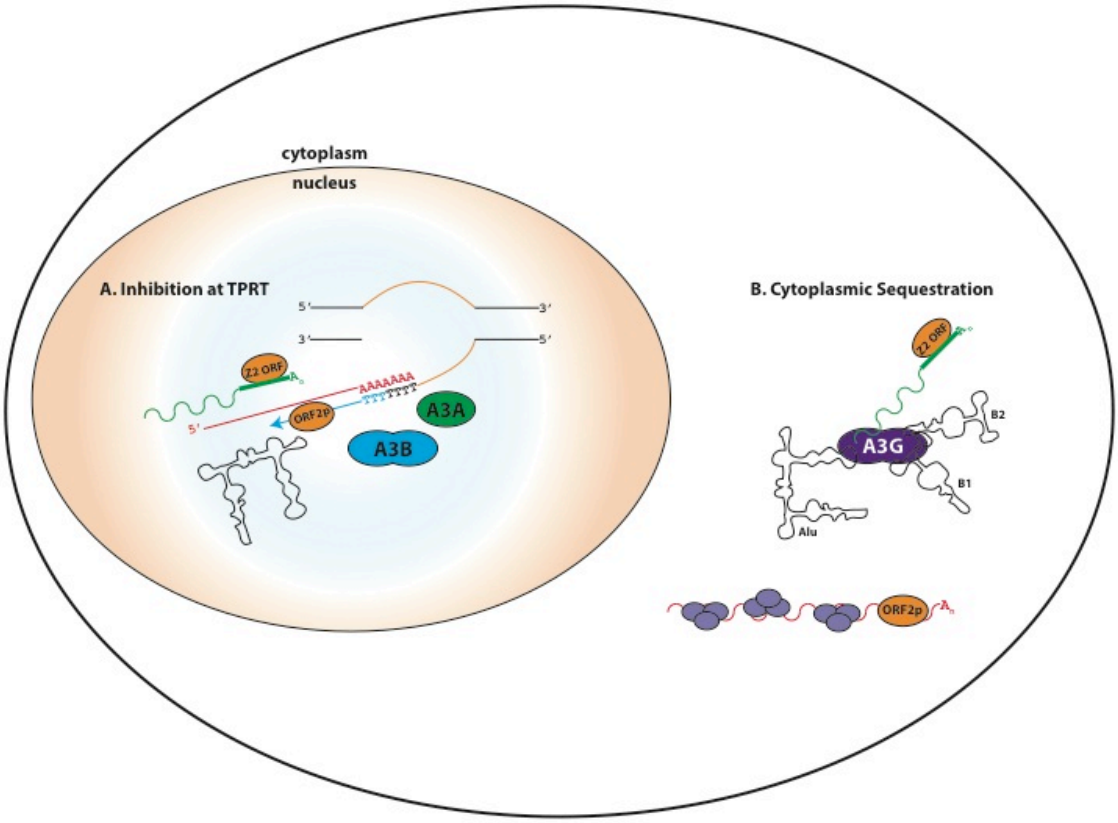


Figure 2.7: Hypotheses for retroelement inhibition by A3A, A3B, and A3G

A. Inhibition at target-primed reverse transcription (TPRT). A3A (green) and A3B (blue) can access the nucleus and are predicted to interfere with TPRT of autonomous and non-autonomous retroelements. Single deaminase-domain A3s are depicted as a single oval; double deaminase-domain A3s are depicted as a two-lobed shape. L1 RNA is depcited in red, nascent L1 cDNA is depicted in blue. Zfl2-2 RNA is depicted in green with a bold green line indicating the Zfl2-2 3'UTR critical for Z2 ORFp binding [36]. The L1 ORF2p and Zfl2-2 ORFp are depicted in orange. Representative non-autonomous retroelement RNA is depicted in black.

B. Inhibition by cytoplasmic sequestration. A3G (purple) is cytoplasmic, and is predicted to restrict the mobilization of non-autonomous retroelements and Zfl2-2 by sequestering retroelement RNAs in the cytoplasm. Zfl2-2 RNA and protein are depicted in green and orange, respectively. Alu, B1, and B2 RNAs are shown in black. In contrast, LINE-1 ORF1p (lavendar ovals) is hypothesized to protect the L1 RNA (red) from sequestration.

References

1. Chiu, Y.L., et al., *High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15588-93.
2. Hulme, A.E., et al., *Selective inhibition of Alu retrotransposition by APOBEC3G.* Gene, 2007. **390**(1-2): p. 199-205.
3. Refsland, E.W. and R.S. Harris, *The APOBEC3 Family of Retroelement Restriction Factors.* Current topics in microbiology and immunology, 2013. **371**: p. 1-27.
4. Navaratnam, N., et al., *The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase.* The Journal of biological chemistry, 1993. **268**(28): p. 20709-12.
5. Teng, B., C.F. Burant, and N.O. Davidson, *Molecular cloning of an apolipoprotein B messenger RNA editing protein.* Science, 1993. **260**(5115): p. 1816-9.
6. Conticello, S.G., et al., *Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases.* Mol Biol Evol, 2005. **22**(2): p. 367-77.
7. Muramatsu, M., et al., *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.* Cell, 2000. **102**(5): p. 553-63.
8. Muramatsu, M., et al., *Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells.* J Biol Chem, 1999. **274**(26): p. 18470-6.
9. Arakawa, H., J. Hauschild, and J.M. Buerstedde, *Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion.* Science, 2002. **295**(5558): p. 1301-6.
10. Liao, W., et al., *APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family.* Biochem Biophys Res Commun, 1999. **260**(2): p. 398-404.
11. Rogozin, I.B., et al., *Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase.* Nat Immunol, 2007. **8**(6): p. 647-56.
12. Conticello, S.G., *The AID/APOBEC family of nucleic acid mutators.* Genome Biol, 2008. **9**(6): p. 229.
13. Jarmuz, A., et al., *An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.* Genomics, 2002. **79**(3): p. 285-96.
14. Dang, Y., et al., *Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family.* J Virol, 2006. **80**(21): p. 10522-33.
15. OhAinle, M., et al., *Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H.* J Virol, 2006. **80**(8): p. 3853-62.
16. Chiu, Y.L. and W.C. Greene, *The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements.* Annu Rev Immunol, 2008. **26**: p. 317-53.

17. Sawyer, S.L., M. Emerman, and H.S. Malik, *Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G.* PLoS Biol, 2004. **2**(9): p. E275.
18. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
19. Beck, C.R., et al., *LINE-1 Elements in Structural Variation and Disease.* Annu Rev Genomics Hum Genet, 2010.
20. Bogerd, H.P., et al., *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8780-5.
21. Chen, H., et al., *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.* Curr Biol, 2006. **16**(5): p. 480-5.
22. Muckenfuss, H., et al., *APOBEC3 proteins inhibit human LINE-1 retrotransposition.* J Biol Chem, 2006. **281**(31): p. 22161-72.
23. Kinomoto, M., et al., *All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.* Nucleic Acids Res, 2007. **35**(9): p. 2955-64.
24. Niewiadomska, A.M., et al., *Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association.* J Virol, 2007. **81**(17): p. 9577-83.
25. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.* J Biol Chem, 2006. **281**(25): p. 16837-41.
26. Esnault, C., et al., *APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses.* Nature, 2005. **433**(7024): p. 430-3.
27. Turelli, P., S. Vianin, and D. Trono, *The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay.* J Biol Chem, 2004. **279**(42): p. 43371-3.
28. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
29. Wei, W., et al., *A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events.* Anal Biochem, 2000. **284**(2): p. 435-8.
30. Stenglein, M.D., et al., *APOBEC3 proteins mediate the clearance of foreign DNA from human cells.* Nat Struct Mol Biol, 2010. **17**(2): p. 222-9.
31. Bulliard, Y., et al., *Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities.* J Virol, 2011. **85**(4): p. 1765-76.
32. Landry, S., et al., *APOBEC3A can activate the DNA damage response and cause cell-cycle arrest.* EMBO Rep, 2011. **12**(5): p. 444-50.
33. Goodier, J.L., et al., *A novel active L1 retrotransposon subfamily in the mouse.* Genome Res, 2001. **11**(10): p. 1677-85.
34. Naas, T.P., et al., *An actively retrotransposing, novel subfamily of mouse L1 elements.* EMBO J, 1998. **17**(2): p. 590-7.
35. Han, J.S. and J.D. Boeke, *A highly active synthetic mammalian retrotransposon.* Nature, 2004. **429**(6989): p. 314-8.

36. Sugano, T., M. Kajikawa, and N. Okada, *Isolation and characterization of retrotransposition-competent LINEs from zebrafish.* Gene, 2006. **365**: p. 74-82.
37. Esnault, C., et al., *Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses.* Nucleic Acids Res, 2006. **34**(5): p. 1522-31.
38. Dewannieux, M., C. Esnault, and T. Heidmann, *LINE-mediated retrotransposition of marked Alu sequences.* Nat Genet, 2003. **35**(1): p. 41-8.
39. Weiner, A.M., *An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome.* Cell, 1980. **22**(1 Pt 1): p. 209-18.
40. Ullu, E. and C. Tschudi, *Alu sequences are processed 7SL RNA genes.* Nature, 1984. **312**(5990): p. 171-2.
41. Rogers, J.H., *The origin and evolution of retroposons.* International review of cytology, 1985. **93**: p. 187-279.
42. Weiner, A.M., P.L. Deininger, and A. Efstratiadis, *Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information.* Annual review of biochemistry, 1986. **55**: p. 631-61.
43. Dewannieux, M. and T. Heidmann, *L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells.* J Mol Biol, 2005. **349**(2): p. 241-7.
44. Wei, W., et al., *Human L1 retrotransposition: cis preference versus trans complementation.* Mol Cell Biol, 2001. **21**(4): p. 1429-39.
45. Sisodia, S.S., B. Sollner-Webb, and D.W. Cleveland, *Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation.* Molecular and cellular biology, 1987. **7**(10): p. 3602-12.
46. Smit, A.F., *Interspersed repeats and other mementos of transposable elements in mammalian genomes.* Current opinion in genetics & development, 1999. **9**(6): p. 657-63.
47. Gibbs, R.A., et al., *Genome sequence of the Brown Norway rat yields insights into mammalian evolution.* Nature, 2004. **428**(6982): p. 493-521.
48. Olsen, L.C., et al., *Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme.* The EMBO journal, 1989. **8**(10): p. 3121-5.
49. Shaper, N.L. and L. Grossman, *Purification and properties of the human placental apurinic/apyrimidinic endonuclease.* Methods in enzymology, 1980. **65**(1): p. 216-24.
50. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-8.
51. Bogerd, H.P., et al., *APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.* Nucleic Acids Res, 2006. **34**(1): p. 89-95.

52. Dombroski, B.A., A.F. Scott, and H.H. Kazazian, Jr., *Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element.* Proc Natl Acad Sci U S A, 1993. **90**(14): p. 6513-7.
53. Sassaman, D.M., et al., *Many human L1 elements are capable of retrotransposition.* Nat Genet, 1997. **16**(1): p. 37-43.
54. Kulpa, D.A. and J.V. Moran, *Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition.* Hum Mol Genet, 2005. **14**(21): p. 3237-48.
55. Wallace, M.R., et al., *A de novo Alu insertion results in neurofibromatosis type 1.* Nature, 1991. **353**(6347): p. 864-6.
56. Esnault, C., J.F. Casella, and T. Heidmann, *A Tetrahymena thermophila ribozyme-based indicator gene to detect transposition of marked retroelements in mammalian cells.* Nucleic acids research, 2002. **30**(11): p. e49.
57. Kaiser, S.M. and M. Emerman, *Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G.* J Virol, 2006. **80**(2): p. 875-82.
58. Miller, A.D. and G.J. Rosman, *Improved retroviral vectors for gene transfer and expression.* BioTechniques, 1989. **7**(9): p. 980-2, 984-6, 989-90.
59. Holmes, R.K., M.H. Malim, and K.N. Bishop, *APOBEC-mediated viral restriction: not simply editing?* Trends Biochem Sci, 2007. **32**(3): p. 118-28.
60. Harris, R.S. and M.T. Liddament, *Retroviral restriction by APOBEC proteins.* Nat Rev Immunol, 2004. **4**(11): p. 868-77.
61. Freeman, J.D., N.L. Goodchild, and D.L. Mager, *A modified indicator gene for selection of retrotransposition events in mammalian cells.* BioTechniques, 1994. **17**(1): p. 46, 48-9, 52.
62. Alisch, R.S., et al., *Unconventional translation of mammalian LINE-1 retrotransposons.* Genes Dev, 2006. **20**(2): p. 210-24.
63. Bishop, K.N., et al., *Cytidine deamination of retroviral DNA by diverse APOBEC proteins.* Curr Biol, 2004. **14**(15): p. 1392-6.

Chapter 3

A Molecular Mechanism for APOBEC3A-Mediated LINE-1 Inhibition

I designed and carried out the retrotransposition assays, LEAP assays, and L1 insertion recoveries detailed in this chapter. Dr. Inigo Narvaiza and Dr. Matt Weitzman generated and characterized the purified recombinant A3A and A3A mutant proteins (Figures 3.7 and 3.8). Randy Planegger provided technical assistance for the retrotransposition assay shown in Figure 3.14. The L1 insertion recoveries were carried out using modifications to the original protocol [1, 2], which were designed and optimized by Santiago Morell in the laboratory of Dr. Jose Luis Garcia-Perez.

Abstract

In this chapter, a molecular mechanism for A3A-mediated L1 inhibition was investigated. I found that A3A inhibits LINE element retrotransposition in a sequence-independent manner, and that A3A does not directly block L1 endonuclease or reverse transcriptase activities. In an *in vitro* assay for L1 TPRT (L1 element amplification protocol, or LEAP), I observed that A3A can

deaminate exposed single-stranded DNA in TPRT intermediates. I therefore hypothesized that A3A inhibits L1 in a deaminase-dependent manner, by editing TPRT intermediates and thereby triggering their degradation by the cellular repair factors uracil DNA glycosylase (UNG) and apurinic/apyrimidinic endonuclease (APE). I predicted that expression of uracil glycosylase inhibitor protein (UGI) in the cell culture assay would simultaneously alleviate A3A-mediated L1 inhibition, and preserve evidence of editing within retrotransposed sequences. Indeed, UGI alleviates A3A-mediated L1 inhibition in cultured cells. Furthermore, L1 insertions generated in the presence of A3A and UGI bear substantial evidence of A3A-mediated editing. Thus, I conclude that A3A effects inhibition of L1, at least in part, by deaminating single-stranded DNA transiently exposed during L1 TPRT, providing the first mechanistic explanation for APOBEC3-mediated autonomous retroelement inhibition.

Introduction

The seven human APOBEC3 (A3) cytidine deaminases are intracellular defenders of genome integrity. The identification of APOBEC3G (A3G) as a restriction factor against *vif*-deficient HIV infectivity [3] first drew attention to the A3 family as a component of innate immunity against exogenous pathogens. Subsequent reports demonstrated that APOBEC3F (A3F) also restricts *vif*-deficient HIV [4, 5], while A3B inhibits wild-type HIV [6]. In addition, A3B, A3F, and A3G restrict hepatitis B virus (HBV) [7-10], and APOBEC3A (A3A) can inhibit replication of adeno-associated virus (AAV) [11]. Based on evidence for ancient

positive selection of the A3 genes, however, it is hypothesized that conflict with endogenous retroelements has been a crucial force in A3 evolution [12]. Indeed, several studies have demonstrated that various A3 proteins, including A3A, can inhibit Long INterspersed Element-1 (LINE-1 or L1) mobility in cultured cells [11, 13-17], yet the mechanism by which A3A restricts L1 retrotransposition remains enigmatic. In this chapter, we investigate the molecular mechanism of A3A-mediated L1 inhibition.

Sequence-Independent Inhibition of LINE Retrotransposition by A3A

To investigate the mechanism of A3A-mediated inhibition of L1 retrotransposition, we first asked whether inhibition depends on the nucleotide sequence of the retroelement. We employed a cell culture retrotransposition assay [18, 19], and compared inhibition of human L1.3 [20] to a natural mouse L1 element (TG_f21) [21], a synthetic mouse L1 element (L1SM) [22], and a LINE-2 element from zebrafish (Zfl2-2) [23] (Figure 3.1 A). The nucleotide sequences of the ORFs encoded by these elements differ from L1.3 to varying extents (25.8%-59.3%)(Figure 3.1 B). In contrast to the human and mouse L1s, Zfl2-2 contains only one ORF, which encodes endonuclease and reverse transcriptase activities required for retrotransposition (Figure 3.1 A) [23].

Consistent with previous reports, A3A expression inhibited retrotransposition of L1.3 to 25.5% of β-arrestin control levels [13]. A3A inhibited TGf21 and Zfl2-2 to 22.2% and 19.1% of β-arrestin, respectively. A3A inhibited L1SM to a modest extent (60.6% of β-arrestin)(Figure 3.1 D). The partial

resistance of L1SM to A3A-mediated inhibition may be attributable to the elevated transcript levels, RNP production, and retrotransposition frequency of this element [22]. The deaminase-deficient mutant A3A-C106S did not significantly restrict retrotransposition of any of the elements tested, consistent with previous reports (Figure 3.1 D)[11]. Thus, the inhibition of LINE retrotransposition by A3A occurs independently of retroelement nucleotide sequence. Additional controls confirmed that the reduction of G418-resistant colonies observed in the presence of A3A reflects a specific inhibition of retrotransposition, beyond a non-specific cytotoxicity (Figure 3.2 A-C). Finally, A3A inhibited retrotransposition when L1 was expressed from a non-episomal vector (pBSKS) and when L1 expression was driven by the native L1 promoter; thus, inhibition does not arise from A3A-mediated interference with a replicating episome or from inhibition of transcription from a viral promoter (Figure 3.3).

A3A does not inhibit L1 endonuclease activity

The LINE elements inhibited by A3A (Figure 3.1) all require element-encoded endonuclease (EN) and reverse transcriptase (RT) activities for their retrotransposition, suggesting that A3A may inhibit a discrete step in TPRT [24]. While EN-deficient L1s cannot mobilize in wild-type cells, they retrotranspose efficiently in certain cells deficient in p53 and in the non-homologous end joining (NHEJ) pathway of DNA repair [25-27]. If A3A specifically blocks L1 endonuclease activity, endonuclease-independent (ENi) retrotransposition events should escape inhibition (Figure 3.4 A). Thus, we assessed the ability of A3A to

inhibit wild-type and ENi retrotransposition in wild-type (4364a) and XRCC4-deficient (XR-1) Chinese Hamster Ovary (CHO) cells. As predicted, A3A inhibited wild-type L1 retrotransposition in 4364a and XR-1 cells, to 3.7% and 10.8% of β-arr, respectively (Figure 3.4 B), demonstrating that A3A can effect L1 inhibition in non-human cells. We found that A3A inhibited ENi retrotransposition in XR-1 cells to 1.8% of control levels (Figure 3.4 C). Thus, we conclude that A3A does not specifically inhibit L1 endonuclease activity.

A3A does not block L1 reverse transcription

To determine if A3A inhibits L1 RT activity, we tested the effect of purified recombinant A3A protein (rA3A or rA3A_C106S) in an *in vitro* assay for L1 reverse transcription termed L1 Element Amplification Protocol (LEAP), which mimics the initial steps of TPRT (Figure 3.6 A)[28]. We purified recombinant A3A (rA3A) fused to a histidine tag from E. coli by Ni^{++}-affinity followed by gel filtration. (Figure 3.7). Control experiments using a fluorescein-labeled ssDNA oligonucleotide containing a single cytosine demonstrate that rA3A is active in UDG-dependent deaminase assays (Figure 3.8 A, left), while purified rC106S did not show detectable levels of deaminase activity (Figure 3.8 A, right). In agreement with previous reports [11], rA3A did not show activity on dsDNA (Figure 3.8 B).

Our initial experiments revealed a dramatic loss of LEAP products in the presence of rA3A, but not rA3A_C106S, which we interpreted as A3A-mediated inhibition of L1 RT procession (Figure 3.5 A). However, we were concerned that

the presence of a cytidine residue in the LEAP adapter (5np1) might be allowing rA3A to interfere with the LEAP reaction indirectly; for example, by binding and sequestering the LEAP primer away from the L1 enzymatic machinery. We therefore designed a LEAP adapter with no cytidine residues (5np0), and carried out LEAP reactions in the presence of rA3A. Strikingly, rA3A did not result in a decrease in LEAP products primed from the 5np0 adapter (Figure 3.5 A). Moreover, sequencing of rA3A LEAP products primed with 5np0, and residual products from reactions primed with 5np1, revealed no evidence of deamination of the adapter or the L1 cDNA. Thus, we were faced with an rA3A-mediated reduction of LEAP products that was dependent on the presence of a cytidine residue in the LEAP adapter, but which appeared to occur without deamination of LEAP products.

Based on the above result, we hypothesized that rA3A could deaminate the cytidine residue in the 5np1 adapter, and that a uracil DNA glycosylase activity present in the L1 RNP preparation was responsible for removing the resulting uracil, leading to degradation of the adapter at the abasic site during the 15-minute 100° C incubation step between the L1 RT reaction and PCR amplification (Figure 3.6 A). Inclusion of recombinant uracil glycosylase inhibitor (UGI) protein in the LEAP reaction with rA3A did not restore LEAP products, leading us to speculate that the cellular uracil DNA glycosylase activity must be provided not by UNG2 but by SMUG1, which is not affected by UGI [29] (Figure 3.5 A). Consistently, in the absence of rA3A, no LEAP products were obtained from a “pre-deaminated” adapter containing a single uracil residue (5np1.U)

(Figure 3.5 B). We next sought to confirm that a uracil DNA glycosylase activity was present in the L1 RNP, by using the 5np1.U adapter to prime MMLV and AMV reverse transcription reactions on purified RNA from L1 RNP preparations, followed by PCR amplification. These reactions contain no rA3A protein and no undefined components (*i.e.*, no putative cellular uracil DNA glycosylase activity), and we therefore expected to generate robust RT products from the 5np1.U adapter. Strikingly, no MMLV RT products were detectable from the 5np1.U adapter (Figure 3.5 C).

At this point, we speculated that the presence of a uracil residue in the adapter was not interfering with reverse transcription, but with the PCR amplification step required to detect RT products. Indeed, we ultimately realized that *PfuTurbo* Hotstart DNA polymerase (Agilent), which we had used in the PCR amplification step of LEAP, cannot process through uracils in the template DNA strand. We reasoned that cytidine residues in LEAP products generated in the presence of rA3A were likely being deaminated by rA3A, resulting in uracil residues that impede the procession of *PfuTurbo*, blocking PCR amplification and detection of LEAP products. We therefore switched to a variant of this enzyme, *PfuTurbo* C_x Hotstart DNA polymerase (Agilent), which can process through uracils in template DNA. When *PfuTurbo* C_x was used to amplify LEAP products, we detected robust LEAP products in the presence of rA3A, regardless of whether the adapter contained a cytidine residue (Figure 3.6 B). Control experiments demonstrated that rA3A and rA3A_C106S do not interfere with *in*

vitro MMLV reverse transcription (Figure 3.9). Thus, we conclude that A3A does not directly block L1 RT procession.

A3A deaminates L1 TPRT intermediates *in vitro*

Previous studies did not detect deamination of L1 insertions generated in the presence of A3A [13]; however, rapid degradation of deaminated substrates by uracil DNA glycosylase (UNG) and apurinic/apyrimidinic endonuclease (APE) may confound detection of such editing *in vivo.* To determine if A3A can deaminate L1 RT products *in vitro,* we characterized 100 LEAP products generated in the presence or absence of rA3A or rA3A_C106S. We did not detect significant editing of control RNP-only or rA3A_C106S products (Figure 3.10 A and C). In rA3A-containing reactions, we expected to uncover evidence of rA3A-mediated editing throughout the LEAP products. To our surprise, we detected little evidence of editing within the L1 cDNA portion of the LEAP products. However, we uncovered robust editing of the single-stranded adapter used to prime the LEAP reaction. For example, among 100 products generated in the presence of rA3A, 91 exhibited a C-to-T change within the single 5'-TCA-3' trinucleotide (a previously published preferred A3A target trinucleotide [11]) in the single-stranded LEAP adapter (Figure 3.10 B, far right; Table 3.1). To make a relevant comparison, we quantified how frequently 5'-TCA-3' trinucleotides within the L1 cDNA were deminated in the presence of rA3A. The L1 cDNA portion of each LEAP product contains six 5'-TCA-3' trinucleotide sequences. We therefore consider 100 L1 cDNAs to contain a total of 600 5'-TCA-3' substrates

that could potentially be edited by rA3A. Among 600 total 5'-TCA-3' substrates in the cDNA portion of 100 products, 25 were deaminated, representing editing of only 4.2% of available 5'-TCA-3' substrates (Table 3.1).

The above results suggested that the L1 RNA remains annealed to the L1 cDNA during reverse transcription, protecting the L1 cDNA and leaving only the single-stranded LEAP adapter vulnerable to deamination. To test this hypothesis, we included recombinant RNaseH in LEAP reactions with rA3A. In RNP-only and rA3A_C106S control reactions, 25 out of 600 (4.2%) and 40 out of 600 (6.7%) of available 5'-TCA-3' substrates within the cDNA portion were deaminated, respectively (Figure 3.10 D and F). This result is consistent with the presence of an endogenous deaminase activity in HeLa cell RNP preparations; notably, APOBEC3B (A3B) is expressed in HeLa cells, and siRNA-mediated A3B knockdown is correlated with a moderate increase in engineered L1 retrotransposition [30]. In contrast, in the presence of rA3A and RNaseH, 228 out of 600 (38.0%) 5'-TCA-3' substrates within the L1 cDNA were deaminated (Table 3.1, Figure 3.10 E).

When assessing the extent of A3A-mediated editing on LEAP products amplified by PCR, it is critical to confirm that sequenced PCR products actually represent unique, independent LEAP products and not clonal amplification of a few LEAP products. Therefore, for the 100 LEAP products generated in the presence of rA3A and RNaseH, we compared the number and location of edited bases on each of the products, along with the poly-A tail length of each product. As illustrated in Table 3.2, the 100 LEAP products were generated from four

independent LEAP reactions. Among the products from each independent LEAP reaction, no two cloned products were identical with respect to number of deaminations, location of deaminations, and poly-A tail length. Thus, we confirm that our sequenced PCR products represent unique L1 RT products. The frequency of deamination events per LEAP product, for all experimental conditions, is shown in Figure 3.11.

Uracil glycosylase inhibitor (UGI) alleviates A3A-mediated L1 inhibition

The above data suggest that rA3A can deaminate single-stranded DNA exposed during TPRT *in vitro*. We therefore hypothesized that A3A inhibits L1 retrotransposition in cultured cells by deaminating TPRT intermediates, leading to their degradation by UNG and APE. Blockage of UNG by a uracil glycosylase inhibitor (UGI) [31] should stabilize deaminated TPRT intermediates, simultaneously alleviating A3A-mediated inhibition and preserving signatures of deamination at nascent L1 insertions (Figure 3.12 A). To test this prediction, we first examined the impact of UGI expression on L1 retrotransposition efficiency in the presence of A3A (Figure 3.12 B, left branch). As predicted, co-expression of UGI and A3A resulted in a ~2-fold alleviation of A3A-mediated inhibition relative to A3A and empty vector control (Figure 3.12 C). Colony counts indicate that UGI expression has no effect on L1 retrotransposition in the absence of A3A (Figure 3.13).

To confirm that alleviation of inhibition by UGI is deaminase-dependent, we employed APOBEC3B (A3B) and A3B catalytic mutants, which have been

demonstrated to inhibit L1 in a deaminase-independent manner [13, 16]. UGI expression marginally alleviated L1 inhibition by wild-type A3B, but had no effect on L1 inhibition by A3B deaminase-deficient mutants (Figure 3.14).

Over-expression of UGI also modestly relieved off-target effects of A3A in NEO control assays (Figure 3.15 B). To confirm that we were observing a *bona fide* alleviation of retrotransposition inhibition and not solely relief of off-target effects of A3A expression, we carried out retrotransposition assays using a reduced amount of A3A plasmid (0.25 µg instead of 1.0 µg). In this case, we observed virtually no effect by A3A on NEO control experiments (Figure 3.15 E), yet a ~2-fold inhibition of retrotransposition persisted (Figure 3.15 D). We incrementally increased the amount of UGI plasmid in the co-transfections, and found that UGI expression consistently relieved A3A-mediated inhibition to ~80% of control levels (Figure 3.15 C).

Additional controls demonstrated that UGI specifically alleviated A3A-mediated inhibition of L1, independently of relieving A3A cytotoxicity (Figure 3.15 A-E). In sum, we conclude that UGI expression specifically alleviates A3A-mediated inhibition of L1 by mitigating the downstream consequences of DNA deamination.

A3A deaminates single-stranded DNA transiently exposed during TPRT

We next examined L1 insertions occurring in the presence of A3A and UGI for signatures of deamination. We co-transfected cells with A3A, UGI, and JM140-L1.3-Δ^2-k7, a retrotransposition indicator plasmid which delivers a ColE1

bacterial origin of replication along with the NEO gene, allowing recovery of L1 insertions and flanking genomic DNA as autonomously replicating plasmids in bacteria [1]. We selected G418-resistant colonies to establish twelve insertion-containing cell lines (Figure 3.12 B, right branch). Upon characterizing the insertions, we observed four incidents of G-to-A mutation within the retrotransposed sequence, which correspond to deamination events on the first strand L1 cDNA (Figure 3.16 A). The 5' target-site duplication (TSD) of one insertion contained a C-to-T change within the preferred A3A target site 5'-TCA-3', suggesting that A3A can act on transiently single-stranded genomic DNA during TPRT (Figure 3.16 B).

To amass further evidence for A3A-mediated editing of L1 retrotransposition events, we employed a stable UGI-expressing U2OS cell line [32] (Figure 3.17 A, left branch). This strategy ensures that UNG activity is blocked before introduction of A3A and L1 plasmids into the cell, as determined by subjecting cell lysates to an assay for uracil DNA glycosylase-dependent cleavage of a fluorescein-labeled oligonucleotide bearing a single uracil residue [32]. As compared to a control U2OS cell line, U2OS_UGI cell lysates contained no detectable uracil DNA glycosylase activity in this assay [32]. In control U2OS cells, 25 ng of A3A plasmid per transfection inhibited retrotransposition to 26.0% of β-arrestin control levels without substantial toxicity (Figure 3.17 A and B, Figure 3.18 A). In U2OS_UGI cells, retrotransposition in the presence of 25 ng of A3A was alleviated to 87.6% of β-arrestin control levels (Figure 3.17 B; Figure

3.18 B). Thus, we conclude that A3A inhibits L1 retrotransposition in U2OS cells, and that this inhibition is relieved by stable UGI expression.

We next used JM140-L1.3-Δ^2-k7 to generate a panel of clonal U2OS_UGI cell lines harboring L1 insertions (Figure 3.17 A, right branch). Among 33 insertions generated in the presence of A3A, representing ~123 kb of retrotransposed sequence, we detected 39 G-to-A changes on the plus-strand of the L1 insertions, far outnumbering any other nucleotide change (Figure 3.19, Figure 3.20). In contrast, among 30 insertions (~107 kb) generated in the presence of C106S, only five G-to-A changes were detected, and among 24 insertions with β-arrestin (~81 kb), we did not detect any G-to-A changes (Figure 3.18, Figure 3.20, Figure 3.21). Thus, we conclude that A3A can deaminate single-stranded DNA during L1 TPRT, providing a probable mechanistic explanation for A3A-mediated L1 inhibition.

Discussion

The data indicate that A3A inhibits L1 retrotransposition in a sequence-independent manner, and without directly inhibiting L1 EN activity or RT procession. We find that recombinant A3A protein (rA3A) can deaminate L1 RT products *in vitro*, and that the availability of single-stranded DNA dictates the specific signatures of deamination on L1 RT products.

In the LEAP reaction *in vitro,* in the absence of RNaseH, the L1 RNA and first-strand cDNA remain annealed in a heteroduplex, and deamination is restricted to the single-stranded LEAP adapter which simulates the free genomic

DNA end used to prime TPRT *in vivo*. Thus, we speculate that when cellular RNaseH is not abundant, the L1 RNA and cDNA could remain annealed in a protective heteroduplex protecting the first-strand L1 cDNA until second strand synthesis occurs. In contrast, when recombinant RNaseH is included in the LEAP reaction, C-to-U editing by A3A is evident throughout the L1 cDNA. Thus, when RNaseH is available *in vivo,* the L1 cDNA may be rendered vulnerable to A3A-mediated deamination. Notably, the L1-related R2 element of *Bombyx mori* can displace RNA during second-strand DNA synthesis *in vitro* [33], obviating the need for RNaseH activity during integration.

We next extended our *in vitro* findings to L1 retrotransposition events in cultured cells by employing a uracil glycosylase inhibitor (UGI) [31], since deaminated TPRT intermediates *in vivo* are predicted to be degraded by UNG and APE. Indeed, expression of UGI in cultured cells alleviates A3A-mediated L1 inhibition and reveals evidence for C-to-U editing of integrated L1 cDNAs as well as a genomic target-site duplication flanking an L1 integration. Remarkably, we observed a strand bias with regards to deamination events within retrotransposed insertions, with 39 G-to-A changes compared one C-to-T change on the plus-strand L1 DNA (Figure 3.19). This result is consistent with deamination of the minus-strand cDNA during TPRT, and is not consistent with deamination of plasmid DNA by A3A, which reportedly occurs on both DNA strands [34]. Furthermore, deamination of 5' flanking genomic DNA cannot easily be explained by editing of transfected plasmid DNA. Finally, among the G-to-A changes observed within retrotransposed sequence in U2OS_UGI cells,

one mutation resulted in a premature termination codon (Figure 3.20; cell line A0-43), which truncates the L1 ORF2p protein by 145 amino acids and reduces retrotransposition efficiency to ~7% of control levels (Figure 3.23), making it unlikely that this mutation occurred on the plasmid prior to retrotransposition. We therefore conclude that deamination of TPRT intermediates is responsible, at least in part, for A3A-mediated inhibition of L1 retrotransposition.

We propose a model for L1 inhibition by A3A in which A3A acts opportunistically on single-stranded DNA during L1 TPRT (Figure 3.24 and 3.25). A3A can deaminate the first-strand L1 cDNA if the L1 RNA has been removed, presumably through the action of cellular RNaseH (Figure 3.24 A, left). When UNG activity is inhibited by UGI, these C-to-U changes become fixed as C-to-T transitions, and are revealed as G-to-A changes on the top strand of the integrated L1 insertion (Figure 3.19, Figure 3.24 A, right). In the absence of UGI, removal of the uracil base by UNG and subsequent cleavage by APE would presumably result in degradation of the L1 cDNA (Figure 3.25 A). Notably, such cleavage of the L1 cDNA could give rise to a 5' truncated L1 insertion (Figure 3.25 A).

The transiently single-stranded genomic DNA regions, which ultimately give rise to L1 target site duplications, are a second potential substrate for A3A-mediated deamination during L1 TPRT. Deamination of the 5' flanking genomic DNA in the presence of UGI is predicted to result in mismatched TSDs, with the deamination event evident as a C-to-T change in the 5' TSD (Figure 3.24 B). Among the L1 insertions generated in HeLa cells, we uncovered one example of

editing consistent with deamination in this region (Figure 3.16 B). Therefore, deamination of 5' flanking genomic DNA may represent a mechanism by which A3A can interfere with L1 TPRT. Notably, the putative deamination event took place within the context 5'-TCA-3', which has been previously reported as a preferred A3A substrate [11], and also was preferentially deaminated by rA3A in our LEAP products (Table 3.1). In the absence of UGI, deamination in this region and resultant cleavage of the top-strand DNA could potentially substitute for second-strand cleavage during TPRT (Figure 3.25 B). The L1 integration event could then go to completion and would perhaps give rise to an insertion bearing shorter target-site duplications than if second-strand cleavage took place at the single-strand/double-strand DNA junction 3' of the TPRT intermediate. Notably, the structure of the transiently single-stranded genomic DNA 5' of the L1 insertion resembles the non-transcribed strand of an R-loop, which is acted upon by the APOBEC3-related activation induced deaminase (AID) *in vitro* [35], and which may be targeted by AID *in vivo* during class-switch recombination [36]. Thus, activity on a transiently single-stranded DNA substrate at a genomic locus may represent a conserved feature of APOBEC-related cytidine deaminases. Indeed, the cytotoxic effects of A3A may arise from a DNA damage response triggered by deamination and UNG-dependent degradation of single-stranded genomic DNA at structures such as transcription bubbles and replication forks [32].

A3A-mediated deamination of the 3' flanking genomic DNA of a TPRT intermediate in the presence of UGI is predicted to result in mismatched TSDs,

with the deamination event evident as a G-to-A change in the 3' TSD (Figure 3.23 C). We did not detect evidence for editing in this location among the retrotransposition events characterized from cultured cells. However, extensive editing of the single-stranded 3' RACE adapter in the LEAP assay suggests that 3' flanking genomic DNA may be a viable target for A3A *in vivo*, especially in the absence of cellular RNaseH. It is intriguing to speculate that in the absence of UGI, A3A-mediated editing within this region could lead to loss of the entire integration intermediate (Figure 3.25 C). Notably, the ERCC1/XPF endonuclease has been implicated in limiting L1 retrotransposition by cleaving this transiently single-stranded 3' genomic DNA during TPRT [37]. The DNA lesion resulting from cleavage of this region could be repaired using the top strand DNA as a template, leaving no detectable evidence of the thwarted retrotransposition event.

The data indicate that A3A inhibits L1 retrotransposition by deaminating transiently single-stranded DNA exposed during TPRT, providing the first mechanistic explanation for APOBEC3-mediated inhibition of an autonomous retroelement. Notably, engineered L1 insertions represent a readily recoverable sequence target, which we have exploited in conjuction with UGI to preserve and directly observe A3A-mediated editing of genomic substrates, without resorting to 3D-PCR-based techniques [34, 38, 39].

Given the expression profile of A3A in human tissues [11, 34], it is unlikely that L1, which must retrotranspose in the germline or early embryo in order to generate heritable insertions, is the physiological target of A3A. However,

endogenous APOBEC3B has been demonstrated to restrict L1 retrotransposition in human embryonic stem cells (hESCs) [30]. Thus, mechanistic insights from A3A-mediated L1 inhibition can be extended to A3B, which represents a likely physiological restriction factor for L1 retrotransposition.

Materials and Methods

Some of the text describing materials and methods in this chapter is duplicated from Chapter 2 of this thesis.

Plasmids

All plasmids were grown in DH5α (F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1) competent *E. coli* (Invitrogen; Carlsbad, CA. Prepared in house as described in [40]) and prepared using the Qiagen Plasmid Midi Kit (QIAGEN; Hilden, Germany) according to the manufacturer's protocol.

APOBEC3 expression constructs:

The pK_βarr, A3A, A3B, A3B_Nterm and A3B_CS expression plasmids have been described previously [41]. The APOBEC3A mutant C106S, described in [11] was received from Dr. Matt Weitzman and Dr. Inigo Narvaiza in pcDNA 3.1^+ (Invitrogen), and were subcloned into the pK vector using *Hind*III and *Xho*I restriction sites.

LINE expression constructs:

JM101/L1.3 has been described previously [20, 42], and consists of the pCEP4 backbone (Invitrogen/Life Technologies; Carlsbad, CA) containing a full-length copy of the L1.3 element and the *mneoI* indicator cassette.

pDK101 has been described previously [43], and consists of JM101/L1.3 modified by PCR mutagenesis to contain the T7 *gene 10* epitope tag at the C-terminus of ORF1p.

pDK135 has been described previously [28], and is identical to pDK101, except it contains the $D_{702}A$ mutation in the putative ORF2p reverse transcriptase active site [44].

pJJH230A/L1.3 has been described previously [45], and is similar to JM101/L1.3, except it contains the *mblastI* indicator cassette ([26]) in the 3' UTR instead of *mneoI*, and it contains the $H_{230}A$ mutation in the ORF2p endonuclease domain [44].

KS101/L1.3/sv+ consists of the pBSKS-II backbone (Stratagene/Agilent Technologies; Santa Clara, CA) containing a full-length L1.3 element with the *mneoI* indicator cassette in the 3'UTR, and the SV40 late polyadenylation signal.

JM102/L1.3 has been described previously [26], and is identical to JM101/L1.3, except it lacks the L1.3 5'UTR.

pCEP4/L1SM has been described previously [22], and consists of a synthetic mouse L1 sharing the same amino acid seqeunce as L1spa [46], but with 24% of its nucleotide sequence replaced for optimal GC-richness. It contains the *mneoI* indicator cassette in the 3'UTR, and is cloned in the pCEP4 backbone.

pCEP4/TGf21 has been described previously [21] and consists of a natural mouse element with the *mneoI* indicator cassette in the 3' UTR, cloned in the pCEP4 backbone.

pCEP4/Zfl2-2 has been described previously [23] and consists of a zebrafish LINE-2 element cloned in the pCEP4 backbone. The Zfl2-2 3'UTR is cloned 3' of the *mneoI* cassette.

pAD2TE1 has been described previously [47] and is similar to pDK101 except it contains aTAP tag on the C-terminus of ORF2p as well as the T7 *gene 10* epitope tag on the C-terminus of ORF1p.

pAD136 has been described previiously [47] and is identical to pAD2TE1 except it contains the $H_{230}A$ mutation in the ORF2p endonuclease domain.

JM140/L1.3/Δ^2/k7 has been described previously [1] and contains the L1.3 element cloned into pCEP4 backbone, but lacks the CMV promoter and the SV40 polyA signal. The *mneoI* indicator cassette and a ColE1 bacterial origin of replication are inserted in the L1.3 3'UTR.

LGCX vector has been described previously [31] and is a variant of the LNCX retroviral vector [48] in which the neomycin phosphotransferase gene has been replaced by GFP.

LGCX/UGI has been described previously [31] and consists of the LGCX vector containing a uracil glycosylase inhibitor (UGI) gene codon-optimized for expression in human cells (hUGI).

pU6i NEO is a pBSKS-based plasmid with the neomycin phosphotransferase (NEO) gene from pEGFPN1 (Clontech) introduced into the backbone. The multi-cloning site contains the U6 promoter. To generate linearized plasmid for control transfections, pU6i NEO was digested with *Bgl*II (New England Biolabs; Ipswitch, Massachussets), which does not disrupt NEO gene expression, and run on an agarose gel to confirm linearization. The restriction digest reactions were purified using the Qiagen gel extraction kit.

Cell Culture

HeLa cells: Cells were grown at 37°C in Dulbecco's modified Eagle medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine calf serum (FBS) (Invitrogen), 1X penicillin/streptomycin/glutamine (Invitrogen) in the presence of 7% CO_2 and 100% humidity.

CHO cells: Cells were grown at 37°C in DMEM-low glucose medium (Invitrogen) supplemented with 10% FBS (Invitrogen), 1X penicillin/streptomycin/glutamine (Invitrogen), and 1X non-essential amino acids (Invitrogen) in the presence of 7% CO_2 and 100% humidity.

U2OS_UGI and U2OS control Cells: Cells were grown at 37°C in DMEM (Invitrogen) supplemented with 10% FBS (Invitrogen) and 1X penicillin/streptomycin/glutamine (Invitrogen) in the presence of 7% CO_2 and 100% humidity.

The L1 Retrotransposition Assay

HeLa and U2OS cell retrotransposition assays were carried out as previously described [18, 19]. Cells were plated at an appropriate density in 6-well dishes (BD Biosciences; San Jose, California), T-75 flasks (BD Biosciences) or 10cm dishes (BD Biosciences or Corning; Corning, New York) to obtain quantifiable colonies for the retroelement expression construct used (see figure legends). Eighteen hours later, transfections were carried out using the FuGene

6 transfection reagent (Roche; Penzberg, Germany) and Opti-MEM (Invitrogen), according to manufacturer's protocol (3 µl FuGene and 97 µl Opti-MEM per µg of DNA transfected). Media was replaced the following day. At 72h post-transfection, cells were subjected to selection with 400 µg/ml G418 (Invitrogen) Selection was carried out for 12-14 days, replacing the selection media every other day. Colonies were washed with 1X phosphate buffered saline (PBS) (Gibco), fixed with 37% paraformaldehyde/8% glutaraldehyde, and stained with 0.1% crystal violet solution.

CHO cell retrotransposition assays were carried out in 4364a and XR-1 cell lines as previously described [26]. Cells were plated in T-75 flasks (BD Biosciences) at an appropriate density to obtain quantifiable colonies for the retroelement construct used (details are given for each experiment in the figure legends). Eight hours later, transfections were carried out using the FuGene 6 transfection reagent (Roche) and Opti-MEM (Invitrogen) according to manufacturer's protocol (3 µl FuGene and 97 µl Opti-MEM per µg of DNA transfected). Media was replaced the following day. At 72h post-transfection, cells were subjected to selection with 400 µg/ml G418 (Invitrogen). Selection was carried out for 12-14 days, replacing the selection media every other day. Colonies were washed with 1X PBS, fixed with 37% paraformaldehyde/8% glutaraldehyde, and stained with 0.1% crystal violet solution.

Protein expression and purification

A3A and A3A-mutant C106S cDNAs were inserted into pET28 and pET21 histidine-tag expression plasmids and expressed in *E. coli,* strain C43(DE3) (Lucigen). Recombinant protein was produced by expression at 37° C in LB plus antibiotics for 3 to 5 hrs after induction with 1mM IPTG. Bacterial cultures were spun down and cell pellets frozen prior to cell lysis. Cell pellets were resuspended in lysis buffer {50 mM NaCl, 200mM KCl, 10% glycerol, 1mM DTT, 35mM imidazole protease and Complete protease inhibitor (Roche)} and lysed using a fluidifier. The soluble (supernatant) fraction was separated by centrifugation at 40,000xg for 1 hr at 6° C. Recombinant protein from the supernatant was purified by binding to Ni-sepharose (GE healthcare), washing with lysis buffer and eluting with lysis buffer containing 0.5M imidazole. Protein was concentrated using Amicon Ultracell filters (Millipore) and further purified by gel filtration on a HiLoad 16/60 Superdex 200 prep grade column (Amersham), dialyzed in Slide-A-Lyzer cassettes (Pierce) in dialysis buffer (50 mM NaCl, 200mM KCl, 10% glycerol, 1mM DTT and, finally, concentrated with Amicon Ultracell filters (Millipore). Protein purification was monitored by SDS PAGE, coomassie blue staining, and/or immunoblotting with anti-His-tag (Sigma) or anti-APOBEC3A antibodies [49]. UV-spectrophotometry and Image J software was used to quantify protein levels.

In vitro Deaminase Assays

To determine rA3A deaminase activity in UDG-dependent deaminase assays, increasing concentrations of rA3A or rC106S were incubated with a 5'-end FITC labeled single strand deoxyoligonucleotide (0.4 mM) in a final reaction volume of 30 ml containing (40 mM Tris, pH 8.0, 10% glycerol, 40 mM KCl, 50 mM NaCl, 5 mM EDTA, and 1 mM DTT). The reactions were incubated at 37°C for 4-8 hrs, stopped by heating to 90°C for 5 min, cooled on ice, and then centrifuged at 10,000 x g for 1 minute. Twenty ml of the supernatant was then incubated with uracil DNA glycosylase (UDG, New England Biolabs) in buffer containing 20 mM Tris, pH 8.0, 1 mM DTT for 1 h at 37°C and treated with 150 mM NaOH for 30 minutes at 37°C. Then, samples were incubated at 95°C for 5 min, 4°C for 2 min and separated by 15% TBE/urea-PAGE. Gels were directly analyzed using a FLA-5100 scanner (Fuji). The PAGE-purified ssDNA oligonucleotide (Invitrogen) used for the deaminase assays (FITC-TCA) contains a single cytosine in the A3A specific target trinucleotide 5'-TCA.

(FITC-5'-TATTATTATTATTATTATTCATTTATTTATTTATTTATTT-3').

For deaminase assays on dsDNA (Fig. 2C) the target FITC-TCA oligonucleotide was pre-incubated with a complementary oligonucleotide, or a non-complementary oligonucleotide (ns) at the indicated ratios.

LEAP Assay

LEAP assays were carried out generally as described in [28].

Primer sequences: All oligos used in this study were syntesized by Integrated DNA Technologies (IDT; Coralville, Iowa)

LEAP adapter 5np1; purified by high-performance liquid chromatography (HPLC): 5'-GATGGATGATGAATAAAGTGT GGGATGAT**C**ATGATGTATGGATAGGTTTTTTTTTTTT-3'

LEAP adapter 5np0; purified by high-performance liquid chromatography (HPLC): 5'-GATGGATGATGAATAAAGTGT GGGATGAT**G**ATGATGTATGGATAGGTTTTTTTTTTTT-3'

LEAP adapter 5np1.U; purified by high-performance liquid chromatography (HPLC): 5'-GATGGATGATGAATAAAGTGT GGGATGAT**U**ATGATGTATGGATAGGTTTTTTTTTTTT-3'

5np1 outer: 5'-GATGGATGATGAATAAAGTG-3'

L1 3' end: 5'-GGGTTCGAAATCGATAAGCTTGGATCCAGAC-3' [28]

Preparation of L1 RNPs: $6x10^6$ HeLa cells were plated per T-175 flask (BD Biosciences). Eighteen hours later, they were transfected with 30 µg of L1 plasmid DNA using FuGene 6 transfection reagent (Roche) and Opti-MEM (Invitrogen) according to manufacturer's protocol (3 µl FuGene and 97 µl Opti-MEM per µg of DNA transfected). At 72 hours post-transfection, cells were subjected to selection with 200 µg/ml hygromycin B (Invitrogen). Hygromycin media was replaced daily for 9-11 days, until harvest day. Untransfected control HeLa cells were plated 2-3 days before harvest day. On harvest day, cells were washed three times with 10 ml cold 1x PBS and collected by scraping into 10 ml cold 1x PBS. Cells were pelleted in a 15-ml conical tube (BD Biosciences) at 3,000xg for 5 minutes at 4°C. The PBS was removed, and cells were lysed for 15 minutes on ice in 1 ml lysis buffer (1.5 mM KCl, 2.5 mM $MgCl_2$, 5 mM Tris-HCl (pH 7.4) 1% w/v deoxycholic acid, 1% w/v Triton X-100 and 1x Complete EDTA-free protease inhibitor cocktail (Roche). Following lysis, cell debris and nuclei were pelleted by centrifugation at 3,000xg for 5 minutes at 4°C, and whole-cell lysate was transferred to a new tube.

Isolation of RNPs by Ultracentrifugation: 1 ml whole-cell lysate was centrifuged through a sucrose cushion consisting of 8.5% and 17% w/v sucrose in 80 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-Cl (pH 7.5), 1 mM DTT, and 1x Complete EDTA-free protease inhibitor cocktail (Roche). Cells were ultracentrifuged for 2h at 178,000xg at 4°C. The resulting pellet was resuspended in 50-100 µl (depending on pellet volume) dH_2O with EDTA-free protease inhibitor cocktail (Roche).

Protein concentration was determined by Bradford assay (Bio-Rad; Hercules, California), and the sample was brought to 1.5 mg/ml final concentration. The portion of the sample intended for LEAP reactions was diluted to 50% v/v glycerol, aliquoted, flash-frozen on dry ice/ethanol, and stored at -80°C. Additional portions were reserved for RNA extraction.

RNA isolation and RT-PCR: RNA was extracted from RNP preps using the Qiagen RNA Easy columns following the manufacturer's instructions, omitting the cell lysis step and including on-column DNase treatment. RNA was quantified using a Nano-Drop spectrophotometer and diluted to 0.5 µg/µl. RT-PCR was carried out on 0.5 µl of purified RNA, using MMLV reverse transcriptase (Promega) and 0.8 µM LEAP adapter (5np1) at 42°C for 30 minutes. For A3A containing reactions, 100-300 ng of rA3A or rA3A_C106S protein in 10% glycerol were included in the MMLV RT reaction. For control reactions containing "heat-killed" rA3A, rA3A samples were incubated at 100° C for 15 minutes prior to inclusion in the MMLV RT reaction. Following the MMLV RT reaction, reaction mixtures were incubated at 100°C for 15 minutes to denature rA3A protein. O.5 µl MMLV-RT reaction product was PCR amplified using *PfuTurbo* C_x Hotstart DNA polymerase (Agilent) and the following conditions: 95° C for 2 minutes, followed by 35 cycles of 30s at 95° C, 30s at 58° C, and 30s at 72° C, with a final extension time of 7 minutes at 72°C. PCR products were visualized on 2% agarose gels.

LEAP Reactions: One µl of 0.75 µg/ul (50% v/v glycerol) RNP sample was incubated with 50 mM Tris-Cl (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, 10 mM DTT, 0.4 µM LEAP primer, 20 U RNasin (Promega), 0.2 mM dNTPs (Invitrogen) and 0.05% v/v Tween 20 in a final volume of 50 µl, for 30 minutes at 37°C. A "no RNP" reaction was prepared as a negative control. For A3A containing reactions, 5-100 ng of rA3A or rA3A_C106S protein in 10% glycerol were included in the LEAP reaction. For RnaseH-containing reactions, 2 units recombinant RNaseH (Invitrogen) were included in the LEAP reaction. For control reactions containing "heat-killed" rA3A, rA3A samples were incubated at 100° C for 15 minutes prior to inclusion in the LEAP reaction.

Following the L1 RT reaction, reactions were incubated at 100°C for 15 minutes to denature rA3A protein. 1 µl of LEAP product was then included in the PCR amplification step using *PfuTurbo* Hotstart DNA polymerase (Agilent) or *PfuTurbo* C_x Hotstart DNA polymerase (Agilent), and the following conditions: 95° C for 2 minutes, followed by 35 cycles of 30s at 95° C, 30s at 58° C, and 30s at 72° C, with a final extension time of 7 minutes at 72°C. PCR products were visualized on 2% agarose gels.

Product characterization: PCR products were excised from agarose gels and purified using the QIAquick gel extraction kit (Qiagen). Products were cloned into ZERO Blunt PCR cloning vector (Invitrogen), transformed, and plasmid DNA recovered by mini-prep (Promega SV Mini-Prep kit; Promega, Fitchburg, Wisconsin). Individual clones were then sequenced. Sequence alignments were

produced using MegAlign software from the Lasergene DNASTAR suite (DNASTAR; Madison, Wisconsin).

Characterization of retrotransposition events

Primers: All oligos used in this study were syntesized by Integrated DNA Technologies (IDT; Coralville, Iowa)

NEOasReco: 5'-TCTATCGCCTTCTTGACGAG-3'

Rescue3seq: 5'-ACTCACGTTAAGGGATTTTGGTCA-3'

PolyAseq: 5'-AAAAAAAAAAAAAAAAAAABN-3'

Generation of clonal cell lines: $2x10^5$ HeLa cells were plated per 15 cm dish (BD Biosciences) and transfected using FuGene 6 (Roche) and Opti-MEM (Invitrogen) with 3 µg each of A3A expression vector, pLGCX/UGI, and JM140/L1.3/Δ^2/k7 according to manufacturer's protocol (3 µl FuGene and 97 µl Opti-MEM per µg of DNA transfected). For U2OS_UGI cells, $5x10^5$ cells were plated per 15 cm dish and transfected using FuGene 6 (Roche) and Opti-MEM (Invitrogen) with 2 µg A3A (or A3A_C106S, or β-arr) plasmid and 8 µg of JM140/L1.3/Δ^2/k7 according to manufacturer's protocol (3 µl FuGene and 97 µl Opti-MEM per µg of DNA transfected). Selection with 400 µg/ml of G418 (Invitrogen) was initiated at 72 hours post-transfection, and carried out for 14 days. G418 media was replaced every other day. On day 14, individual colonies were manually picked to individual wells of 12-well culture dishes (BD

Biosciences). These colonies were expanded for 2-3 weeks to establish clonal cell lines. Confluent T-75 flasks (BD Biosciences) of each cell line were harvested by trypisinization, and genomic DNA was prepared using the Qiagen Blood and Cell Culture Midi Kit, according to manufacturer's instructions.

Recovery of L1 insertions: Insertions were recovered generally as previously described [2]. Briefly, 8 µg of genomic DNA was digested with excess restriction enzyme (*Hind*III, *Ssp*I, *Bgl*II, *Bam*HI: NEB; *Bcl*I: Promega) overnight at 37° C. The following morning, an additional 2 µl of enzyme was added to the digest, and incubated at 37° C for 2 hours. Digestion reactions were heat-inactivated (65° C, 15 minutes), or in the case of *Bgl*II digest, cleaned up with the Wizard DNA clean-up kit (Promega). The entire digest was then ligated with 8 µl T4 DNA ligase (NEB) under dilute conditions (500 µl total) overnight at 16° C. The next morning, 2 µl additional T4 DNA ligase was added to the reaction and incubated at room temperature for 4 hours. Ligations were concentrated on an Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-100 membrane (Millipore; Billerica, Massachussets) at 8,000xg for 5 minutes, followed by a wash with 500µl dH_2O, 5 minutes at 8,000xg. 10-100 µl of concentrated DNA was recovered with a 10 second spin. The entire ligation was then added to 500 µl of XL-10 gold ultra-competent *E. coli* (Stratagene; prepared in-house as described in [40]) and incubated on ice for 1-3 hours. Transformations were heat-shocked at 42° C for 38-45 seconds, and allowed to recover on ice for 2 minutes. 2 ml of warm (37° C) LB (no antibiotic) was added to each transformation. Transformations were

incubated overnight at room-temperature on an orbital shaker at 100 rpm. The next morning, transformations were pelleted (300xg) and gently resuspended in 500 µl fresh LB. About 400 µl of this resuspension was plated on a 15 cm Kanamycin plate (30 µg/ml), and the remaining ~100 µl was used to seed a 2 ml liquid culture (30 µg/ml Kanamycin). Plates and liquid cultures were incubated for 18-24 hours at 37° C. Mini-preps (Promega SV mini-prep kit) were prepared from 2 ml cultures. From 15 cm plates, individual colonies were picked to inoculate 2 ml cultures (30 µg/ml Kanamycin), grown overnight at 37° C, and plasmid DNA was prepared by mini-prep the following day. Plasmid DNA recovered from mini-preps was digested with the original enzyme used for recovery, to confirm intramolecular ligation. To characterize flanking genomic DNA, recovered insertions were sequenced using primers annealing to the 5' (NEOasReco) and 3' (Rescue3seq) ends of the NEO_ColE1 recovery cassette, as well as an oligo dA primer (polyAseq). The genomic locations of insertions were determined by aligning flanking sequence to the human genome (Feb. 2009; GRCh37/hg19), using the BLAT function of the UCSC genome browser.

Acknowledgements

I would like to thank the Moran lab for helpful discussion of the data presented in this chapter. Inigo Narvaiza and Matt Weitzman generated and characterized purified recombinant A3A protein. Randy Planegger helped design and carried out the retrotransposition assays with A3B deaminase-deficient mutants (Figure 3.14).

Figure 3.1: A3A-mediated L1 inhibition is sequence-independent

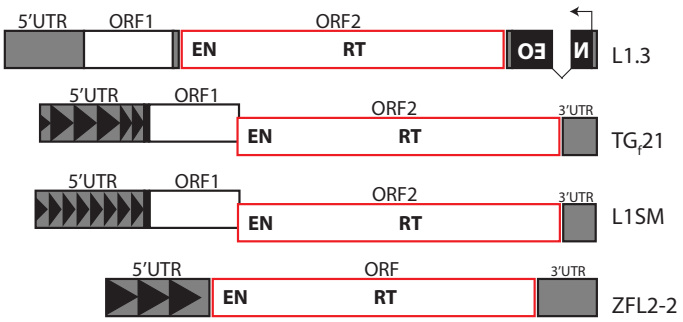
A. Schematic of the LINE elements used in this experiment: The *mneoI* cassette [18, 50] is shown within the 3'UTR of human L1.3. Black triangles represent the monomers in the 5'UTR of the mouse L1s (TG_f21 and L1SM), and repetitive units in the 5'UTR of the zebrafish LINE-2 element (Zfl2-2). L1.3, TG_f21, and L1SM each encode ORF1 and ORF2 (outlined in red). Zfl2-2 contains only one ORF (outlined in red) which encodes endonuclease (EN) and reverse transcriptase (RT) activities.

B. Percent nucleotide identity between ORF1 or ORF2 of human L1.3, and the corresponding ORF of each of the elements used. For Zfl2-2, n/a indicates not applicable, as this element lacks an ORF1 orthologue.

C. Experimental strategy. Approximately $2x10^4$ (L1SM), $1x10^5$ (L1.3), $2x10^5$ (NEO control), or $4x10^5$ (TGf21 and Zfl2-2) HeLa cells were plated per T-75 flask and co-transfected in duplicate with 2 µg each of LINE and A3A (or β-arrestin control) expression vector.

D. Sequence-independent inhibition of retrotransposition by A3A. The impact of wild-type A3A (white bars), A3A_C106S (grey bars), and β-arr positive control (black bars) on LINE retrotransposition is shown. The y-axis indicates percent retrotransposition, with the β-arr positive control set to 100% for each element. The x-axis indicates the LINE element used. Error bars indicate percent standard deviation derived from duplicate transfections. The same experiment is shown in Figure 4.2, with the addition of A3A_F75L. This experiment was performed twice, representing biological replicates.

A.

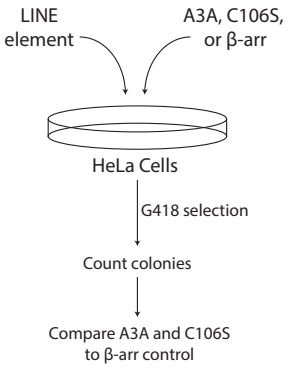


B.

% Nucleotide Identity with L1.3

	ORF1	ORF2
TGf21	28.2	59.3
L1SM	32.6	51.7
Zfl2-2	n/a	25.8

C.



D.

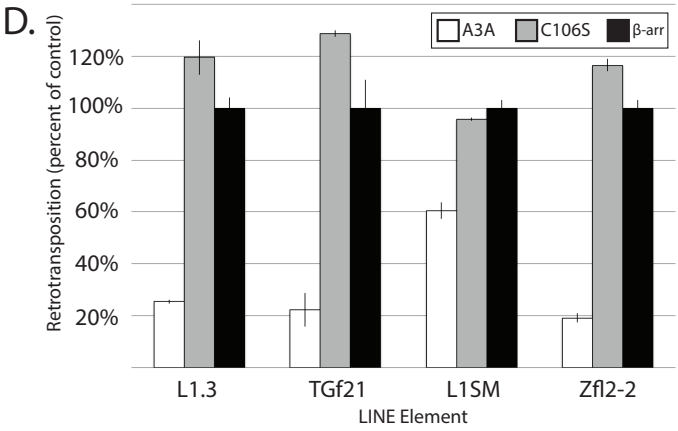
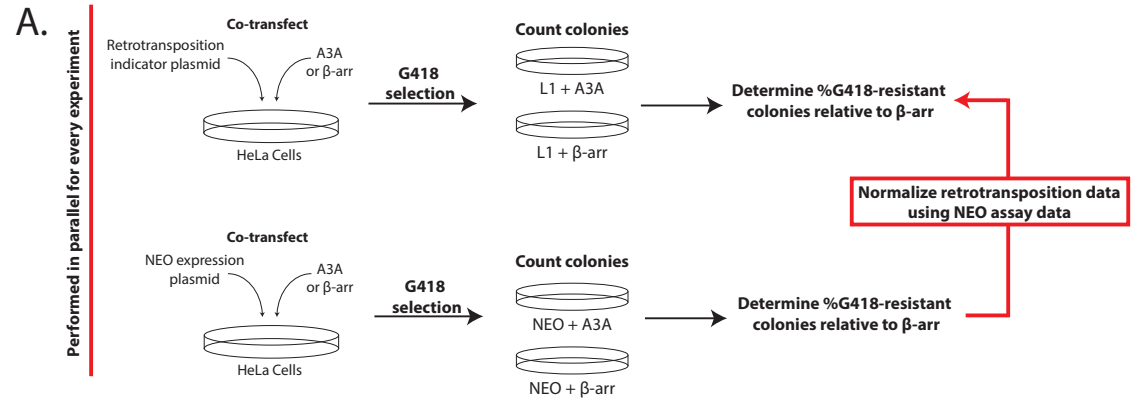


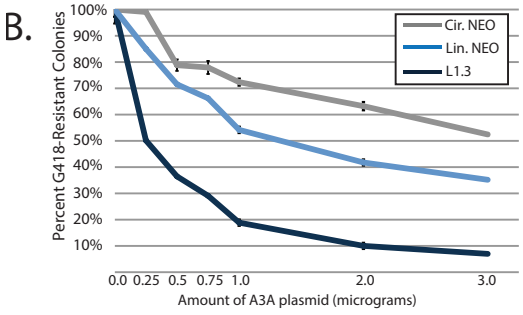
Figure 3.2: A3A expression has cytotoxic effects that do not account for the full impact of A3A expression on L1 retrotransposition

A. Schematic for A3A toxicity controls. For each assay in this thesis, co-transfections with a L1 retrotransposition indicator and A3A/β-arr plasmid, and a NEO expression vector and A3A/β-arr plasmid, were carried out in parallel. At 72 hours post-transfection, G418 selection (400μg/ml) was initiated and carried out for 14 days. Colonies were fixed, stained, and counted. The percentage of G418-resistant colonies relative to β-arr control was determined for L1 and NEO experiments. The results from each NEO co-transfection were used to normalize data from the corresponding L1 co-transfection.

B. A3A expression diminishes the number of G418-resistant colonies generated by co-transfection of a circular or linearized NEO expression vector, and by L1 retrotransposition. Approximately $2x10^5$ (retrotransposition assay) or $4x10^5$ (NEO assay) HeLa cells per T-75 flask were transfected in duplicate with 4 μg total plasmid DNA. Each transfection included 1 μg of L1 or NEO expression vector, and increasing amounts of A3A expression vector (from 0.0 μg to 3.0 μg). The remaining plasmid mass was made up with β-arr expression vector so that each co-transfection consisted of a total plasmid mass of 4 μg. The y-axis indicates the percentage of G418-resistant colonies obtained from each transfection, with 0.0 μg A3A (4.0 μg β-arr) set to 100%. The x-axis indicates the amount of A3A plasmid present in the co-transfection. The gray line indicates colony counts for circular NEO co-transfection experiments; the light blue line indicates colony counts for linear NEO co-transfection experiments, and the dark blue line indicates counts for L1.3 retrotransposition. Error bars indicate percent standard deviation between duplicate transfections.

C. Dose-dependent A3A-mediated L1 retrotransposition inhibition. The colony counts generated in the retrotransposition assay (Supplemental Figure 1A) were normalized using the colony counts generated in the linear or circular NEO control co-transfections. The y-axis indicates normalized percent retrotransposition, with 0.0 μg A3A (4.0 μg β-arr) set to 100%. The x-axis indicates the amount of A3A plasmid in the co-transfection. Dark blue bars show the results of normalization of L1 data with circular NEO co-transfection data; gray bars show the results of normalization with linear NEO co-transfection data. Error bars indicate percent standard deviation between duplicate transfections.
This experiment was performed twice, representing biological replicates.


A.
Performed in parallel for every experiment
Co-transfect
Retrotransposition indicator plasmid
A3A or β-arr
HeLa Cells
G418 selection
Count colonies
L1 + A3A
L1 + β-arr
Determine %G418-resistant colonies relative to β-arr
Normalize retrotransposition data using NEO assay data
Co-transfect
NEO expression plasmid
A3A or β-arr
HeLa Cells
G418 selection
Count colonies
NEO + A3A
NEO + β-arr
Determine %G418-resistant colonies relative to β-arr


B.
Percent G418-Resistant Colonies
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
Cir. NEO
Lin. NEO
L1.3
0.0 0.25 0.5 0.75 1.0 2.0 3.0
Amount of A3A plasmid (micrograms)

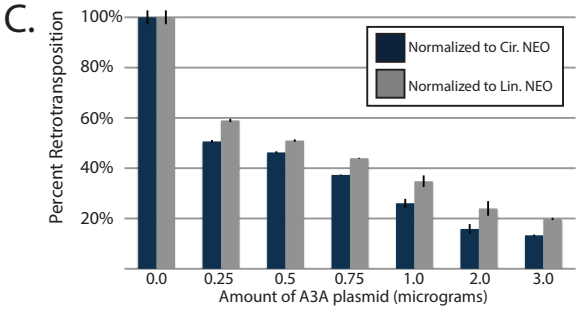

C.
Percent Retrotransposition
100%
80%
60%
40%
20%
Normalized to Cir. NEO
Normalized to Lin. NEO
0.0 0.25 0.5 0.75 1.0 2.0 3.0
Amount of A3A plasmid (micrograms)

Figure 3.3: Inhibition of L1 by A3A is not attributable to interference with L1 expression from a viral (CMV) promoter, or by interference with the replication of an episomal L1 expression vector

Approximately $2x10^3$ HeLa cells were plated per well of a 6-well dish and co-transfected in triplicate with 0.5 µg A3A expression vector and 0.5 µg of L1 expression vector per well. L1 expression vectors: pDK101 L1.3, a pCEP4-based episomal construct harboring a CMV promoter in addition to the L1 5'UTR [43]. pBSKS L1.3 is a non-episomal vector from which L1 expression is driven solely by the L1 5'UTR. pJM102 L1.3 is a pCEP4-based episomal construct harboring a CMV promoter. pJM102 L1.3 does not contain the L1 5'UTR; L1 expression is driven only by a CMV promoter. The y-axis indicates linear NEO-normalized percent retrotransposition, with the β-arr positive control set to 100%. The L1 expression construct is indicated on the x-axis. Dark blue bars show % retrotransposition in the presence of A3A; gray bars indicate the positive control β-arr. Error bars indicate percent standard deviation derived from triplicate transfections. Data were normalized to linear NEO control co-transfections as shown in figure 3.2.

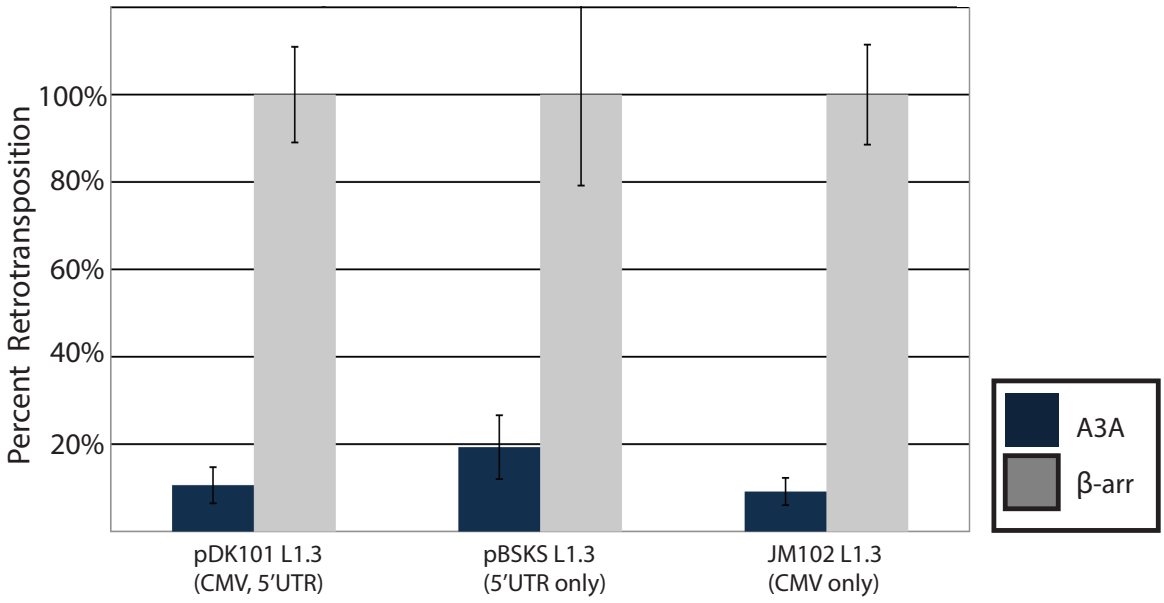

Percent Retrotransposition
100%
80%
60%
40%
20%
pDK101 L1.3
(CMV, 5'UTR)
pBSKS L1.3
(5'UTR only)
JM102 L1.3
(CMV only)
A3A
β-arr

Figure 3.4: A3A-mediated L1 inhibition does not occur by blockage of L1 endonuclease activity

A) Experimental rationale. if A3A specifically blocks L1 EN activity, ENi retrotransposition events are expected to escape inhibition. If A3A does not specifically block L1 EN activity, ENi events are expected to be susceptible to inhibition.

B) Control retrotransposition assay in 4364a (wild-type) CHO cells. Approximately $1x10^5$ 4364a CHO cells were plated in a T-75 flask and co-transfected with 2 µg of wild-type L1 retrotransposition indicator (pAD2TE1) and 2 µg of A3A or control (β-arr) expression vector. The y-axis indicates percent retrotransposition, with the β-arr positive control set to 100%.

C) Wild-type and ENi retrotransposition assay in XR-1 CHO cells. Approximately $1x10^5$ XR-1 CHO cells were plated in a T-75 flask and co-transfected with 2 µg of wild-type (pAD2TE1, gray bars) or endonuclease mutant (pAD136, black bars) L1 retrotransposition indicator plasmid and 2 µg of A3A or control (β-arr) expression vector. The y-axis indicates percent retrotransposition, with the β-arr positive control set to 100%. Error bars indicate standard deviation between duplicate transfections.

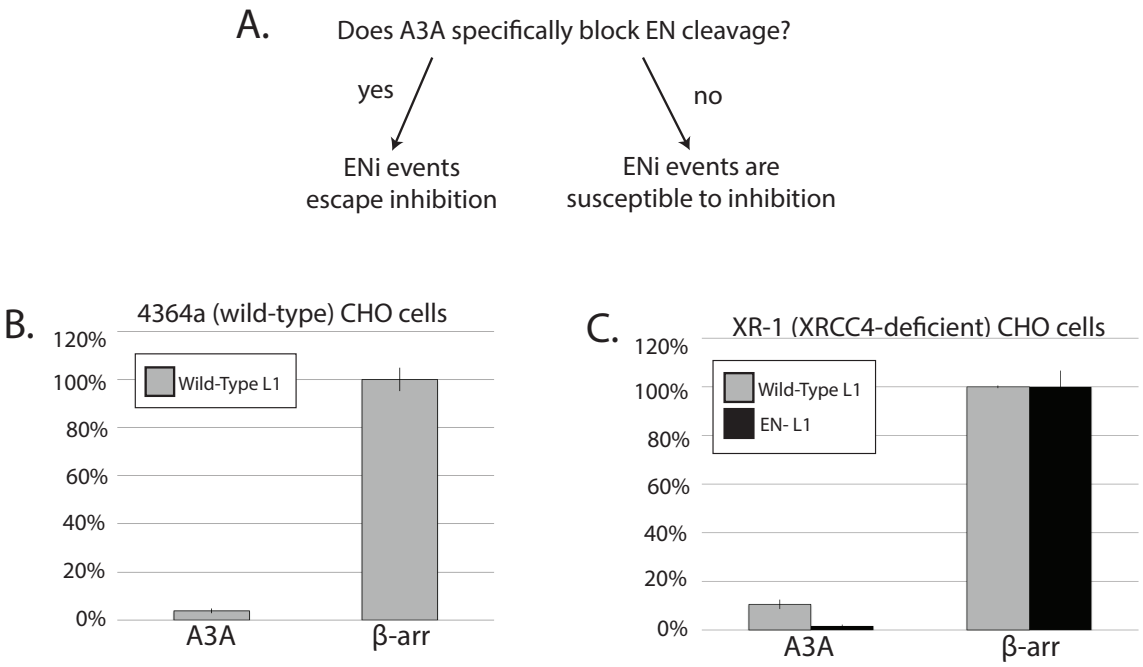

A.
Does A3A specifically block EN cleavage?
yes
no
ENi events
escape inhibition
ENi events are
susceptible to inhibition
B.
4364a (wild-type) CHO cells
120%
100%
80%
60%
40%
20%
0%
Wild-Type L1
A3A
β-arr
C.
XR-1 (XRCC4-deficient) CHO cells
120%
100%
80%
60%
40%
20%
0%
Wild-Type L1
EN- L1
A3A
β-arr

Figure 3.5: *Pfu* cannot process through uracils in the template strand during PCR amplification

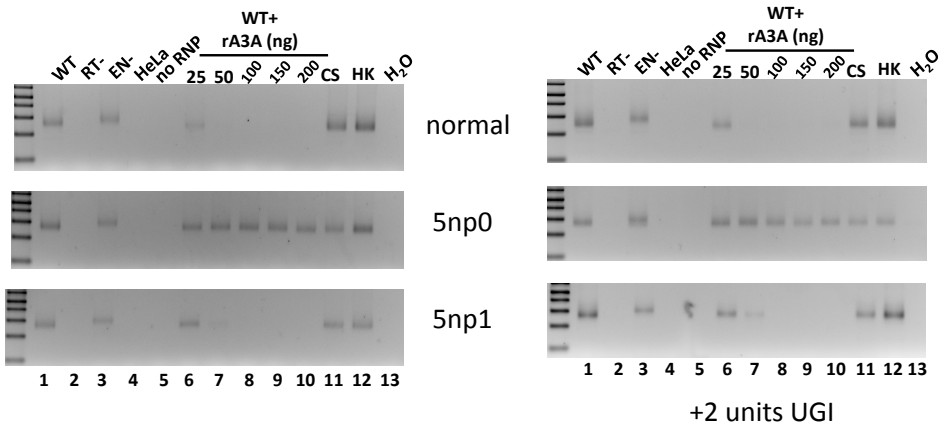
For each panel, lanes are labeled across the top. For LEAP reactions (A and B), RNPs of the indicated L1 construct were used; for MMLV RT reactions (C), RNA purified from the indicated type of RNP was used. WT indicates pDK101 [43], a wild-type L1, RT- indicates pDK135 [28], an L1 bearing a mutation in the reverse transcriptase domain (D702A), EN- indicates JJ10.3-H230A [45], an endonuclease-deficient L1 bearing the mutation H230A in the endonuclease active site. This construct also contains a blasticidin-resistance indicator cassette instead of *mneoI;* a small difference in cassette length causes LEAP products from this construct to be slightly larger, giving rise to a higher molecular weight LEAP product. HeLa indicates untransfected HeLa cells; CS indicates 100 ng of rA3A_C106S, a deaminase deficient A3A mutant protein. HK indicates 100 ng of heat-killed (15 minutes at 100°C) rA3A protein. H2O is a water control (no RT product) for the PCR amplification step.

A. A3A-dependent loss of LEAP products occurs only when there is a cytidine residue within the LEAP adapter. Increasing nanogram amounts of recombinant A3A protein (rA3A) were added to wild-type L1 RNP reactions. On the right, 2 units of recombinant uracil glycosylase inhibitor protein (UGI) were included in each reaction. For each set of experiments, the LEAP adapter used is indicated. Top: "Normal" indicates the original LEAP/3'RACE adapter [28], which contains seven cytidines. Middle: 5np0 is a modified LEAP adapter containing no cytidine residues. Bottom: 5np1 is identical to 5np0, except that it contains a single cytidine residue.

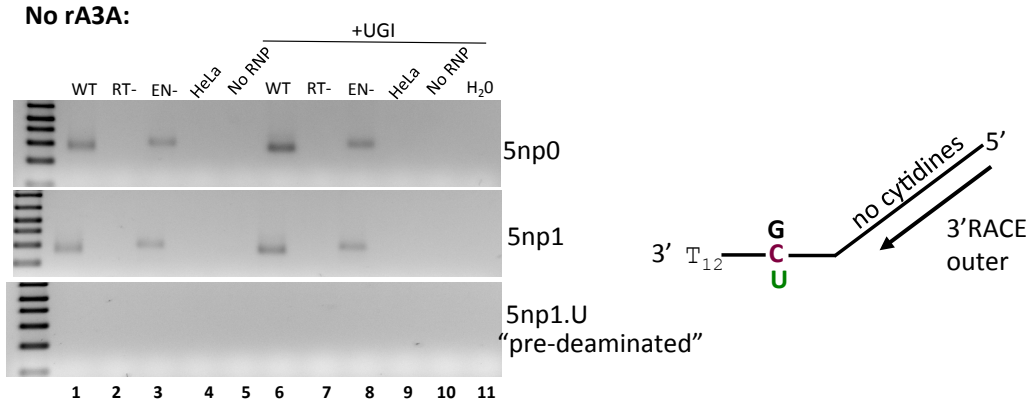
B. LEAP products are lost when the LEAP adapter contains a uracil residue. No rA3A protein was added to the reactions in this panel. On the right, the 5np0 ("G"), 5np1 ("C"), and 5np1.U ("U") LEAP adapters are diagrammed. On the left, LEAP reactions were performed with 5np0 (top), 5np1 (middle), and the "pre-deaminated" LEAP adapter 5np1.U (bottom). Reactions were performed both in the absence (lanes 1-5) and presence (lanes 6-10) of 2 units of recombinant UGI.

C. MMLV RT products are lost when the LEAP adapter contains a uracil residue. No rA3A protein was added to the reactions in this panel. Above, lanes 1-6 contain MMLV RT products generated from the original ("normal") LEAP/3'RACE adapter. Lanes 7-12 were generated from the 5np0 adapter. Below, lanes 1-6 were generated from the 5np1 adapter, and lanes 7-12 were generated from the 5np1.U adapter.

A.


WT+ rA3A (ng)
WT RT- EN- HeLa no RNP 25 50 100 150 200 CS HK H2O
normal
5np0
5np1
1 2 3 4 5 6 7 8 9 10 11 12 13
WT+ rA3A (ng)
WT RT- EN- HeLa no RNP 25 50 100 150 200 CS HK H2O
1 2 3 4 5 6 7 8 9 10 11 12 13
+2 units UGI

B.


No rA3A:
+UGI
WT RT- EN- HeLa No RNP WT RT- EN- HeLa No RNP H20
5np0
5np1
5np1.U
"pre-deaminated"
1 2 3 4 5 6 7 8 9 10 11
3' T12
G
C
U
no cytidines
5'
3'RACE outer

C.

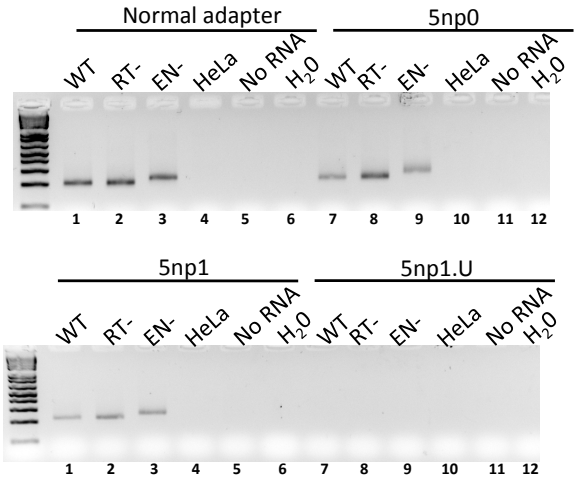
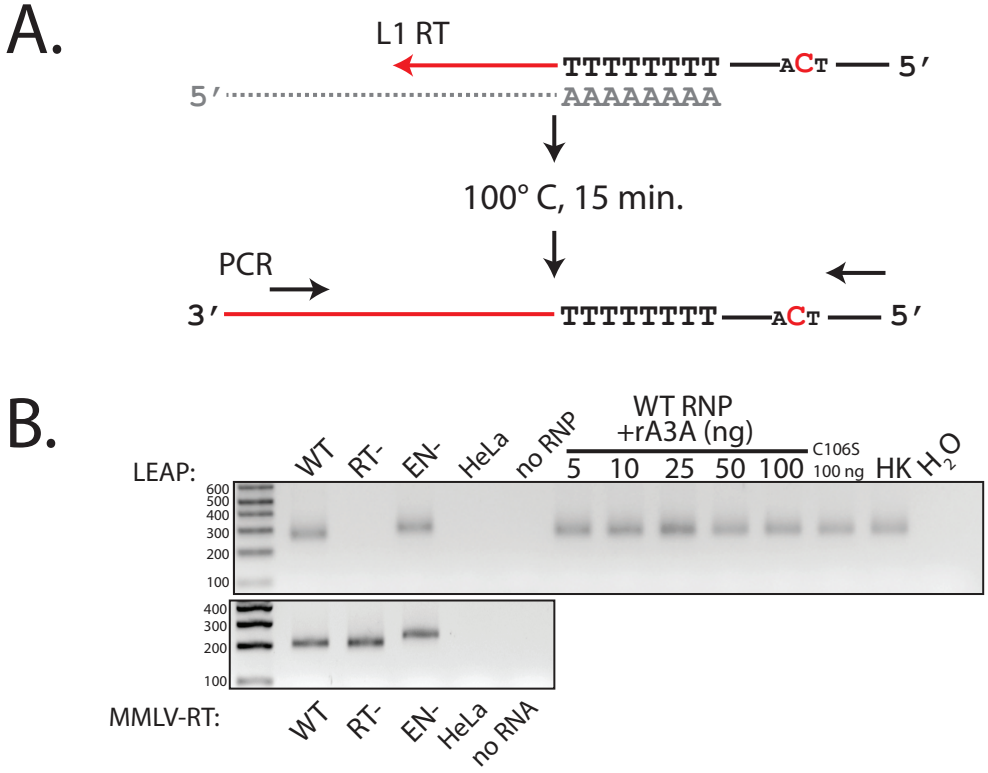

Normal adapter
5np0
WT RT- EN- HeLa No RNA H2O WT RT- EN- HeLa No RNA H2O
1 2 3 4 5 6 7 8 9 10 11 12
5np1
5np1.U
WT RT- EN- HeLa No RNA H2O WT RT- EN- HeLa No RNA H2O
1 2 3 4 5 6 7 8 9 10 11 12

Figure 3.6: rA3A does not block L1 RT activity

A) The LEAP assay. L1 RNPs, consisting of the L1 RNA (dashed grey line) and the L1-encoded proteins (not shown) are isolated by ultracentrifugation from transfected cells. The L1 RNPs are provided with a 3' RACE primer (5np1, shown in black) consisting of a 3' oligo dT and a 5' unique adapter sequence. The 5np1 adapter contains a single cytidine (red) in the unique adapter portion. Following the RT step, reactions are boiled at 100° C for 15 minutes to denature rA3A protein. L1 cDNAs were amplified by PCR using primers to the engineered L1 3' end and the unique adapter sequence (black arrows).

B) rA3A does not inhibit the LEAP reaction. Control reactions, left to right: wild-type (pDK101) RNP, RT mutant (pDK135 [28]; bearing RT domain mutation D702A) RNP, EN mutant (JJ101.3-H230A [45]) RNP; this construct also contains a blasticidin-resistance indicator cassette instead of *mneoI;* a small difference in cassette length causes LEAP products from this construct to be slightly larger, giving rise to a higher molecular weight LEAP product. HeLa indicates untransfected HeLa cell RNP prep, no RNP control. rA3A-containing reactions: wild-type (pDK101) RNP with increasing nanogram amounts (5-100 ng) of rA3A protein. CS: 100 ng of rA3A_C106S. HK: 100 ng of "heat-killed" (100° C for 15 minutes) rA3A protein. H_2O: no L1 RT product PCR negative control. Below, MMLV RT reactions illustrate the integrity of purified RNA isolated from the RNP prep used in the LEAP assay. MMLV RT reactions were carried out using the 5np1 3'RACE adapter, and PCR amplified using the 5np1 outer primer and the L1 3' end primer. All PCR reactions were carried out using *Pfu* Turbo C_x.


A.
L1 RT
TTTTTTTT
ACT
5′
5′
AAAAAAAA
100° C, 15 min.
PCR
3′
TTTTTTTT
ACT
5′
B.
LEAP:
WT
RT-
EN-
HeLa
no RNP
WT RNP
+rA3A (ng)
5
10
25
50
100
C106S
100 ng
HK
H2O
600
500
400
300
200
100
400
300
200
100
MMLV-RT:
WT
RT-
EN-
HeLa
no RNA

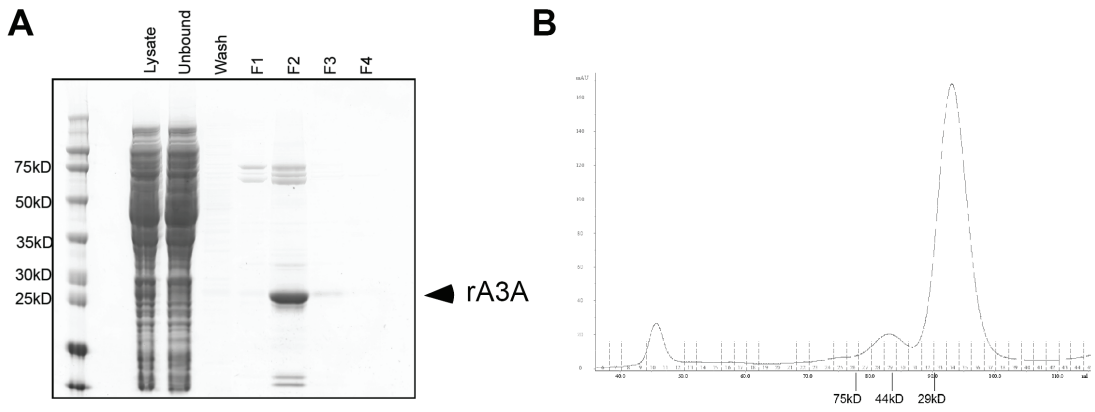


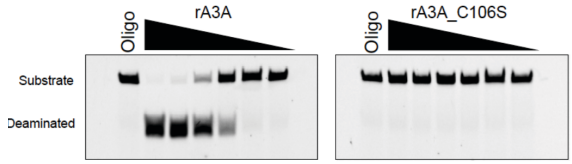
Figure 3.7: Purification of recombinant A3A protein

A) Purification of rA3A from E. coli by Ni affinity. Purification of His-tagged rA3A protein was monitored by SDS-4-12% polyacrylamide gel electrophoresis and coomassie blue staining. Input bacterial lysate was loaded onto the Ni-Sepharose column (Lysate). Flowthrough lysate and wash fractions were collected (Unbound and Wash, respectively). rA3A was eluted with lysis buffer containing 0.5M imidazole and consecutive elution fractions were collected (F1-4). Eluted His-tagged rA3A was mostly eluted in F2 (arrow). Approximate molecular sizes are indicated on the left.

B) Purification of rA3A by gel filtration. rA3A purified by Ni-Affinity was further purified by gel filtration by FPLC on a Superdex 200 column. 280nm absorbance profile of the FPLC fractions is shown. Elution volume and fractions are indicated in the X-axis. Approximate molecular weight is indicated.

This figure was generated by Dr. Inigo Narvaiza.

A.



B.

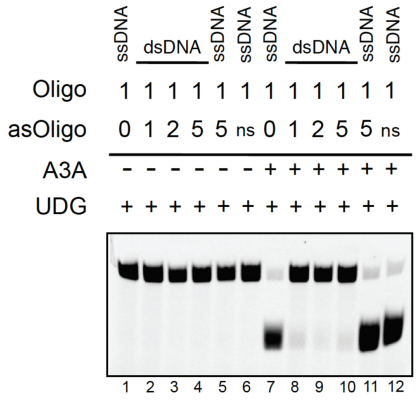


Figure 3.8: Characterization of rA3A deaminase activity *in vitro*

A. rA3A but not the C106S mutant has deaminase activity in vitro.
Increasing amounts of rA3A (left panel) or rC106S (right panel) were incubated with a FITC-labeled single strand DNA oligonucleotide containing a single cytidine residue for 4 hrs at 37° C. Samples were then incubated with recombinant uracil DNA glycosylase (UNG) and NaOH, and the products were resolved by gel electrophoresis using a 15% TBE/urea polyacrylamide gel. Molecular sizes of substrate oligonucleotide and deaminated, cleaved product are indicated. A control reaction (Oligo) was performed in the absence of recombinant protein.

B. Recombinant A3A does not deaminate dsDNA. UDG dependent Deaminase assays with rA3A in the presence of increasing concentrations of a oligonucleotide complementary (asOligo) to the FITC labeled deamination substrate (1:1, 1:2, 1:5 ratio substrate:asoligo). In control reactions the antisense oligonucleotide was absent (lanes 1 and 7), antisense oligonucleotide was added after incubation with A3A (lanes 5 and 11), or a non specific oligonucleotide (ns) was added (lanes 6 and 12). A band corresponding showing a product of deamination was only detected after UDG treatment when the substrate is accessible to rA3A as ssDNA: in the absence of asOligo (lanes 7 and 12) or when the antisense oligonucleotide was added after incubation of the substrate with rA3A (lane 11).

This figure was generated by Dr. Inigo Narvaiza.

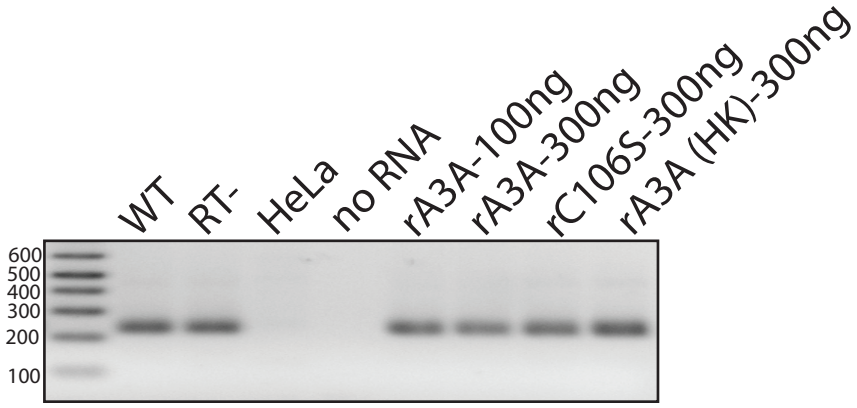


Figure 3.9: Recombinant A3A does not inhibit MMLV-RT activity

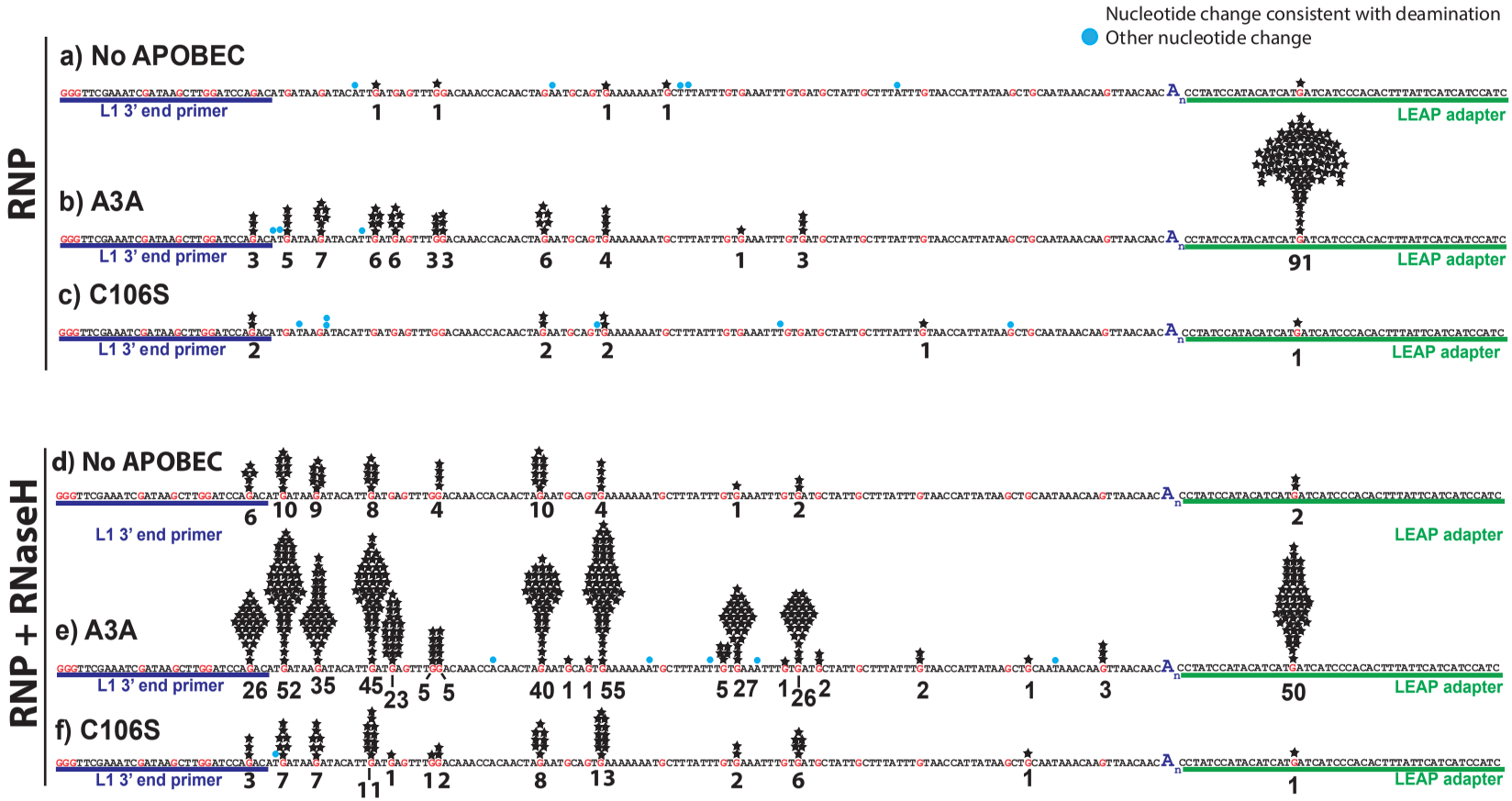
From right to left: MMLV RT reactions were carried out using the 5np1 3'RACE adapter, on purified RNA isolated from RNPs harvested from cells transfected with pDK101 (WT) or pDK135 (RT-). RNA isolated from untransfected HeLa cells (HeLa) and a no RNA sample were included as controls. Recombinant A3A (100 ng and 300 ng), rA3A_C106S (300 ng) and "heat-killed" (15 minutes, 100° C) rA3A (300 ng) were included in MMLV RT reactions. Products were amplified using the 5np1 outer primer and the L1 3' end primer.

Figure 3.10: Summary of 100 LEAP products per experimental condition

The plus-strand sequence of the unedited LEAP product is depicted, 5' to 3; therefore, G-to-A changes correspond to C-to-U deaminations on the minus-strand L1 cDNA. The location of the L1 3' end PCR primer is underlined in blue. A_n indicates the LEAP product polyA tail, followed by the 5np1 LEAP adapter sequence, underlined in green. Within the LEAP product sequence, all G nucleotides are highlighted in red. Black stars piled above G nucleotides graphically represent the number times among 100 LEAP products that a nucleotide change consistent with deamination was observed at that position. Counts of deamination events are represented numerically below. Blue circles piled above the sequence represent nucleotide changes not consistent with deamination.

Top panel, a-c: Summary of 100 LEAP products generated using wild-type L1 RNP plus: a) no rA3A protein, b) 100 ng of wild-type rA3A, c) 100 ng of rA3A_C106S.

Bottom panel, d-f: Summary of 100 LEAP products generated using wild-type L1 RNP with 2 Units of RNaseH, plus: d) no rA3A protein, e) 100 ng of wild-type rA3A, f) 100 ng of rA3A_C106S.


Nucleotide change consistent with deamination
Other nucleotide change
RNP
a) No APOBEC
b) A3A
c) C106S
L1 3' end primer
LEAP adapter
RNP + RNaseH
d) No APOBEC
e) A3A
f) C106S
L1 3' end primer
LEAP adapter

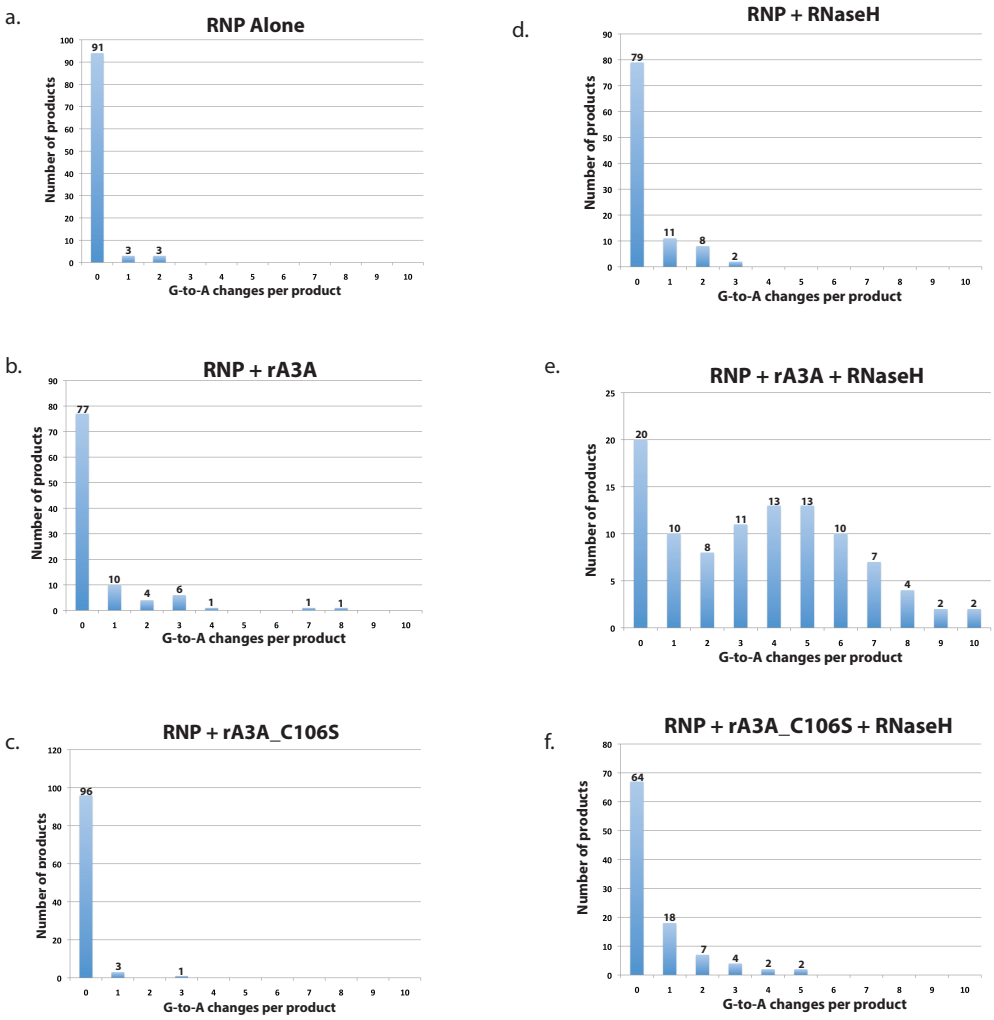


Figure 3.11: Numerical distribution of deamination events per L1 cDNA portion of the LEAP product, excluding the LEAP adapter

For each experimental condition depicted in Figure 5, a histogram quantifies G-to-A changes per product (x-axis) versus the number of LEAP products (out of 100 per condition) harboring each quantity of G-to-A changes (y-axis). The L1 cDNA portion of the LEAP product, but not the LEAP adapter, is considered in this analysis. The abundance of LEAP products harboring multiple deaminations suggests that A3A may act processively on single-stranded DNA.

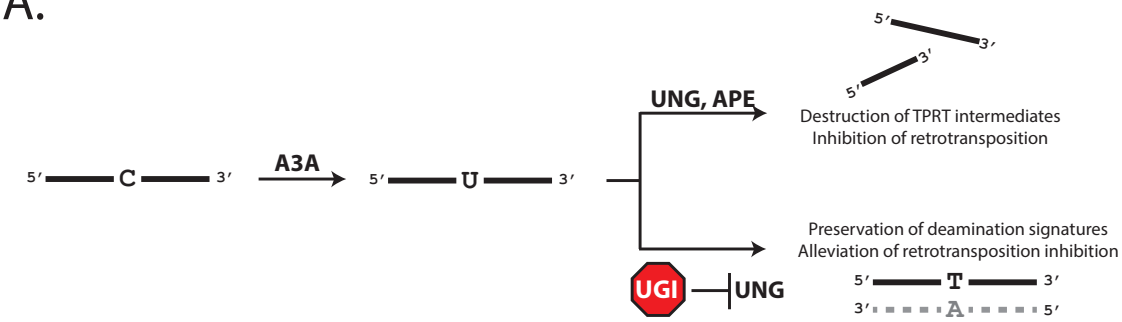
Figure 3.12: UGI expression alleviates A3A-mediated inhibition of retrotransposition in HeLa cells

A) Experimental rationale. From left: A3A is predicted to deaminate cytidine to uracil in single-stranded DNA in TPRT intermediates. Upper branch: uracils in DNA are recognized and removed by UNG, resulting in an abasic site that can be cleaved by APE, which is predicted to result in destruction of TPRT intermediates, thereby inhibiting retrotransposition. Lower branch: UGI expression blocks the activity of UNG, preserving uracils in DNA. Thus, deaminated TPRT intermediates are predicted to persist, and harbor signatures of deamination which become fixed as C-to-T changes (G-to-A on the opposite DNA strand), and A3A-mediated inhibition of retrotransposition should thereby be alleviated.

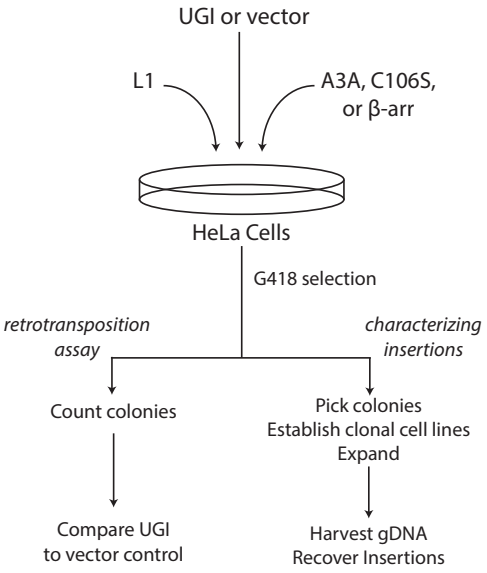
B) Schematic for testing the effect of UGI expression on A3A-mediated L1 inhibition, and generation of stable HeLa cell lines harboring retrotransposition events, using A3A and UGI expression vectors. Left branch, retrotransposition assay: Approximately $1x10^5$ HeLa cells were plated per 10 cm dish and co-transfected with 1 µg each of L1 expression plasmid (JM101 L1.3) A3A or β-arr control vector, and pLGCX/UGI or pLGCX empty vector. G418 selection (400 µg/ml) was initiated at 72h post-transfection and carried out for 14 days. Right branch, characterizing insertions: $2x10^5$ HeLa cells were plated per 15 cm dish, and transfected 24 hours later with 3 µg each of JM140/L1.3/Δ^2/k7, A3A, and LGCX/UGI plasmids. G418 selection (400 µg/ml) was initiated at 72h post-transfection and carried out for 14 days. Individual colonies were then isolated and expanded to establish clonal cell lines. Genomic DNA was prepared and subjected to the recovery procedure.

C) UGI expression alleviates A3A-mediated retrotransposition inhibition. The y-axis depicts percentage of L1 retrotransposition relative to control. For each experimental condition, the appropriate β-arr control was set to 100%. Gray bars indicate empty pLGCX vector; black bars indicate pLGCX_UGI. Error bars represent standard deviation between duplicate transfections.

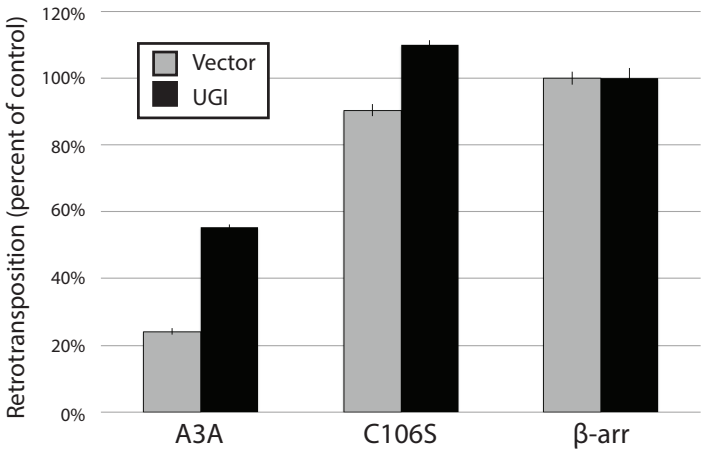
A.


5′ C 3′
A3A
5′ U 3′
UNG, APE
5′
3′
3′
5′
Destruction of TPRT intermediates
Inhibition of retrotransposition
Preservation of deamination signatures
Alleviation of retrotransposition inhibition
UGI
UNG
5′ T 3′
3′ A 5′

B.


UGI or vector
L1
A3A, C106S,
or β-arr
HeLa Cells
G418 selection
retrotransposition
assay
characterizing
insertions
Count colonies
Pick colonies
Establish clonal cell lines
Expand
Compare UGI
to vector control
Harvest gDNA
Recover Insertions

C.


Retrotransposition (percent of control)
120%
100%
80%
60%
40%
20%
0%
Vector
UGI
A3A
C106S
β-arr

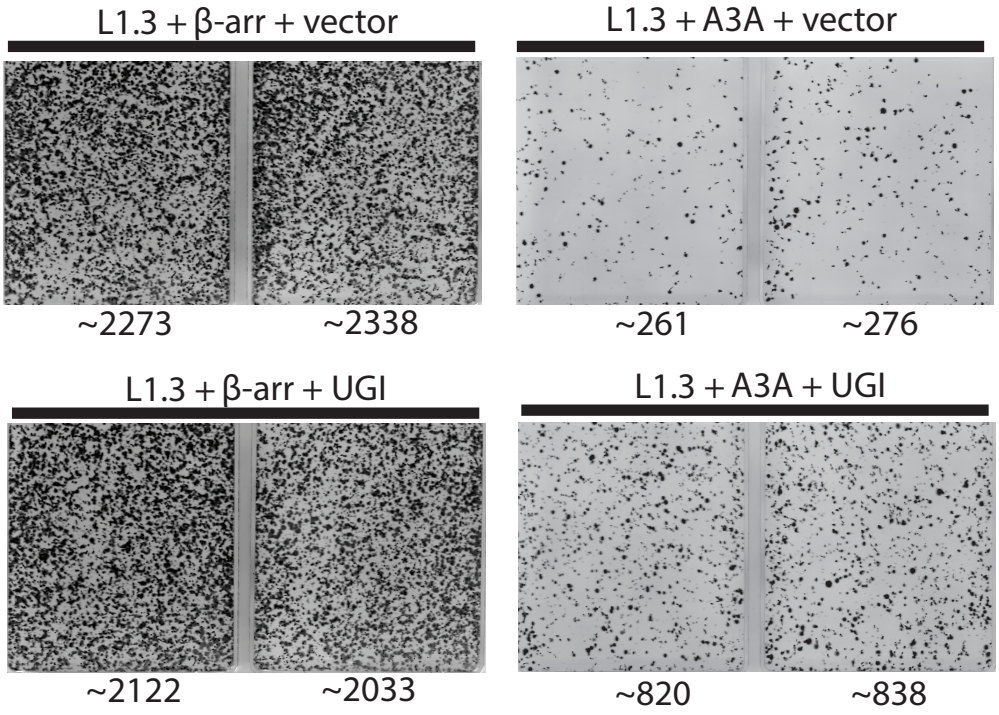


Figure 3.13: UGI co-expression does not enhance retrotransposition efficiency independently of A3A co-expression

Above left, raw colony counts from duplicate L1.3+β-arr+LGCX_vector co-transfections. Below left, raw colony counts from duplicate L1.3+β-arr+LGCX_UGI co-transfections. Above right, colony counts from duplicate L1.3+A3A+LGCX_vector co-transfections. Below right, colony counts from duplicate L1.3+A3A+LGCX_UGI co-transfections.

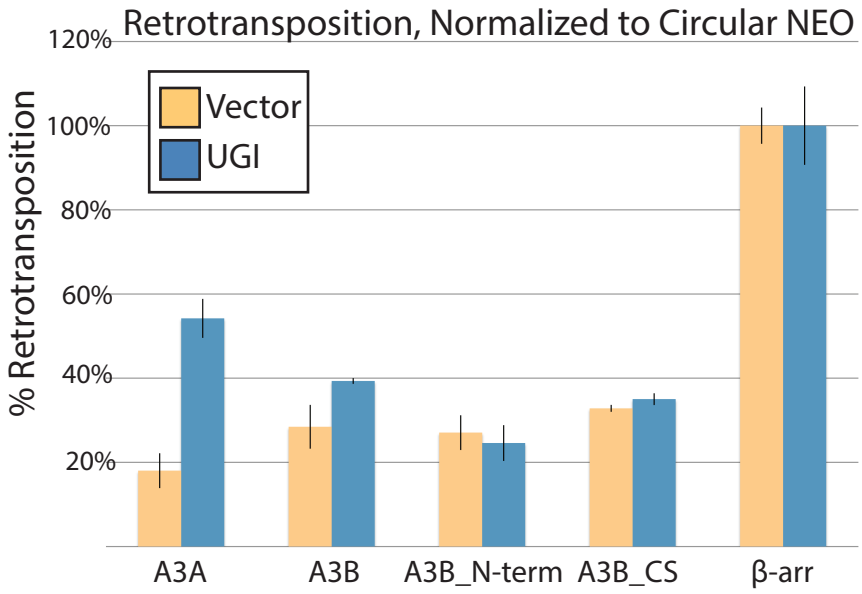


Figure 3.14: UGI does not alleviate inhibition by cytidine deaminase deaminase mutants of APOBEC3B

Approximately $1x10^5$ HeLa cells were plated per 10 cm dish, and transfected with 1 µg each of APOBEC3 vector, L1 vector, and pLGCX_UGI or pLGCX empty vector. Y-axis indicates percentage of retrotransposition relative to control, with the appropriate β-arr control for each experimental condition set to 100%. The APOBEC3 construct used is indicated along the x-axis. Dark blue bars show vector-only (pLGCX) control, yellow bars show inclusion of pLGCX_UGI. Data were normalized to circular NEO control co-transfections.

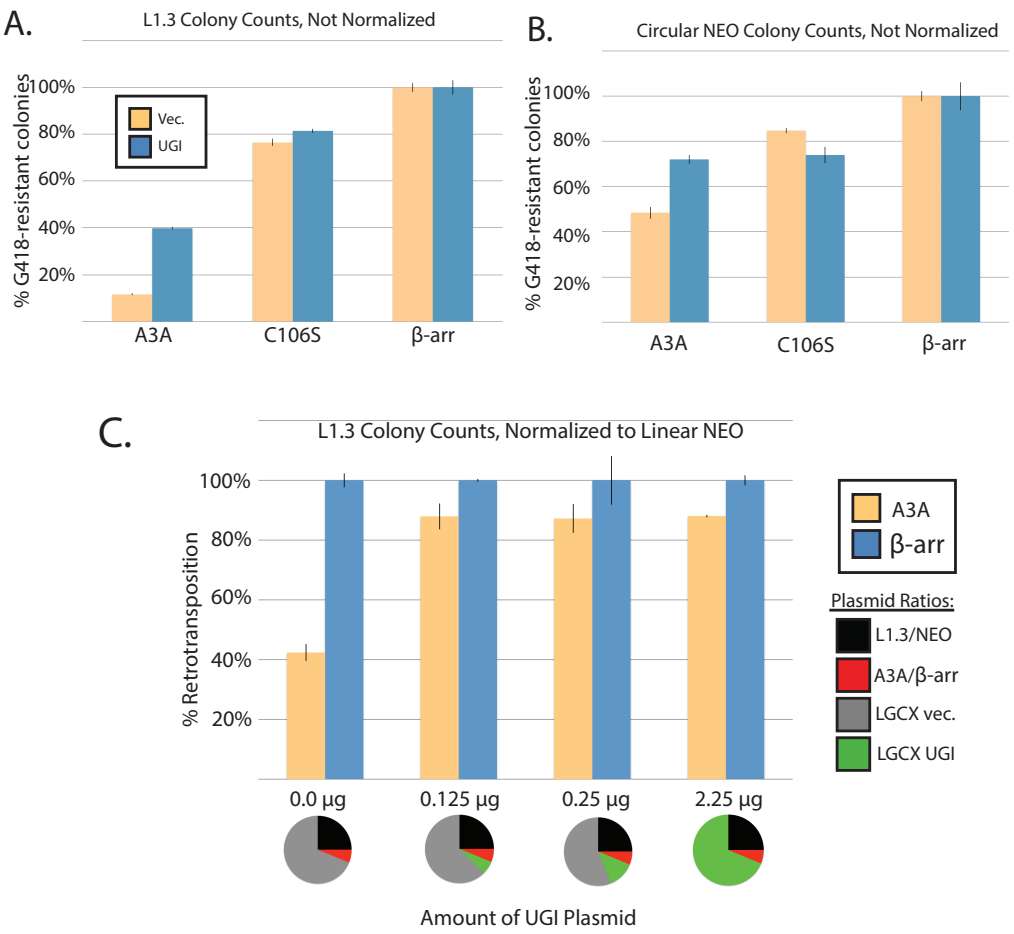
Figure 3.15: UGI expression specifically alleviates A3A-mediated L1 inhibition in HeLa cells

A) Colony counts, not normalized, corresponding to Figure 3.12 C. Colony counts from L1 retrotransposition assays, not normalized. Gold bars indicate LGCX vector control, blue bars indicate LGCX-UGI co-transfection. The y-axis indicates percentage of G418-resistant colonies relative to β-arr control. The x-axis shows the A3A expression vector used.

B) Colony counts from circular NEO control co-transfections, corresponding to Figure 3C. UGI co-expression partially alleviates A3A cytotoxicity. Gold bars indicate LGCX vector control, blue bars indicate LGCX-UGI co-transfection. The y-axis indicates % G418-resistant colonies relative to β-arr control. The x-axis shows the A3A expression vector used.

C) UGI alleviates A3A-mediated inhibition of retrotransposition independently of its impact on off-target effects. Here, we used .25 µg of A3A plasmid, to achieve minimal effect on NEO control experiments, while substantially inhibiting L1 retrotransposition (Figure 3.2, B-C). This control demonstrates that UGI expression specifically alleviates inhibition of L1, independently of effects on A3A cytotoxicity. Approximately $1x10^5$ HeLa cells were transfected with a total plasmid mass of 4 µg, the composition of which is represented in the pie charts below each bar graph. Assays were carried out in parallel with L1.3 and linear NEO control, which were held constant at 1 µg per transfection (black wedge). The amount of A3A plasmid (or β-arr control) was held constant at .25 µg (red wedge). LGCX_UGI expression vector (green wedge) was added in increasing amounts (0 µg, 0.125 µg, 0.25 µg, 2.25 µg, x-axis); the remaining plasmid mass for each co-transfection was supplied by LGCX empty vector (gray wedge). Retrotransposition assay was normalized to linear NEO. Y-axis indicates percent retrotransposition relative to β-arr control for each condition. Error bars represent percent standard deviation between duplicate transfections.

D) Retrotransposition assay colony counts and E) linear NEO colony counts, from the experiment shown in Supplemental Figure 3D. Approximately $1x10^5$ HeLa cells were transfected with a total plasmid mass of 4 µg, the composition of which is represented in the pie charts below each bar graph. Assays were carried out in parallel with L1.1 and linear NEO control, which were held constant at 1 µg per transfection (black wedge). The amount of A3A plasmid (or β-arr control) was held constant at .25 µg (red wedge). LGCX_UGI expression vector (green wedge) was added in increasing amounts (0 µg, 0.125 µg, 0.25 µg, 2.25 µg, x-axis); the remaining plasmid mass for each co-transfection was supplied by LGCX empty vector (gray wedge). Y-axes indicate percentage of G418-resistant colonies relative to β-arr control for each condition. In all graphs, error bars indicate percent standard deviation derived from duplicate transfections.


A.
L1.3 Colony Counts, Not Normalized
% G418-resistant colonies
100%
80%
60%
40%
20%
Vec.
UGI
A3A
C106S
β-arr
B.
Circular NEO Colony Counts, Not Normalized
% G418-resistant colonies
100%
80%
60%
40%
20%
A3A
C106S
β-arr
C.
L1.3 Colony Counts, Normalized to Linear NEO
% Retrotransposition
100%
80%
60%
40%
20%
A3A
β-arr
Plasmid Ratios:
L1.3/NEO
A3A/β-arr
LGCX vec.
LGCX UGI
0.0 µg
0.125 µg
0.25 µg
2.25 µg
Amount of UGI Plasmid
D.
L1.3 Colony Counts, Not Normalized
% G418-resistant colonies
100%
80%
60%
40%
20%
0.0 µg
0.125 µg
0.25 µg
2.25 µg
Amount of UGI Plasmid
E.
Linear NEO Colony Counts, Not Normalized
% G418-resistant colonies
100%
80%
60%
40%
20%
0.0 µg
0.125 µg
0.25 µg
2.25 µg
Amount of UGI Plasmid

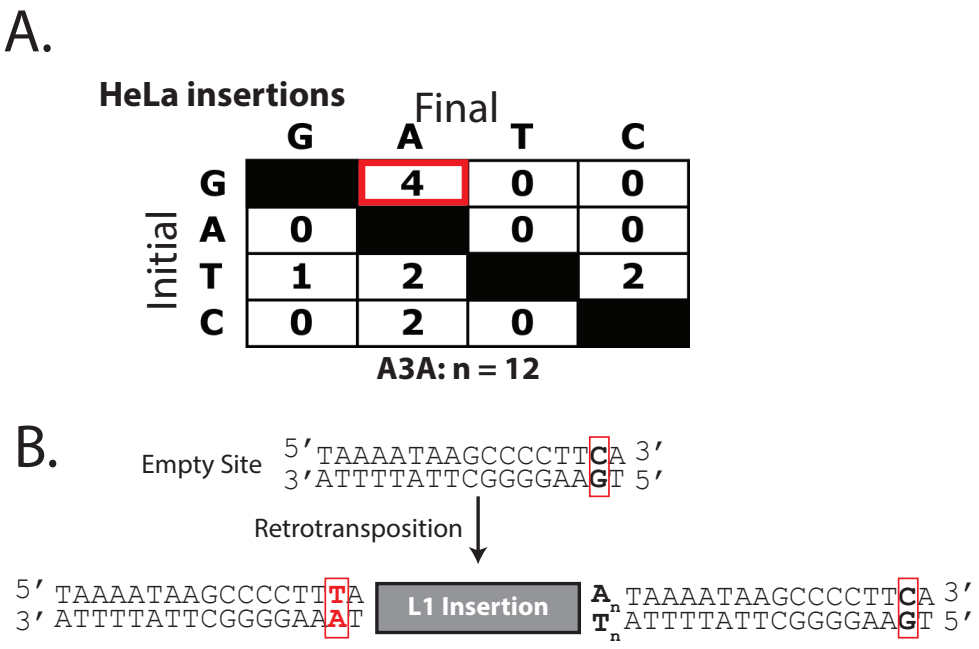


Figure 3.16: UGI expression reveals evidence of A3A-mediated editing of L1 retrotransposition events in HeLa cells

A) Summary of mutations found within retrotransposed L1 sequence among 12 insertions generated in the presence of A3A in HeLa cells. The number of G-to-A changes in the plus-strand sequence (indicative of deamination of the nascent first-strand cDNA) is highlighted in red.

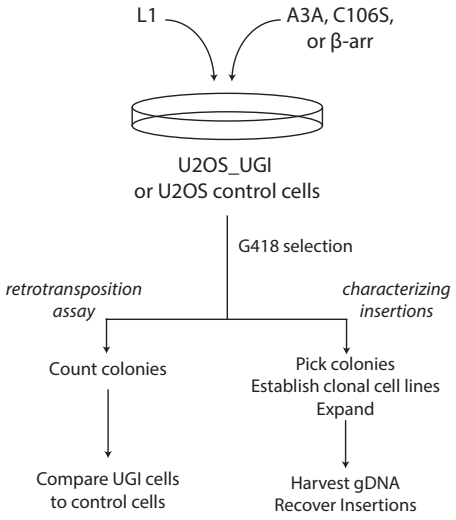
B) An L1 insertion harboring mismatched TSDs indicative of A3A-mediated editing. Above, the empty-site genomic sequence of the eventual target site duplications is shown. The targeted "C" is highlighted in red. Below, the structure of the L1 integration event, flanked by mismatched TSDs. The 3' TSD retains the original genomic sequence, while the 5' TSD bears evidence of A3A-mediated deamination.

Figure 3.17: Stable UGI expression alleviates A3A-mediated inhibition of retrotransposition in U2OS cells

A) Schematic for testing the effect of stable UGI expression on A3A-mediated L1 inhibition in U2OS cells, and generation of stable U2OS_UGI cell lines harboring retrotransposition events generated in the presence of A3A. Left branch, retrotransposition assay: Approximately $5x10^4$ U2OS_UGI or control U2OS cells per 10 cm dish were transfected with a total of 1.25 µg of DNA, including 1.0 µg of JM101 L1.3 plasmid. For the β-arr control transfections, 250 ng of pK_β-arr plasmid was used. For the A3A and C106S transfections, the remaining 250 ng consisted of 25 ng of pK_A3A plasmid or 25 ng of pk_A3A_C106S plasmid, and 225 ng of pK_β-arr plasmid. G418 selection (400 µg/ml) was initiated at 72h post-transfection and carried out for 14 days. Right branch, characterizing insertions: $5x10^5$ U2OS_UGI cells were plated per 15 cm dish, and transfected 24 hours later with 8 µg of JM140_L1.3_Δ2_k7 plasmid, and 2 µg of A3A, C106S, or β-arr plasmid. At 72 hours post-transfection, selection with 400 µg/ml G418 was initiated and carried out for 14 days. Individual colonies were isolated and expanded to establish clonal cell lines. Genomic DNA was prepared and subjected to the recovery procedure.

B) Stable UGI expression alleviates A3A-mediated L1 inhibition. This data point is taken from the titration experiment shown in Supplemental Figure 5. Left, results for control U2OS cells. Right, results for U2OS_UGI cells. The y-axes indicate percent retrotransposition relative to β-arr control. White bars indicate A3A, gray bars indicate C106S, black bars indicate β-arr. Error bars represent standard deviations between duplicate transfections.

A.


L1
A3A, C106S,
or β-arr
U2OS_UGI
or U2OS control cells
G418 selection
retrotransposition
assay
characterizing
insertions
Count colonies
Pick colonies
Establish clonal cell lines
Expand
Compare UGI cells
to control cells
Harvest gDNA
Recover Insertions

B.

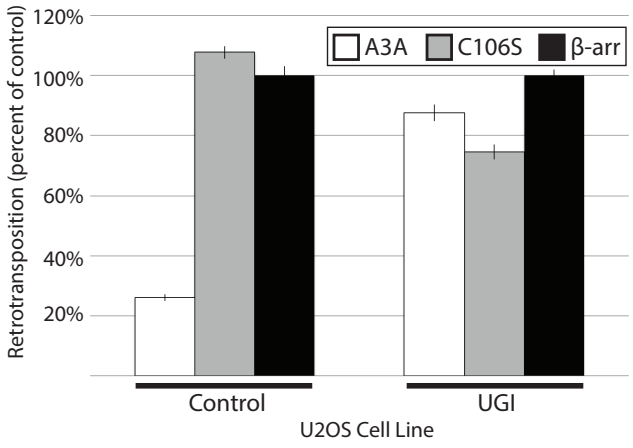

Retrotransposition (percent of control)
120%
100%
80%
60%
40%
20%
A3A
C106S
β-arr
Control
UGI
U2OS Cell Line

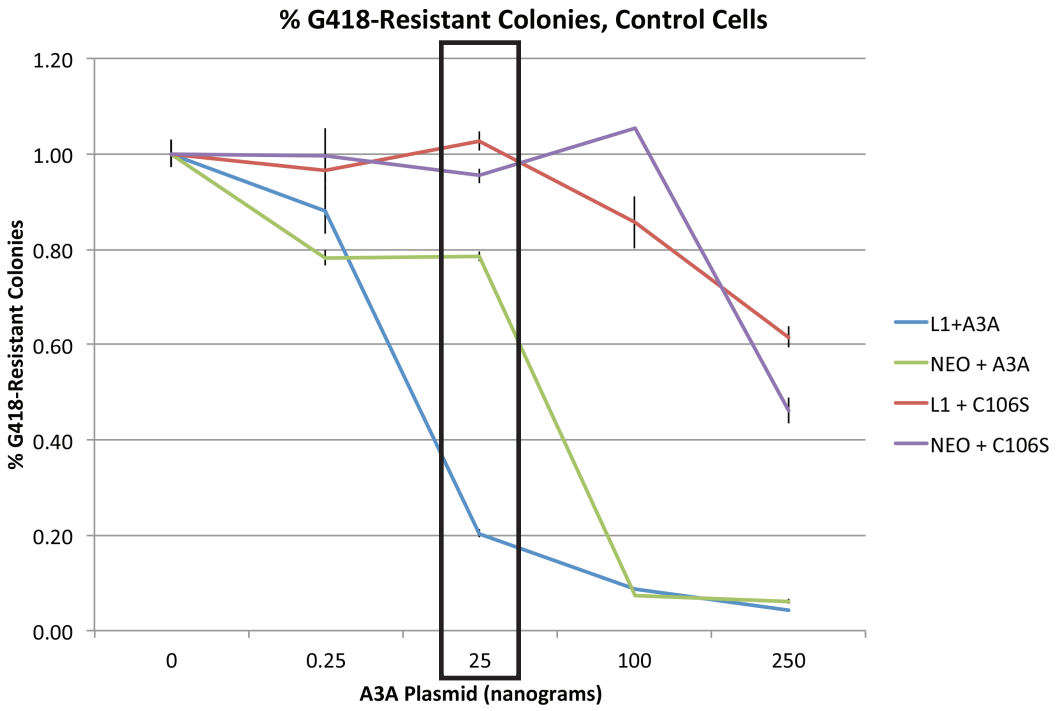
Figure 3.18: Titration to determine the appropriate amount of A3A plasmid to use in U2OS cells

A. Effect of A3A expression on G418-resistant colony generation by L1 retrotransposition or circular NEO plasmid in U2OS control cells. Approximately $5x10^4$ control U2OS cells were transfected with a total of 1.25 µg of DNA, including 1.0 µg of JM101/L1.3 plasmid. For the β-arr control transfections, 250 ng of pK_β-arr plasmid was used. A3A and C106S transfections consisted of, 0.25 ng, 25 ng, 100 ng, or 250 ng of pK_A3A or pk_A3A_C106S expression vector, and the appropriate amount of pK_β-arr plasmid to bring total plasmid mass to 1.25 µg. The y-axis shows %G418-resistant colonies, with β-arr control co-transfection (0 ng A3A) set to 100%. The x-axis shows the amount of A3A plasmid used.

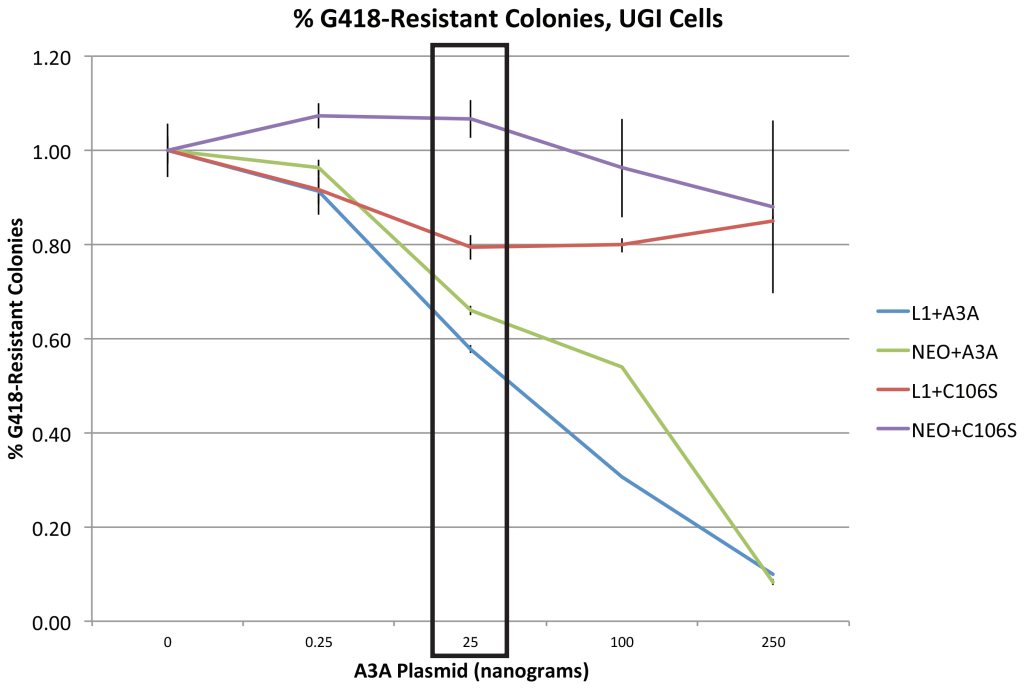
B. *Effect of A3A expression on G418-resistant colony generation by L1 retrotransposition or circular NEO plasmid in U2OS_UGI cells.* Approximately $5x10^4$ U2OS_UGI cells were transfected with a total of 1.25 µg of DNA, including 1.0 µg of JM101_L1.3 plasmid. For the β-arr control transfections, 250 ng of pK_β-arr plasmid was used. A3A and C106S transfections consisted of, 0.25 ng, 25 ng, 100 ng, or 250 ng of pK_A3A or pk_A3A_C106S expression vector, and the appropriate amount of pK_β-arr plasmid to bring total plasmid mass to 1.25 µg. The y-axis shows %G418-resistant colonies, with β-arr control co-transfection (0 ng A3A) set to 100%. The x-axis shows the amount of A3A plasmid used. This experiment was performed three times, representing biological replicates.

The data point at 25 ng of A3A plasmid is illustrated in the bar graph in Figure 3.17.

A.


% G418-Resistant Colonies, Control Cells
% G418-Resistant Colonies
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
0.25
25
100
250
A3A Plasmid (nanograms)
L1+A3A
NEO + A3A
L1 + C106S
NEO + C106S

B.


% G418-Resistant Colonies, UGI Cells
% G418-Resistant Colonies
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
0.25
25
100
250
A3A Plasmid (nanograms)
L1+A3A
NEO+A3A
L1+C106S
NEO+C106S

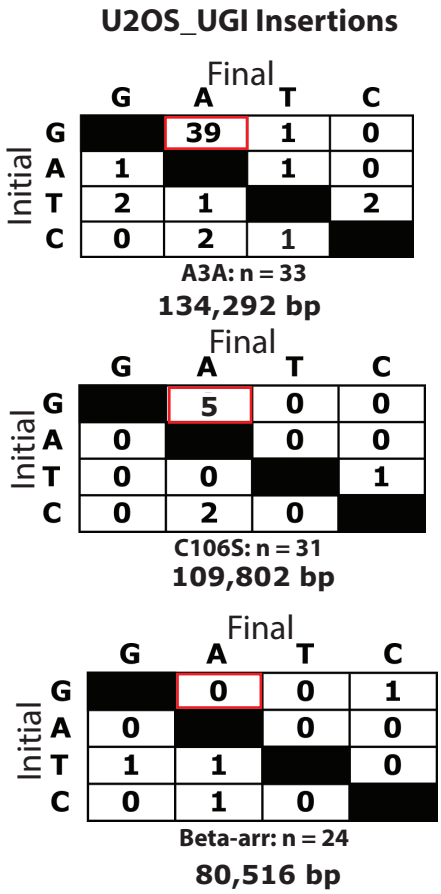


Figure 3.19: Stable UGI expression reveals evidence of A3A-mediated editing of L1 insertions in U2OS cells

Summary of G-to-A and other mutations found within retrotransposed L1 sequence among insertions generated in U2OS_UGI cells. The number of G-to-A changes in the plus-strand sequence (indicative of deamination of the nascent first-strand cDNA) is highlighted in red.

Figure 3.20: Schematic representation of L1 retrotransposition events generated in U2OS_UGI cells in the presence of A3A

Above, the JM140-L1.3-Δ^2-k7 retrotransposition indicator cassette is depicted [1]. Features of the plasmid are shown to scale: 5'UTR (green), ORF1 (light blue), inter-orf spacer (light gray), ORF2 (dark blue), 3'UTR (orange) NEO cassette (purple) with HSV_tk polyA signal (maroon), SV40 promoter/ori (pink), and ColE1 bacterial origin of replication (red). Below, the length and structural features of characterized insertions are represented graphically. The name of each cell line is shown at left. Nucleotide changes from the JM140-L1.3-deltadelta-k7 sequence are depicted as follows: stars indicate G-to-A changes, circles indicate any other nucleotide change. Blue shapes indicate changes in non-coding regions. Green shapes indicate silent mutations, yellow shapes indicate missesnse mutations, and red shapes indicate nonsense mutations.

A3A insertions; JM140_L1.3_deltadelta_k7:

Cell Line Name

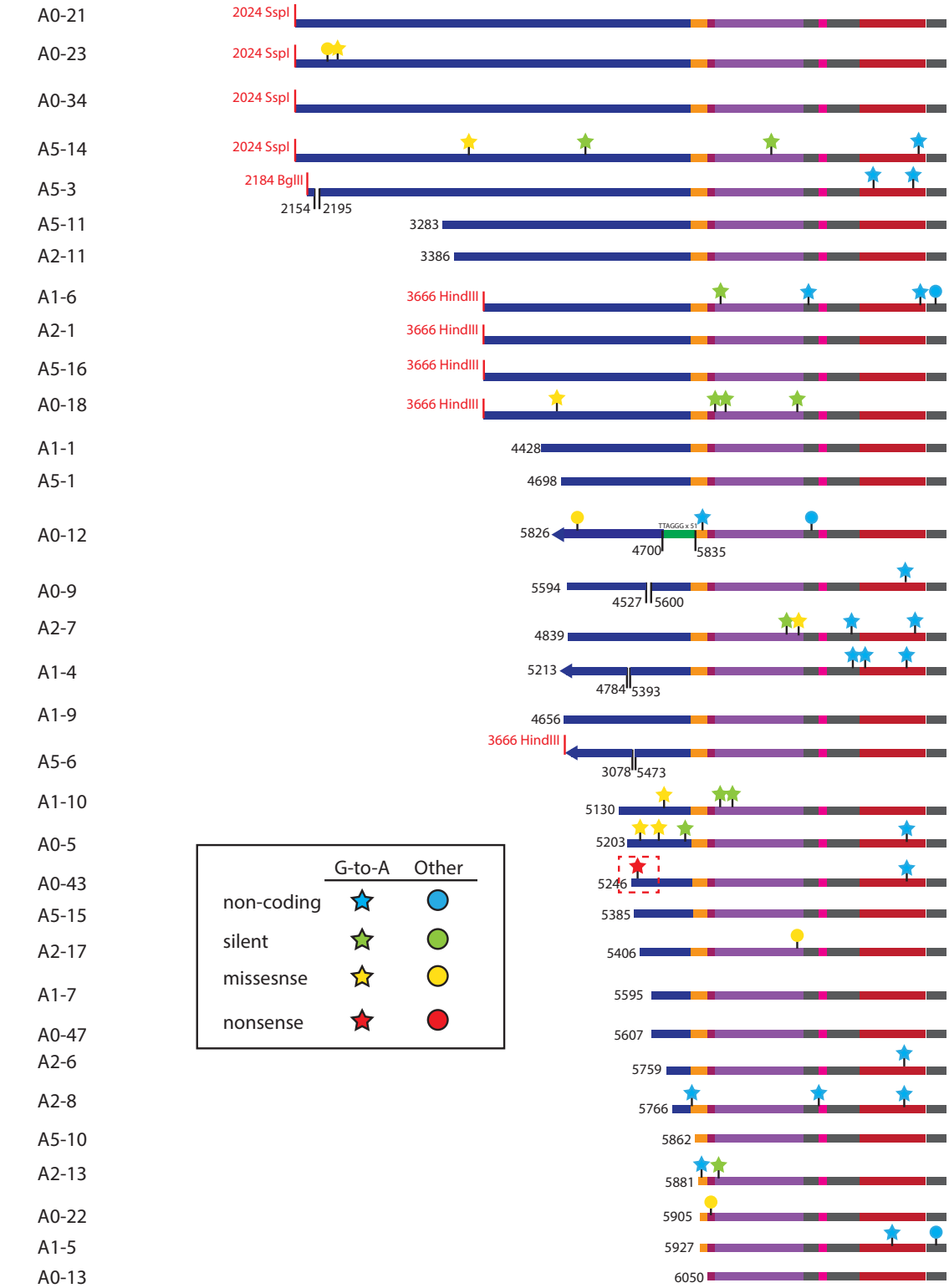


Figure 3.21: Schematic representation of L1 retrotransposition events generated in U2OS_UGI cells in the presence of β-arrestin

Above, the JM140-L1.3-Δ^2-k7 retrotransposition indicator cassette is depicted [1]. Features of the plasmid are shown to scale: 5'UTR (green), ORF1 (light blue), inter-orf spacer (light gray), ORF2 (dark blue), 3'UTR (orange) NEO cassette (purple) with HSV_tk polyA signal (maroon), SV40 promoter/ori (pink), and ColE1 bacterial origin of replication (red). Below, the length and structural features of characterized insertions are represented graphically. The name of each cell line is shown at left. Nucleotide changes from the JM140-L1.3-deltadelta-k7 sequence are depicted as follows: stars indicate G-to-A changes, circles indicate any other nucleotide change. Blue shapes indicate changes in non-coding regions. Green shapes indicate silent mutations, yellow shapes indicate missesnse mutations, and red shapes indicate nonsense mutations.

Beta-arr insertions; JM140_L1.3_deltadelta_k7:

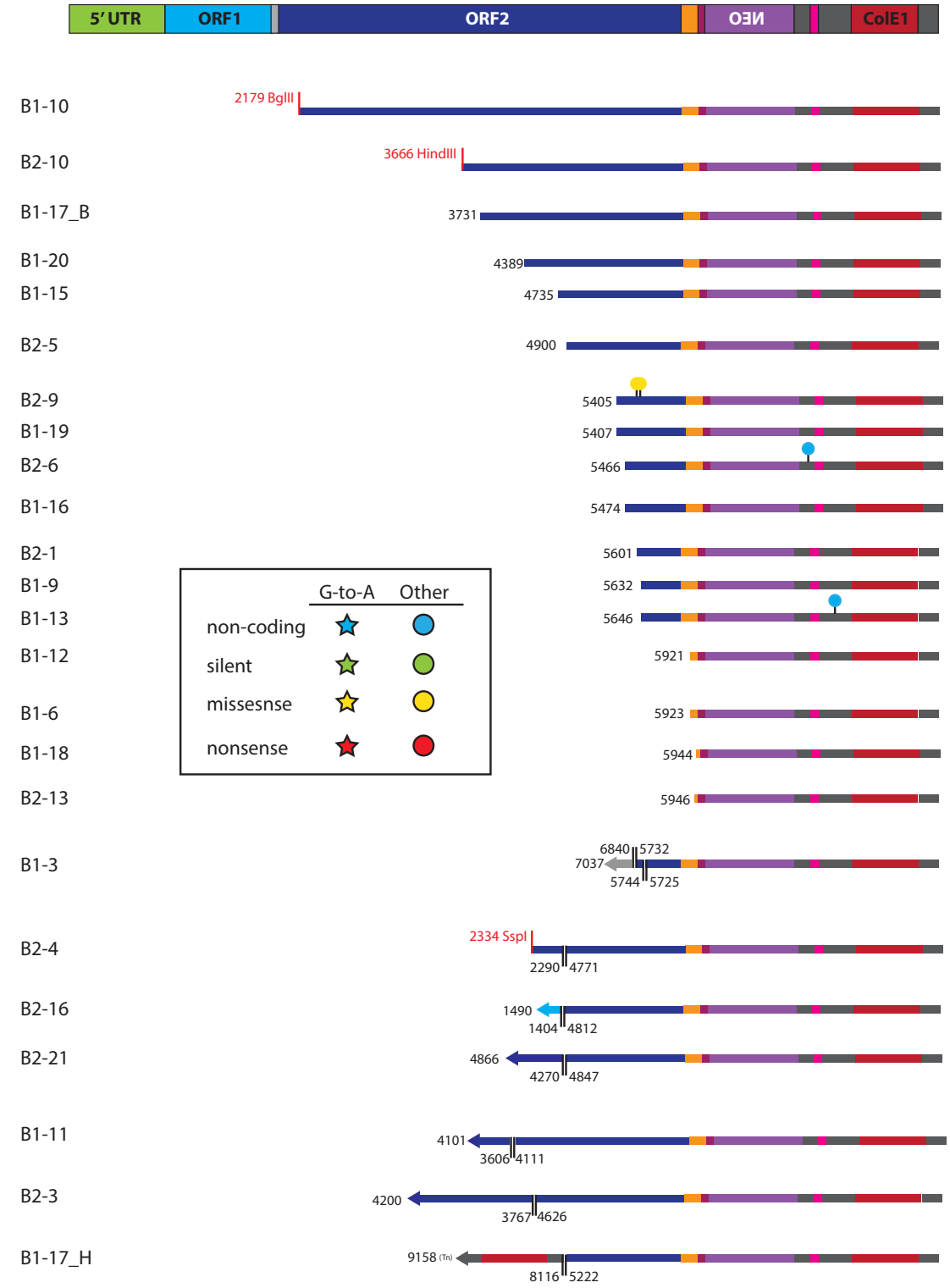
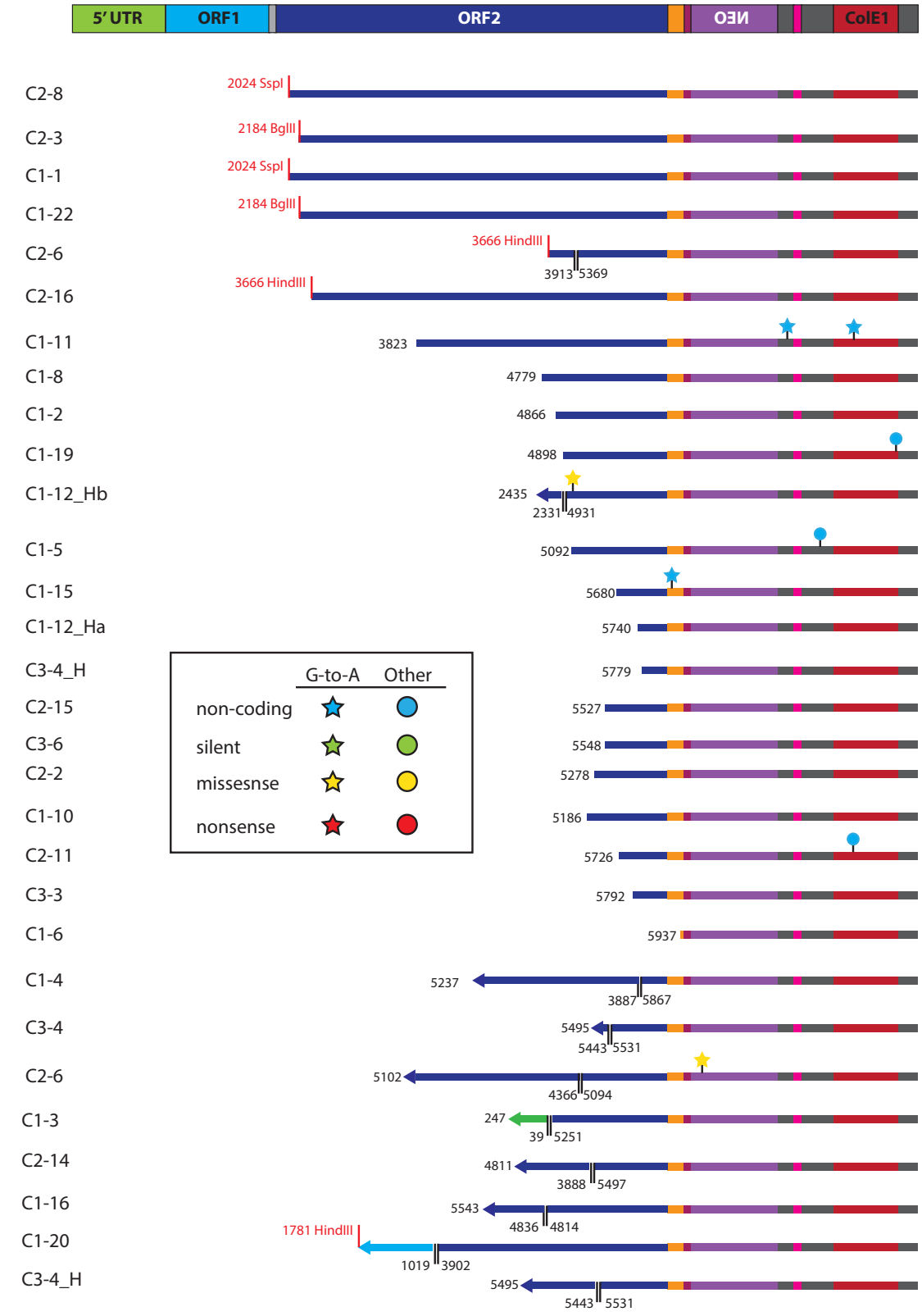


Figure 3.22: Schematic representation of L1 retrotransposition events generated in U2OS_UGI cells in the presence of A3A_C106S

Above, the JM140-L1.3-Δ^2-k7 retrotransposition indicator cassette is depicted [1]. Features of the plasmid are shown to scale: 5'UTR (green), ORF1 (light blue), inter-orf spacer (light gray), ORF2 (dark blue), 3'UTR (orange) NEO cassette (purple) with HSV_tk polyA signal (maroon), SV40 promoter/ori (pink), and ColE1 bacterial origin of replication (red). Below, the length and structural features of characterized insertions are represented graphically. The name of each cell line is shown at left. Nucleotide changes from the JM140-L1.3-deltadelta-k7 sequence are depicted as follows: stars indicate G-to-A changes, circles indicate any other nucleotide change. Blue shapes indicate changes in non-coding regions. Green shapes indicate silent mutations, yellow shapes indicate missesnse mutations, and red shapes indicate nonsense mutations.

C106S insertions; JM140_L1.3_deltadelta_k7:



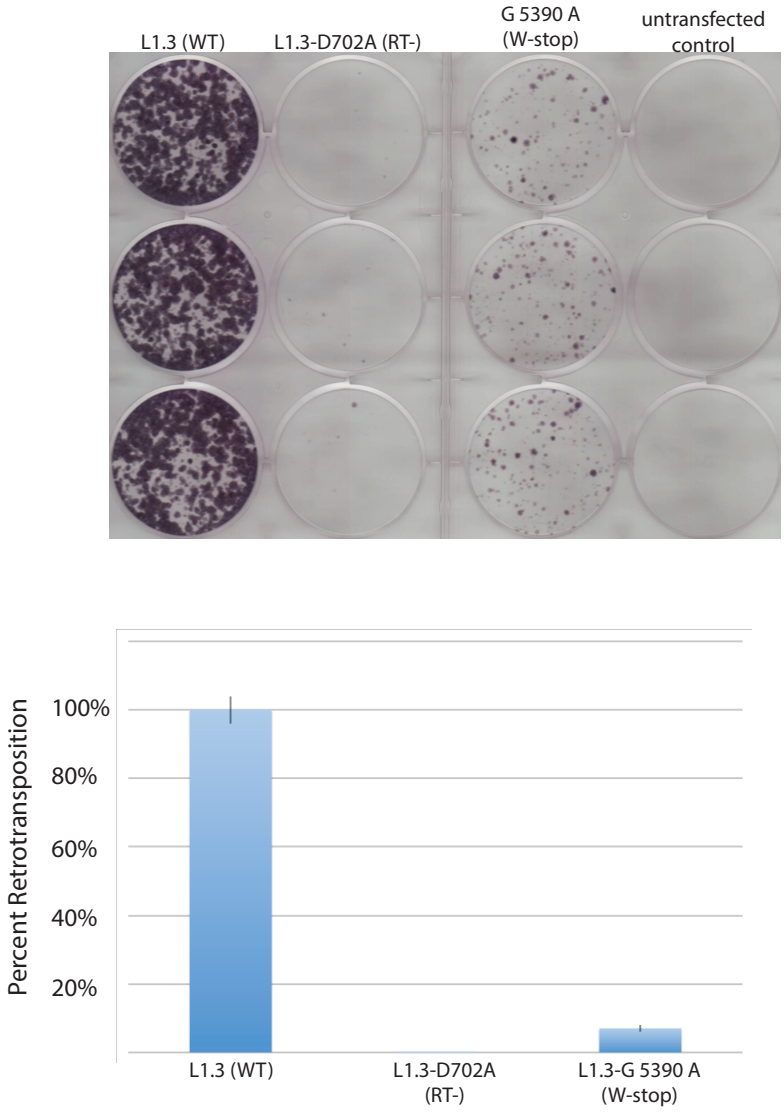


Figure 3.23: G 5390 A mutation substantially decreases L1 retrotransposition efficiency

The mutation G 5390 A was recapitulated by site-directed mutagenesis on wild-type pJM101/L1.3. For the retrotransposition assay, approximately $2x10^4$ U2OS_UGI cells were plated per well of a 6-well dish. 24 hours later, cells were transfected with 1 µg of retrotransposition indicator plasmid. Selection with 400 µg/ml G418 was initiated at 72 h post-transfection and carried out for 14 days. Cells were fixed, stained, and counted. The x-axis shows the elements used in the assay. RT- indicates the negative control pDK135, which bears a mutation in the RT domain that abolishes retrotransposition. The y-axis indicates % retrotransposition relative to pJM101/L1.3 control. Error bars indicate standard deviations derived from triplicate transfections. This experiment was performed twice, representing biological replicates.

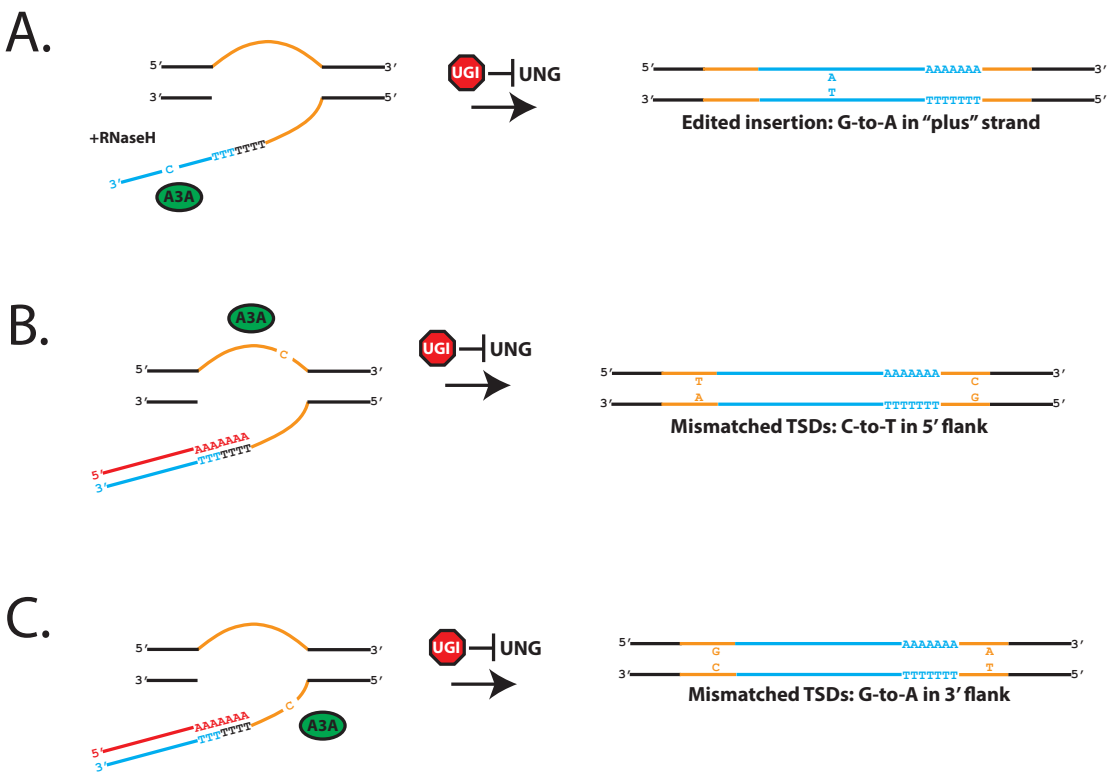


Figure 3.24: Modes of action by A3A at L1 TPRT, in the presence of UGI
Potential TPRT intermediates are illustrated at left. Transiently single-stranded genomic DNA regions that ultimately become target-site duplications are shown in orange. The L1 RNA is depicted in red; the L1 cDNA is shown in blue. The green oval represents A3A protein. The red "stop sign" represents UGI blocking the activity of UNG. On the right, the predicted signatures of deamination on integrated L1 insertions are illustrated.

A. Deamination of the L1 cDNA. If A3A deaminates the minus-strand L1 cDNA (presumably rendered vulnerable by the action of RnaseH) in the presence of UGI, mutations are evident as G-to-A changes on the plus-strand of the integrated L1.

B. Deamination of the 5' flanking genomic DNA. In the presence of UGI, deamination of the transiently single-stranded 5' flanking genomic DNA during TPRT (left side) results in a C-to-T change in the 5' TSD relative to the 3' TSD (right side).

C. Deamination of the 3' flanking genomic DNA. In the presence of UGI, deamination of the transiently single-stranded 3' flanking genomic DNA during TPRT is predicted to result in a G-to-A change in the 3' TSD relative to the 5' TSD.

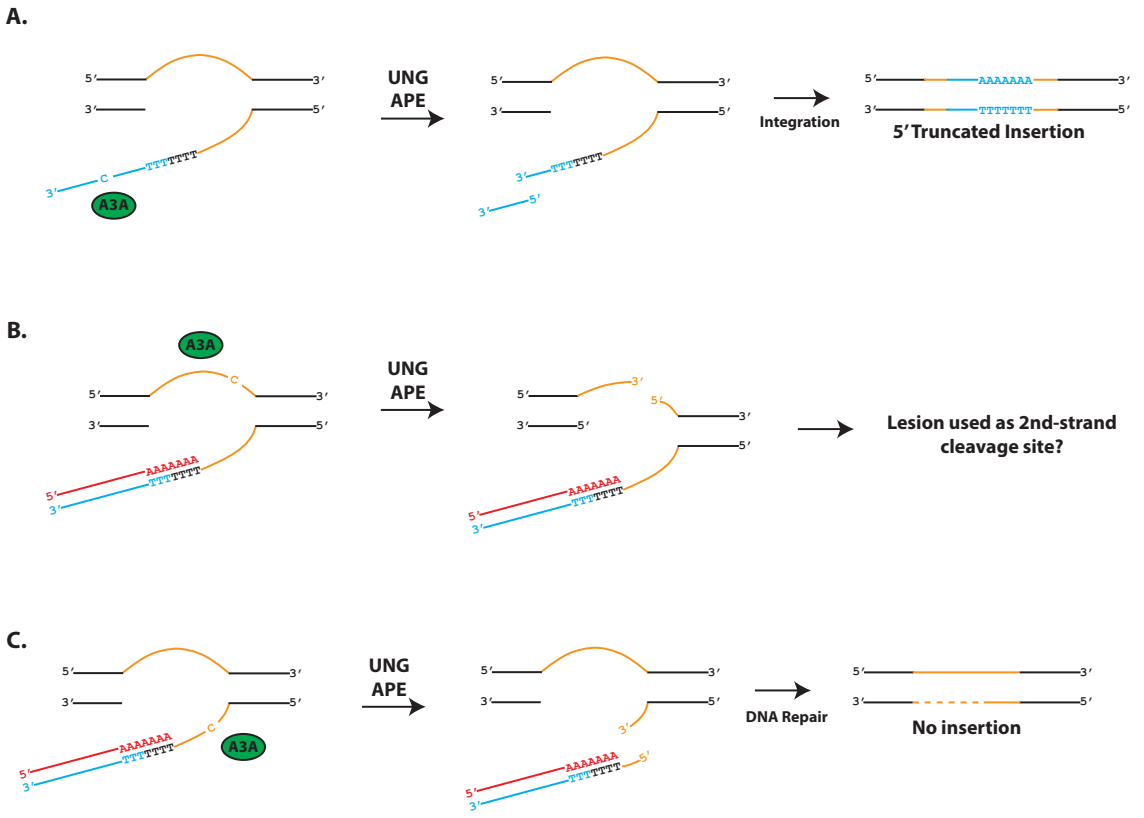


Figure 3.25: Modes of action by A3A at L1 TPRT Potential TPRT intermediates are illustrated at left. Transiently single-stranded genomic DNA regions that ultimately become target-site duplications are shown in orange. The L1 RNA is depicted in red; the L1 cDNA is shown in blue. The green oval represents A3A protein. On the right, predicted outcomes of A3A-mediated deamination followed by TPRT intermediate cleavage are illustrated.

A. Deamination of the L1 cDNA. If A3A deaminates the minus-strand L1 cDNA (presumably rendered vulnerable by the action of RnaseH), the actions of UNG and APE could result in cleavage of the L1 cDNA, giving rise to a 5' truncated L1 insertion.

B. Deamination of the 5' flanking genomic DNA. Deamination of the transiently single-stranded 5' flanking genomic DNA during TPRT (left side), followed by the actions of UNG and APE, could result in top strand cleavage, that could substitute for L1-mediated second-strand cleavage during integration.

C. Deamination of the 3' flanking genomic DNA. Deamination of the transiently single-stranded 3' flanking genomic DNA during TPRT, followed by the actions of UNG and APE, could lead to loss of the entire L1 integration intermediate. The resulting lesion could be repaired using the top strand genomic DNA as a template, leaving no evidence of the thwarted insertion.

Table 3.1: Analysis of rA3A target site preference on LEAP products

Each sub-table lists all possible trinucleotide contexts with a “G” as the middle nucleotide (corresponding to a “C” on the first-strand LEAP cDNA). The first column shows how many times each trinucleotide appears in the LEAP product sequence. Greyed-out rows indicate that the trinucleotide is not present in the LEAP product sequence. The second column lists how many times this each trinucleotide is available in 100 LEAP products. The third column quantifies how many times a deamination event was observed within each trinucleotide. The fourth column quantifies the percentage of available sites deaminated in 100 LEAP products. Below, colored rows express frequencies of deamination as a percentage of total available G nucleotides, total G nucleotides residing in published A3A target sites (TGA, GGA), total published A3A target sites plus A3A, which was prefered by rA3A in this assay (TGA, GGA, AGA), and total TGAs only. To the right, an additional cell lists the percentage of adapter sequences deaminated, at the sole “C” nucleotide which resides within the

	# of sites in LEAP Product	available sites in 100 products	deaminated in RNP alone	percentage of available sites deaminated in 100 products	
GGG					
AGG					
TGG	1	100	1	1.0%	
CGG					
GGA	1	100	0	0.0%	
AGA	2	200	0	0.0%	
TGA	6	600	2	0.3%	
CGA					
GGT					
AGT	3	300	0	0.0%	
TGT	3	300	0	0.0%	
CGT					
GGC					
AGC	1	100	0	0.0%	
TGC	5	500	1	0.2%	
CGC					
	Total G's:	2200	4	0.2%	
	Total published A3A target sites (GGA, TGA):	700	3	0.4%	
	Totalpublished A3A target sites, plus AGA	900	3	0.3%	**adapters deaminated:**
	Total TGA's	600	2	0.3%	1.0%

	# of sites in LEAP Product	available sites in 100 products	deaminated in A3A alone	percentage of available sites deaminated in 100 products	
GGG					
AGG					
TGG	1	100	3	3.0%	
CGG					
GGA	1	100	3	3.0%	
AGA	2	200	13	6.5%	
TGA	6	600	25	4.2%	
CGA					
GGT					
AGT	3	300	0	0.0%	
TGT	3	300	0	0.0%	
CGT					
GGC					
AGC	1	100	0	0.0%	
TGC	5	500	0	0.0%	
CGC					
	Total G's:	2200	44	2.0%	
	Total published A3A target sites (GGA, TGA):	700	28	4.0%	
	Totalpublished A3A target sites, plus AGA	900	41	4.6%	**adapters deaminated:**
	Total TGA's	600	25	4.2%	91.0%

	# of sites in LEAP Product	available sites in 100 products	deaminated in C106S alone	percentage of available sites deaminated in 100 products	
GGG					
AGG					
TGG	1	100	0	0.0%	
CGG					
GGA	1	100	0	0.0%	
AGA	2	200	2	1.0%	
TGA	6	600	2	0.3%	
CGA					
GGT					
AGT	3	300	0	0.0%	
TGT	3	300	1	0.3%	
CGT					
GGC					
AGC	1	100	0	0.0%	
TGC	5	500	0	0.0%	
CGC					
	Total G's:	2200	5	0.2%	
	Total published A3A target sites (GGA, TGA):	700	2	0.3%	
	Totalpublished A3A target sites, plus AGA	900	4	0.4%	**adapters deaminated:**
	Total TGA's	600	2	0.3%	1.0%

	# of sites in LEAP Product	available sites in 100 products	deaminated in RNP+RNaseH	percentage of available sites deaminated in 100 products	
GGG					
AGG					
TGG	1	100	0	0.0%	
CGG					
GGA	1	100	4	4.0%	
AGA	2	200	19	9.5%	
TGA	6	600	25	4.2%	
CGA					
GGT					
AGT	3	300	0	0.0%	
TGT	3	300	0	0.0%	
CGT					
GGC					
AGC	1	100	0	0.0%	
TGC	5	500	0	0.0%	
CGC					
	Total G's:	2200	48	2.2%	
	Total published A3A target sites (GGA, TGA):	700	25	3.6%	
	Totalpublished A3A target sites, plus AGA	900	44	4.9%	**adapters deaminated:**
	Total TGA's	600	25	4.2%	2.0%

	# of sites in LEAP Product	available sites in 100 products	deaminated in A3A+RNaseH	percentage of available sites deaminated in 100 products	
GGG					
AGG					
TGG	1	100	5	5.0%	
CGG					
GGA	1	100	5	5.0%	
AGA	2	200	75	37.5%	
TGA	6	600	228	38.0%	
CGA					
GGT					
AGT	3	300	4	1.3%	
TGT	3	300	8	2.7%	
CGT					
GGC					
AGC	1	100	0	0.0%	
TGC	5	500	5	1.0%	
CGC					
	Total G's:	2200	330	15.0%	
	Total published A3A target sites (GGA, TGA):	700	233	33.3%	
	Totalpublished A3A target sites, plus AGA	900	308	34.2%	**adapters deaminated:**
	Total TGA's	600	228	38.0%	50.0%

	# of sites in LEAP Product	available sites in 100 products	deaminated in C106S+RNase	percentage of available sites deaminated in 100 products	
GGG					
AGG					
TGG	1	100	1	1.0%	
CGG					
GGA	1	100	2	2.0%	
AGA	2	200	15	7.5%	
TGA	6	600	40	6.7%	
CGA					
GGT					
AGT	3	300	0	0.0%	
TGT	3	300	0	0.0%	
CGT					
GGC					
AGC	1	100	0	0.0%	
TGC	5	500	0	0.0%	
CGC					
	Total G's:	2200	58	2.6%	
	Total published A3A target sites (GGA, TGA):	700	41	5.9%	
	Totalpublished A3A target sites, plus AGA	900	56	6.2%	**adapters deaminated:**
	Total TGA's	600	40	6.7%	1.0%

Table 3.2: Detailed summary of the number and location of deamination events on each of 100 rA3A+RNaseH LEAP products, and poly-A tail length

First column: LEAP products generated in the presence of A3A. Products were derived from four independent L1 RT reactions, indicated by colored shading (yellow, orange, pink, purple). Second column: Poly-A tail length of individual LEAP products. Third column: total G-to-A changes in each product. Fourth Column: total non G-to-A changes in each product. The next 18 columns denote the positions of G bases deaminated in rA3A-LEAP products; for each product, a "1" indicates a deamination at that position. The next column, Other, indicates the position of non G-to-A changes. The final column, Adapter TCA, indicates whether the single-stranded LEAP adapter was deaminated for each product; a "1" that the single "C" in the adapter was deaminated in that product.

Sample no.	polyA length	Total GtoA	Total Other	G29	G34	G39	G47	G50	G56	G57	G72	G76	G79	G81	G99	G101	G108	G110	G113	G128	G144	G155	Other	Adapter (TCA)
A3A+RNaseH 1	56	0																						1
A3A+RNaseH 2	68	2			1									1										1
A3A+RNaseH 3	101	8		1	1	1	1	1			1			1				1						1
A3A+RNaseH 4	62	2					1			1														1
A3A+RNaseH 5	85	7			1		1	1	1	1	1			1										
A3A+RNaseH 6	77	9		1	1	1	1	1			1			1	1	1								1
A3A+RNaseH 7	12-C-41	6		1	1	1	1	1						1										1
A3A+RNaseH 8	63	0	1																				A104T	
A3A+RNaseH 9	44	1									1													
A3A+RNaseH 10	91	3	1	1		1										1							T148C	
A3A+RNaseH 11	84	5			1		1	1						1				1						1
A3A+RNaseH 12	94	3			1						1			1										1
A3A+RNaseH 13	61	6			1	1	1				1	1		1										1
A3A+RNaseH 14	38	3									1			1				1						
A3A+RNaseH 15	88	0																						
A3A+RNaseH 16	92	4			1		1				1			1										1
A3A+RNaseH 17	86	7		1	1		1				1			1		1		1						1
A3A+RNaseH 18	48	5			1		1	1						1				1						1
A3A+RNaseH 19	85	6		1	1	1	1							1				1						1
A3A+RNaseH 20	85	2			1								1											1
A3A+RNaseH 21	84	0																						
A3A+RNaseH 22	47	0																						
A3A+RNaseH 23	94	9			1	1	1	1			1			1	1	1	1							1
A3A+RNaseH 24	48	10		1	1	1	1	1			1			1		1		1				1		
A3A+RNaseH 25	61	0																						1
A3A+RNaseH 26	57	1												1										
A3A+RNaseH 27	55	7		1	1	1	1							1		1		1						1
A3A+RNaseH 28	39-GATGGCTAATGAAT-12	7			1	1	1	1								1		1	1					1
A3A+RNaseH 29	45	6			1		1	1			1			1				1						1
A3A+RNaseH 30	50	5			1			1			1			1		1								1
A3A+RNaseH 31	46-GACGATC-15	3			1	1								1										
A3A+RNaseH 32	81	6			1	1	1	1			1			1										1
A3A+RNaseH 33	51	0																						
A3A+RNaseH 34	84	0	1																				A88G	
A3A+RNaseH 35	60	1									1													
A3A+RNaseH 36	99	5		1			1				1			1		1								
A3A+RNaseH 37	79	8			1	1	1	1			1			1		1		1						1
A3A+RNaseH 38	96	3					1				1			1										1
A3A+RNaseH 39	31	10		1	1	1	1	1			1			1		1		1		1				1
A3A+RNaseH 40	31	0																						1
A3A+RNaseH 41	58	6		1		1	1	1						1				1						1
A3A+RNaseH 42	93	4		1	1	1					1													1
A3A+RNaseH 43	30	6		1	1	1	1				1			1										
A3A+RNaseH 44	72	5			1		1							1		1		1						1
A3A+RNaseH 45	54	3			1									1		1								
A3A+RNaseH 46	60	0																						1
A3A+RNaseH 47	58	0																						
A3A+RNaseH 48	59	0																						1
A3A+RNaseH 49	62	1			1																			
A3A+RNaseH 50	31	5			1		1	1						1				1						1
A3A+RNaseH 51	42	4		1	1		1									1								
A3A+RNaseH 52	50	5			1		1		1					1		1								1
A3A+RNaseH 53	48	0																						
A3A+RNaseH 54	12	0																						1
A3A+RNaseH 55	91	4				1	1							1						1				
A3A+RNaseH 56	55	0																						
A3A+RNaseH 57	68	1									1													
A3A+RNaseH 58	51	2			1	1																		1
A3A+RNaseH 59	62	5		1	1		1				1			1										
A3A+RNaseH 60	50	0																						1
A3A+RNaseH 61	49	0																						

A3A+RNaseH 62	51	1										1																			
A3A+RNaseH 63	79	4			1	1	1					1																			1
A3A+RNaseH 64	57	6			1	1	1	1							1				1												
A3A+RNaseH 65	52	4						1				1			1						1										1
A3A+RNaseH 66	81	3				1					1				1																
A3A+RNaseH 67	93	2										1						1													1
A3A+RNaseH 68	58	3					1	1	1																						1
A3A+RNaseH 69	54	4				1			1			1			1																1
A3A+RNaseH 70	39	1																1													
A3A+RNaseH 71	107	6			1	1	1			1					1			1													
A3A+RNaseH 72	37	1													1																
A3A+RNaseH 73	62	4				1	1					1			1																
A3A+RNaseH 74	63	7			1	1	1					1			1			1			1										1
A3A+RNaseH 75	40	0																													
A3A+RNaseH 76	76	5						1				1				1					1			1							1
A3A+RNaseH 77	77	6				1		1							1			1			1					1					
A3A+RNaseH 78	100	0																													
A3A+RNaseH 79	50	4				1			1									1			1										1
A3A+RNaseH 80	60	1										1																			1
A3A+RNaseH 81	38	0																													
A3A+RNaseH 82	43	1				1																									1
A3A+RNaseH 83	59	5					1	1							1						1					1					1
A3A+RNaseH 84	54	4						1							1	1	1														
A3A+RNaseH 85	28	3			1	1									1																1
A3A+RNaseH 86	69	5			1		1	1				1			1																
A3A+RNaseH 87	55	7	1		1	1	1	1				1			1						1								A64T		1
A3A+RNaseH 88	32	5			1		1	1							1						1										1
A3A+RNaseH 89	55	2								1		1																			1
A3A+RNaseH 90	69	4			1		1	1							1																
A3A+RNaseH 91	64	8				1	1	1		1		1			1			1			1										
A3A+RNaseH 92	62	4					1	1	1			1																			
A3A+RNaseH 93	104	2					1	1																							
A3A+RNaseH 94	60	7	1			1	1		1	1	1				1			1											T97A		
A3A+RNaseH 95	53	3				1										1	1														
A3A+RNaseH 96	44	2										1			1																
A3A+RNaseH 97	56	5				1		1	1			1			1																
A3A+RNaseH 98	53	3			1	1									1																
A3A+RNaseH 99	47	4				1			1									1			1										
A3A+RNaseH 100	38	8			1	1	1	1				1			1			1			1										
			TOTAL per position:		26	52	35	45	23	5	5	40	1	1	55	5	27	1	26	2	2	1	3								
					AGA	**TGA**	**AGA**	**TGA**	**TGA**	**TGG**	**GGA**	**AGA**	**TGC**	**AGT**	**TGA**	**TGT**	**TGA**	**TGT**	**TGA**	**TGC**	**TGT**	**TGC**	**AGT**								50

References

1. Gilbert, N., et al., *Multiple fates of L1 retrotransposition intermediates in cultured human cells.* Mol Cell Biol, 2005. **25**(17): p. 7780-95.
2. Gilbert, N., S. Lutz-Prigge, and J.V. Moran, *Genomic deletions created upon LINE-1 retrotransposition.* Cell, 2002. **110**(3): p. 315-25.
3. Sheehy, A.M., et al., *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.* Nature, 2002. **418**(6898): p. 646-50.
4. Wiegand, H.L., et al., *A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins.* EMBO J, 2004. **23**(12): p. 2451-8.
5. Zheng, Y.H., et al., *Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication.* J Virol, 2004. **78**(11): p. 6073-6.
6. Doehle, B.P., A. Schafer, and B.R. Cullen, *Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif.* Virology, 2005. **339**(2): p. 281-8.
7. Turelli, P., et al., *Inhibition of hepatitis B virus replication by APOBEC3G.* Science, 2004. **303**(5665): p. 1829.
8. Zhang, W., et al., *Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression.* Cell Microbiol, 2008. **10**(1): p. 112-21.
9. Rosler, C., et al., *APOBEC-mediated interference with hepadnavirus production.* Hepatology, 2005. **42**(2): p. 301-9.
10. Suspene, R., et al., *Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo.* Proc Natl Acad Sci U S A, 2005. **102**(23): p. 8321-6.
11. Chen, H., et al., *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.* Curr Biol, 2006. **16**(5): p. 480-5.
12. Sawyer, S.L., M. Emerman, and H.S. Malik, *Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G.* PLoS Biol, 2004. **2**(9): p. E275.
13. Bogerd, H.P., et al., *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8780-5.
14. Muckenfuss, H., et al., *APOBEC3 proteins inhibit human LINE-1 retrotransposition.* J Biol Chem, 2006. **281**(31): p. 22161-72.
15. Niewiadomska, A.M., et al., *Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association.* J Virol, 2007. **81**(17): p. 9577-83.
16. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.* J Biol Chem, 2006. **281**(25): p. 16837-41.

17. OhAinle, M., et al., *Antiretroelement activity of APOBEC3H was lost twice in recent human evolution.* Cell Host Microbe, 2008. **4**(3): p. 249-59.
18. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
19. Wei, W., et al., *A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events.* Anal Biochem, 2000. **284**(2): p. 435-8.
20. Sassaman, D.M., et al., *Many human L1 elements are capable of retrotransposition.* Nat Genet, 1997. **16**(1): p. 37-43.
21. Goodier, J.L., et al., *A novel active L1 retrotransposon subfamily in the mouse.* Genome Res, 2001. **11**(10): p. 1677-85.
22. Han, J.S. and J.D. Boeke, *A highly active synthetic mammalian retrotransposon.* Nature, 2004. **429**(6989): p. 314-8.
23. Sugano, T., M. Kajikawa, and N. Okada, *Isolation and characterization of retrotransposition-competent LINEs from zebrafish.* Gene, 2006. **365**: p. 74-82.
24. Luan, D.D., et al., *Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.* Cell, 1993. **72**(4): p. 595-605.
25. Feng, Q., et al., *Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.* Cell, 1996. **87**(5): p. 905-16.
26. Morrish, T.A., et al., *DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.* Nat Genet, 2002. **31**(2): p. 159-65.
27. Coufal, N.G., et al., *Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(51): p. 20382-7.
28. Kulpa, D.A. and J.V. Moran, *Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles.* Nat Struct Mol Biol, 2006. **13**(7): p. 655-60.
29. Langlois, M.A. and M.S. Neuberger, *Human APOBEC3G can restrict retroviral infection in avian cells and acts independently of both UNG and SMUG1.* J Virol, 2008. **82**(9): p. 4660-4.
30. Wissing, S., et al., *Endogenous APOBEC3B Restricts LINE-1 Retrotransposition in Transformed Cells and Human Embryonic Stem Cells.* J Biol Chem, 2011. **286**(42): p. 36427-37.
31. Kaiser, S.M. and M. Emerman, *Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G.* J Virol, 2006. **80**(2): p. 875-82.
32. Landry, S., et al., *APOBEC3A can activate the DNA damage response and cause cell-cycle arrest.* EMBO Rep, 2011. **12**(5): p. 444-50.
33. Kurzynska-Kokorniak, A., et al., *DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon.* J Mol Biol, 2007. **374**(2): p. 322-33.

34. Stenglein, M.D., et al., *APOBEC3 proteins mediate the clearance of foreign DNA from human cells.* Nat Struct Mol Biol, 2010. **17**(2): p. 222-9.
35. Yu, K., et al., *Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops.* Molecular and cellular biology, 2005. **25**(5): p. 1730-6.
36. Yu, K., et al., *R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells.* Nature immunology, 2003. **4**(5): p. 442-51.
37. Gasior, S.L., A.M. Roy-Engel, and P.L. Deininger, *ERCC1/XPF limits L1 retrotransposition.* DNA repair, 2008. **7**(6): p. 983-9.
38. Suspene, R., et al., *Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism.* Proc Natl Acad Sci U S A, 2011. **108**(12): p. 4858-63.
39. Suspene, R., et al., *Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR.* J Gen Virol, 2005. **86**(Pt 1): p. 125-9.
40. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-8.
41. Bogerd, H.P., et al., *APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.* Nucleic Acids Res, 2006. **34**(1): p. 89-95.
42. Dombroski, B.A., A.F. Scott, and H.H. Kazazian, Jr., *Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element.* Proc Natl Acad Sci U S A, 1993. **90**(14): p. 6513-7.
43. Kulpa, D.A. and J.V. Moran, *Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition.* Hum Mol Genet, 2005. **14**(21): p. 3237-48.
44. Wei, W., et al., *Human L1 retrotransposition: cis preference versus trans complementation.* Mol Cell Biol, 2001. **21**(4): p. 1429-39.
45. Kopera, H.C., et al., *Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase.* Proc Natl Acad Sci U S A, 2011.
46. Naas, T.P., et al., *An actively retrotransposing, novel subfamily of mouse L1 elements.* EMBO J, 1998. **17**(2): p. 590-7.
47. Doucet, A.J., et al., *Characterization of LINE-1 ribonucleoprotein particles.* PLoS Genet, 2010. **6**(10).
48. Miller, A.D. and G.J. Rosman, *Improved retroviral vectors for gene transfer and expression.* BioTechniques, 1989. **7**(9): p. 980-2, 984-6, 989-90.
49. Narvaiza, I., et al., *Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase.* PLoS pathogens, 2009. **5**(5): p. e1000439.
50. Freeman, J.D., N.L. Goodchild, and D.L. Mager, *A modified indicator gene for selection of retrotransposition events in mammalian cells.* BioTechniques, 1994. **17**(1): p. 46, 48-9, 52.

Chapter 4

A3A_F75L, a Putative Separation-of-Function APOBEC3A Mutant

I designed and carried out the cultured cell retrotransposition assays and *in vitro* LEAP assays described in this chapter. Dr. Inigo Narvaiza, in the laboratory of Dr. Matt Weitzman, generated the purified recombinant A3A and mutant A3A proteins, and performed the *in vitro* deaminase assays shown in Figure 4.3 A. Dr. Rahul Kohli carried out the bacterial mutator assay alluded to in this chapter.

Abstract

In Chapter 3, I undertook a detailed mechanistic examination of APOBEC3A (A3A) mediated inhibition of LINE-1 retrotransposition. I elucidated a deaminase-dependent pathway of inhibition, in which A3A edits transiently-exposed single-stranded DNA during L1 target-primed reverse transcription (TPRT). I conclude that, in concert with the action of cellular repair factors uracil DNA glycosylase (UNG) and AP endonuclease (APE), this mechanism is at least partially responsible for A3A-mediated L1 inhibition. However, these data do not

rule out a deaminase-independent component of A3A-mediated L1 inhibition. In this chapter, I report L1 inhibition by a particular A3A mutant, A3A_F75L. This mutant has previously been demonstrated to lack deaminase activity *in vitro* [1]. Yet, in the experiments detailed in this chapter, I found that A3A_F75L potently inhibits L1 retrotransposition in cultured cells. Thus, I initially regarded A3A_F75L as a separation-of-function mutant which supports a deaminase-independent mechanism for A3A-mediated L1 inhibition. Unexpectedly, however, I found that A3A_F75L mediated L1 inhibition is alleviated by uracil glycosylase inhibitor (UGI) expression. Furthermore, A3A_F75L has above-background deaminase activity in a bacterial mutator assay. I hypothesize that this mutant retains deaminase activity *in vivo,* but not *in vitro.*

Introduction

The human APOBEC3 (A3) proteins contain conserved cytidine deaminase active sites with the consensus sequence H-X-E-X_{23-28}-P-C-X_{2-4}-C, where X represents any amino acid [2]. Based on comparisons with well-characterized zinc-dependent deaminases, the histidine (H) and two cysteine (C) residues are hypothesized to participate in Zn^{2+} ion coordination, while the critical glutamic acid (E) residue is proposed to function as a proton shuttle [3-6] (Figure 4.1 B). Mutation of these conserved residues abolishes cytidine deaminase activity [7-9]. However, A3A is a relatively small protein (199 amino acids; Figure 4.1 A), and disruption of Zn^{2+} ion coordination may severely disrupt protein folding and active site integrity. Thus, mutational analyses to distinguish whether

A3A-mediated inhibition of retroelements and viruses depends specifically on cytidine deaminase activity, or only requires an intact active site, have proved challenging.

In a recent study [1], Narvaiza et al. generated and characterized a panel of A3A mutant proteins, including a mutant termed A3A_F75L. Phenylalanine residue F75 resides within the A3A cytidine deaminase active site (Figure 4.1 A), but is not predicted to participate in Zn^{2+} ion coordination (Figure 4.1 B). Notably, the analogous residue in APOBEC1, which is an RNA editing enzyme, is required for deaminase activity as well as nucleic acid binding [4, 10]. Previous studies indicate that wild-type A3A restricts adeno-associated virus (AAV) replication, whereas A3A_C106S, which bears a mutation in a critical Zn^{2+} coordinating cysteine residue and lacks deaminase activity *in vitro* [7], does not affect AAV replication [1, 7]. A3A_F75L lacks deaminase activity *in vitro,* but retains the ability to inhibit AAV replication in cultured cells [1]. Thus, A3A_F75L putatively separates the capacity to restrict AAV replication from A3A cytidine deaminase activity, and represents a potential reagent to distinguish between deaminase-dependent and deaminase-independent A3A-mediated restriction of other retroelements and exogenous pathogens.

In Chapter 3, we identify a deaminase-dependent mechanism for A3A-mediated L1 inhibition, in which A3A deaminates single-stranded DNA transiently exposed during L1 TPRT. However, our results do not rule out the existence of a deaminase-independent mechanism of inhibition. Indeed, evidence of deaminase-independent L1 inhibition by APOBEC3 factors has been provided by

studies on APOBEC3B (A3B) [8, 11] and APOBEC3F (A3F) [11]. Furthermore, in Chapter 2, we report deaminase-independent inhibition of a zebrafish LINE-2 element [12] by APOBEC3G (A3G). Here, we employ A3A_F75L, a putative separation-of-function A3A mutant, in an attempt to disentangle deaminase-dependent and deaminase-independent modes of A3A-mediated L1 inhibition.

Results

A3A_F75L inhibits LINE element retrotransposition in cultured cells

To investigate a potential deaminase-independent mode of A3A-mediated L1 inhibition, we asked whether A3A_F75L inhibits human L1.3 [13, 14], as well as mouse natural and synthetic L1s (TGf21 [15] and L1SM [16]) and a zebrafish LINE-2 element (Zfl2-2 [12]), in the cell culture retrotransposition assay [17, 18]. We found that F75L potently inhibits retrotransposition of L1.3 (16% of control), TGf21 (16% of control), and Zfl2-2 (13% of control)(Figure 4.2). Inhibition of L1SM by F75L was less pronounced (51% of control). Across all elements tested, F75L effected slightly more potent inhibition than wild-type A3A. In contrast, C106S did not inhibit any of the elements tested, consistent with previous reports (Figure 4.2)[7].

A3A_F75L does not block L1 Reverse Transcription *in vitro*

APOBEC3G (A3G) is hypothesized to inhibit elongation of HIV reverse transcripts by acting as a physical “road-block” to RT procession on the HIV RNA template, in a deaminase-independent manner [19]. We therefore asked

whether wild-type A3A, A3A_F75L, and A3A_C106S block L1 reverse transcription. We employed an *in vitro* assay termed L1 element amplification protocol (LEAP)[20]. In the LEAP assay, L1 RNPs are provided with a single-stranded 3'RACE adapter, which consists of a 3' oligo dT and a unique adapter sequence (Figure 3.6 A). The oligo dT portion of the adapter anneals to the L1 RNA polyA tail, providing a 3'OH from which the ORF2p-encoded reverse transcriptase activity can initiate reverse transcription of its associated RNA. The resultant L1 cDNA is amplified by PCR using primers to the engineered L1 3' end and the unique adapter sequence.

We included purified recombinant A3A protein (rA3A) as well as A3A mutant proteins (rA3A_F75L, rA3A_C106S) in the LEAP reaction. Wild-type rA3A exhibits robust deaminase activity in an *in vitro* assay, while rA3A_F75L and rA3A_C106S exhibit minimal deaminase activity (Figure 4.3 A). We included these recombinant proteins in LEAP reactions, and observed no diminishment of LEAP products (Figure 4.3 B). Thus, we concluded that rA3A and rA3A mutant proteins do not block L1 RT activity.

Recombinant A3A_F75L protein does not efficiently deaminate LEAP products *in vitro*

We next examined LEAP products generated in the presence of rA3A, rA3A_F75L, and rA3A_C106S. As described in Chapter 3, rA3A efficiently deaminated the single-stranded adapter used to prime the L1 reverse transcription reaction, and when recombinant RNaseH was included in the LEAP

reaction, A3A efficiently deaminated the L1 cDNA. In contrast, rA3A_F75L, like rA3A_C106S, did not exhibit robust deaminase activity on LEAP substrates (Figure 4.4). Thus, recombinant A3A_F75L protein lacks robust deaminase activity *in vitro*.

F75L-mediated L1 inhibition is alleviated by UGI in cultured cells

Based on the pattern of deamination observed in the LEAP assay, we hypothesized that wild-type A3A inhibits L1 retrotransposition by deaminating transiently exposed single-stranded DNA during L1 TPRT. Deaminated TPRT intermediates would be degraded by the cellular repair factors uracil DNA glycosylase (UNG) and apurinic/apyrimidinic endonuclease (APE), thereby effecting L1 inhibition. We therefore reasoned that expression of uracil glycosylase inhibitor protein (UGI) [21] would alleviate A3A-mediated retrotransposition inhibition. Indeed, co-expression of UGI with A3A in the L1 retrotransposition alleviated L1 inhibition by ~2-fold over vector control (Figure 3.12 C, Figure 3.17 B). We also find evidence for A3A-mediated editing of L1 TPRT intermediates in cultured cells, in the presence of UGI (Figure 3.16 A and B, Figure 3.19). Unexpectedly, UGI expression also alleviated A3A_F75L-mediated retrotransposition inhibition to a similar extent (Figure 4.5). Notably, UGI expression alone does not increase retrotransposition efficiency (Figure 3.13). Furthermore, UGI expression does not alleviate deaminase-independent retrotransposition inhibition by APOBEC3B (A3B) mutants (Figure 3.14), nor does it alleviate deaminase-independent inhibition of a zebrafish LINE-2 element

(Zfl2-2) by APOBEC3G (A3G) (Figure 2.5). Thus, we hypothesize that A3A_F75L may possess deaminase activity *in vivo*, but not *in vitro.* Indeed, A3A_F75L exhibits cytidine deaminase activity above the level exhibited by A3A_C106S in a bacterial mutator assay (Matthew Weitzman and Rahul Kohli, personal communication).

Discussion

In this Chapter, we examine the impact of a putative A3A separation-of-function mutant, A3A_F75L, on L1 retrotransposition. A3A_F75L lacks deaminase activity *in vitro;* however, like wild-type A3A, A3A_F75L inhibits human, mouse, and zebrafish LINE elements in a cultured cell retrotransposition assay. Recombinant A3A_F75L protein has no effect on L1 reverse transcription in an *in vitro* assay, and does not exhibit robust deaminase activity on L1 RT products. However, like wild-type A3A, A3A_F75L mediated L1 inhibition is alleviated by UGI expression. Furthermore, A3A_F75L has greater deaminase activity than A3A_C106S in a bacterial mutator assay.

Taken together the above results suggest that A3A_F75L lacks deaminase activity in *in vitro* assays, yet retains deaminase activity when expressed in bacteria or cultured mammalian cells. We hypothesize that *in vivo,* unidentified cellular factors may facilitate proper A3A_F75L protein folding or stability (Figure 4.6). Therefore, when removed from the cellular context, A3A_F75L cannot efficiently deaminate single-stranded DNA.

Alternatively, it is possible that A3A_F75L truly lacks deaminase activity *in vitro* and *in vivo*. For this to be the case, UGI expression must alleviate A3A_F75L mediated retrotransposition inhibition by a mechanism that does not involve the stabilization of uracil residues in DNA. Notably, UGI blocks the active site of UNG by mimicking protein-DNA interactions [22, 23]. Therefore, one could hypothesize that UGI also blocks the A3A active site by mimicking a cytidine residue, thereby preventing the A3A active site from functioning in a deaminase-independent manner, but giving the appearance of alleviating the downstream effects of deamination. This scenario is unlikely, however, in light of our finding that UGI does not alleviate L1 inhibition by deaminase-defective A3B mutants (Figure 3.14) or inhibition of Zfl2-2 by a deaminase-defective A3G mutant (Figure 2.5). Furthermore, inclusion of recombinant UGI in the LEAP reaction with wild-type rA3A does not abolish A3A-mediated deamination of LEAP products (Figure 3.5). In sum, we find that A3A_F75L lacks deaminase activity *in vitro*, but retains deaminase activity *in vivo.* The molecular explanation for this difference remains to be uncovered. It is possible that A3A_F75L has weaker enzymatic activity than wild-type A3A, and that this deficiency is reflected in *in vitro* experiments as a lack of detectable activity. However, in the cellular context, A3A_F75L may be able to localize efficiently to TPRT intermediates (or exposed single-stranded plasmid DNA in a bacterial mutator assay). Thus, in cultured cells, any residual A3A_F75L deaminase activity may be sufficient to inhibit L1 retrotransposition in a deaminase-dependent manner. Alternatively, A3A_F75L, in mammalian cells, may interact with an endogenously expressed

A3 factor (such as A3B), and enhance its ability to act at L1 TPRT in a deaminase-dependent manner. Overall, whether L1 inhibition by wild-type A3A involves a deaminase-independent component remains to be elucidated.

Materials and Methods

Some of the text describing materials and methods in this chapter is duplicated from Chapter 2 and Chapter 3 of this thesis.

Plasmids

All plasmids were grown in DH5α (F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1) competent *E. coli* (Invitrogen; Carlsbad, CA. Prepared in house as described in [24]). Plasmids were prepared using the Qiagen Plasmid Midi Kit (QIAGEN; Hilden, Germany) according to the manufacturer's protocol.

APOBEC3 expression constructs:

The pK_βarr and A3A expression plasmids have been described previously [9]. The APOBEC3A mutant C106S, described in [7], and the APOBEC3A mutant F75L, described in [1] were received from Dr. Matt Weitzman and Dr. Inigo Narvaiza in pcDNA 3.1^+ (Invitrogen), and were subcloned into the pK vector using *Hind*III and *Xho*I restriction sites.

Bacterial mutator assay constructs:

I PCR amplified A3A, A3A_F75L, and A3A_C106S cDNAs from the respective pK expression vectors using the following primers:

5'_NdeI_A3A: GGGGCATATGATGGAAGCCAGCCCAGCATC

3'_XhoI_A3A: GCGATCCTCGAGTCAAGCGTAATC

The resultant fragments were cloned into the pET41 expression vector (Novagen) using NdeI and XhoI restriction endonucleases (NEB). pET41_A3A, pET41_F75L, and pET41_C106S were then used by Dr. Rahul Kohli and Dr. Matt Weitzman to carry out the bacterial mutator assay (personal communication).

LINE expression constructs:

pJM101/L1.3 [13, 14], pCEP4/L1SM [16], pCEP4/TGf21 [15], pCEP4/Zfl2-2 [12], LGCX vector [21], LGCX/UGI [21], and pU6i NEO are described in depth in Chapters 2 and 3 of this thesis.

Cell Culture

Cell culture conditions are described in Chapters 2 and 3 of this thesis.

L1 Retrotransposition Assays

Retrotransposition assay methods [17, 18] are described in detail in Chapters 2 and 3 of this thesis.

Production and characterization of Recombinant A3A Proteins

Purification of recombinant A3A and mutant proteins was carried out as described in Chapter 3. *In vitro* deaminase assays were carried out as described in Chapter 3.

LEAP Assay

Methods for RNP preparation and LEAP assays [20] are described in detail in Chapters 2 and 3 of this thesis.

Acknowledgements

I would like to thank members of the Moran lab for helpful discussion. Dr. Inigo Narvaiza and Dr. Matthew Weitzman generated recombinant A3A proteins. Dr. Rahul Kohli performed the bacterial mutator assay.

A.

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQ

AKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQE

NTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQ

PWDGLDEHSQALSGRLRAILQNQGN*

HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC

B.

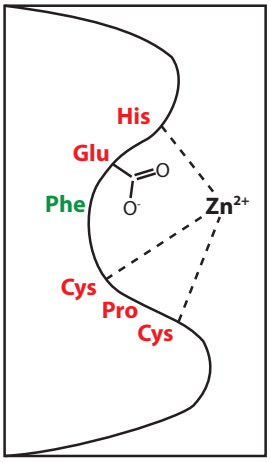


Figure 4.1: The cytidine deaminase active site of APOBEC3A

A. The amino acid sequence of APOBEC3A. The conserved cytidine deaminase active site is underlined within the amino acid sequence, and expanded below. The critical active site residues highlighted in red (H70, E72, P100, C101, C106) [25]. F75 is highlighted in green.

B. Schematic of the A3A active site. H70, C101, and C106 (highlighted in red) participate in zinc ion coordination; E72 is important in catalyzing the removal of the amine group to convert cytidine to uridine (Reviewed in [26] and [2]). F75 is highlighted in green. This figure is re-drawn based on a figure from a review by Harris and Liddament [26].

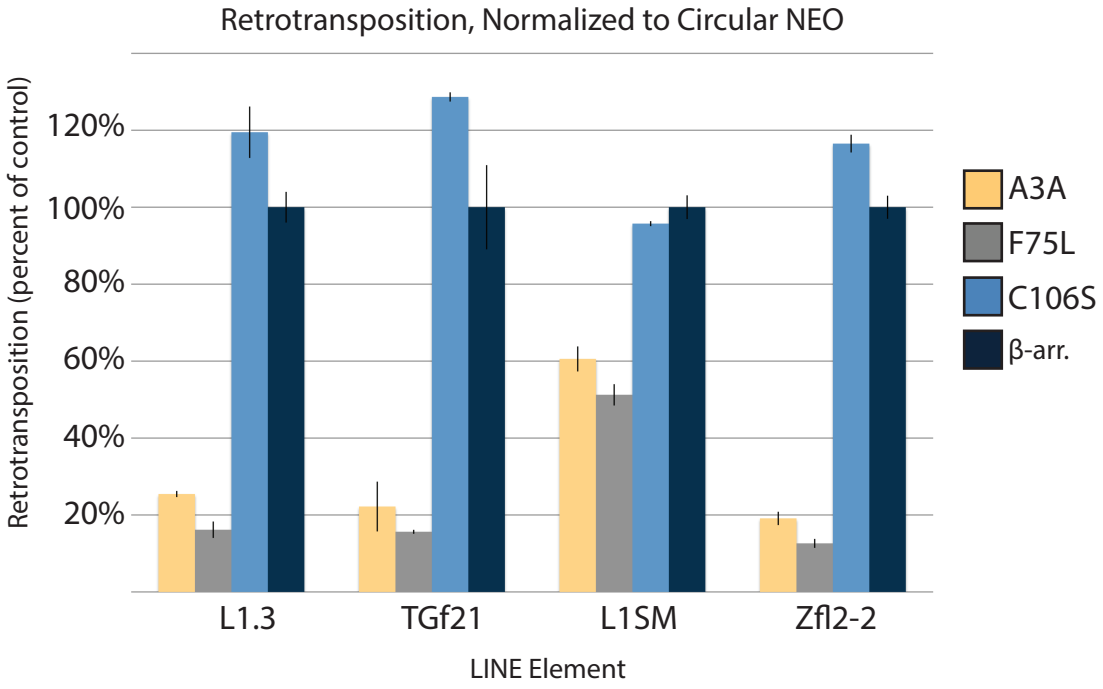


Figure 4.2: A3A_F75L inhibits LINE element retrotransposition

Approximately $2x10^4$ (L1SM), $1x10^5$ (L1.3), $2x10^5$ (NEO control), or $4x10^5$ (TGf21 and Zfl2-2) HeLa cells were plated per T-75 flask and co-transfected in duplicate with 2 µg each of LINE and A3A (or β-arr control) expression vector. The y-axis indicates percent retrotransposition in the presence of each A3A protein, with the β-arr positive control set to 100% for each element. The x-axis indicates the LINE element used. Maize bars show results for wild-type A3A, gray bars show results for A3A_F75L, light blue bars show results for A3A_C106S, and dark blue bars show results for the β-arr positive control. Error bars indicate percent standard deviation derived from duplicate transfections. Data was normalized using the circular NEO control. Figure 3.1 depicts data from the same experiment, omitting A3A_F75L. A representative experiment is shown. This assay was performed twice, representing biological replicates.

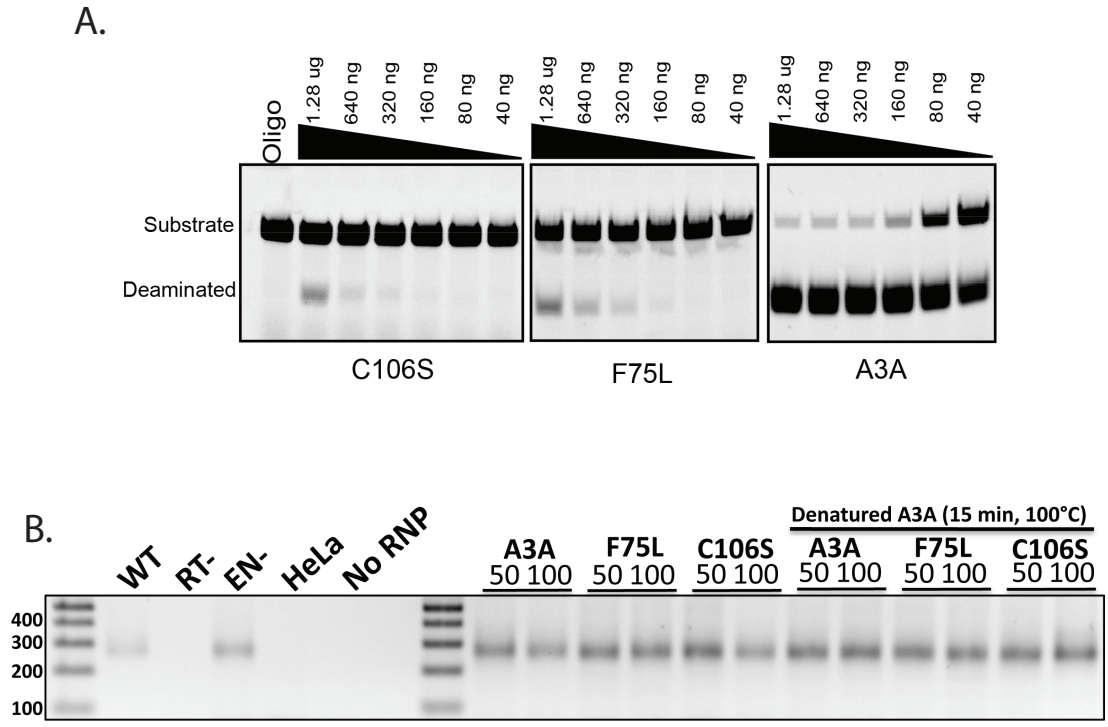


Figure 4.3: Recombinant A3A_F75L protein does not inhibit LEAP activity

A. In vitro *deaminase activity of rA3A and rA3A mutant proteins.* Increasing amounts of rA3A (left panel), rF75L (middle panel) or rC106S (right panel) were incubated with a FITC-labeled single strand DNA oligonucleotide containing a single cytidine residue for 4 hrs at 37° C. Samples were then incubated with recombinant uracil DNA glycosylase (UDG) and NaOH, and the products were resolved by gel electrophoresis using a 15% TBE/urea polyacrylamide gel. Molecular sizes of substrate oligonucleotide and deaminated, cleaved product are indicated. A control reaction (Oligo) was performed in the absence of recombinant protein.

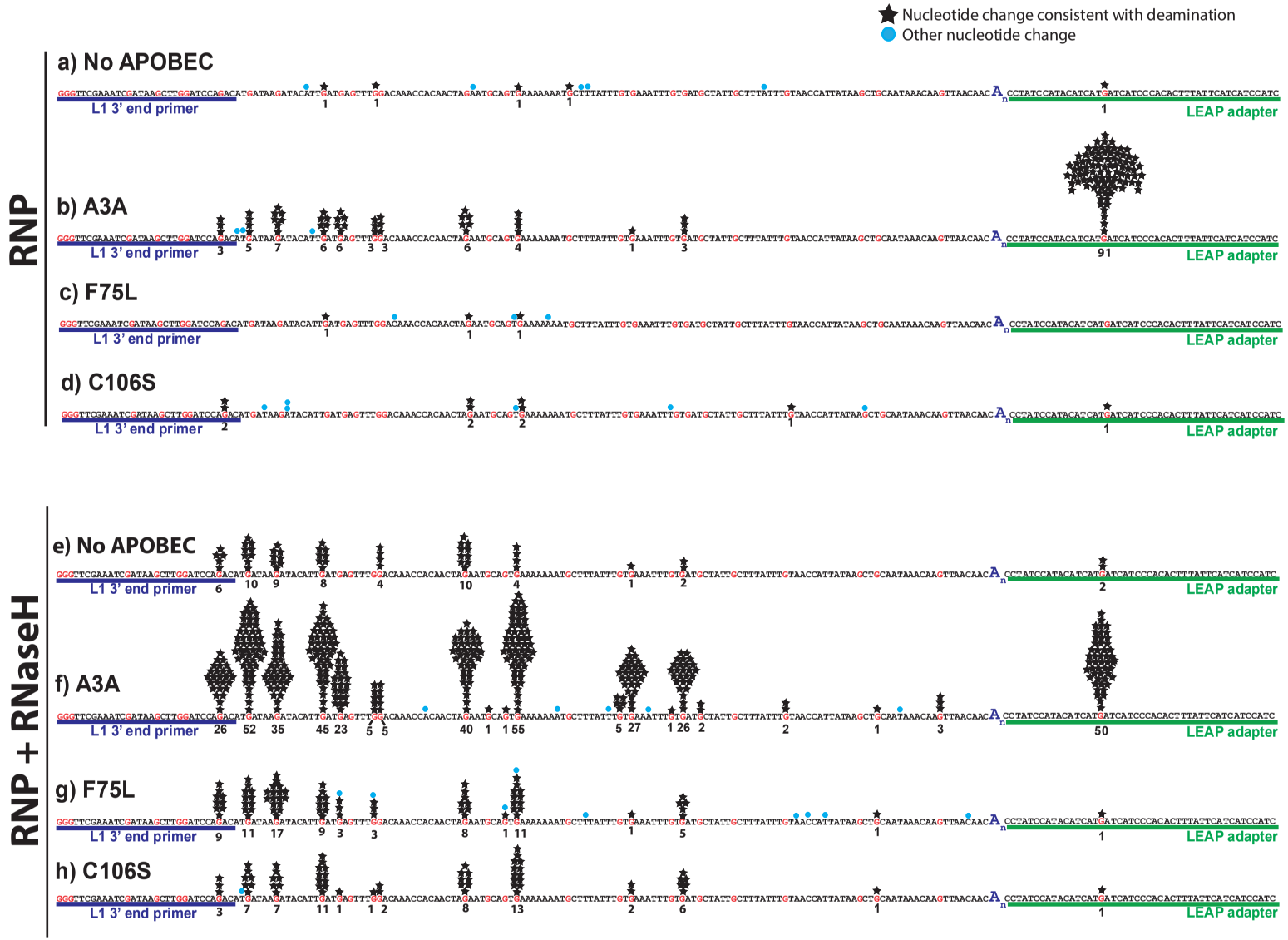
B. Recombinant A3A_F75L protein does not inhibit LEAP activity. Left: controls, left to right: Wild-type RNP (pDK101), RT mutant RNP (pDK135), EN mutant RNP (pDK136), untransfected HeLa cell RNP, no RNP control. *Right:* 50 ng and 100 ng of rA3A, rA3A_F75L, and rA3A_C106S were included in LEAP reactions with wild-type (pDK101) RNPs. As a control, 50 ng and 100 ng of "heat-killed" (15 minutes, 100° C) rA3A and mutant proteins were added to LEAP reactions.

Figure 4.4: rA3A_F75L does not have robust deaminase activity

For each experimental condition, the plus-strand sequence of the unedited LEAP product is depicted, 5' to 3. Therefore, C-to-T changes indicative of editing by A3A on the minus-strand cDNA are revealed as G-to-A changes on the depicted plus strand. The location of the L1 3' end PCR primer is depicted at far left, underlined in blue. At right, the location of the poly-A tail is indicated (A_n), followed by the 5np1 LEAP adapter sequence, underlined in green. Within the LEAP product sequence, all G nucleotides are highlighted in red. Black stars piled above G nucleotides graphically represent the number times among 100 LEAP products that a nucleotide change consistent with deamination was observed at that position. Counts of deamination events are represented numerically below. Blue circles piled above the sequence represent nucleotide changes not consistent with deamination. Data from the same experiment for A3A, A3A_C106S, and β-arr are also presented in Figure 3.10.

Top panel, a-d: Summary of 100 LEAP products generated using wild-type L1 RNP plus: a) no rA3A protein, b) 100 ng of wild-type rA3A, c) 100 ng of rA3A_F75L, d) 100 ng of rA3A_C106S.

Bottom panel, e-h: Summary of 100 LEAP products generated using wild-type L1 RNP with 2 Units of RNaseH, plus: e) no rA3A protein, f) 100 ng of wild-type rA3A, g) 100 ng of rA3A_F75L, d) 100 ng of rA3A_C106S.


Nucleotide change consistent with deamination
Other nucleotide change
RNP
a) No APOBEC
b) A3A
c) F75L
d) C106S
RNP + RNaseH
e) No APOBEC
f) A3A
g) F75L
h) C106S
L1 3' end primer
LEAP adapter

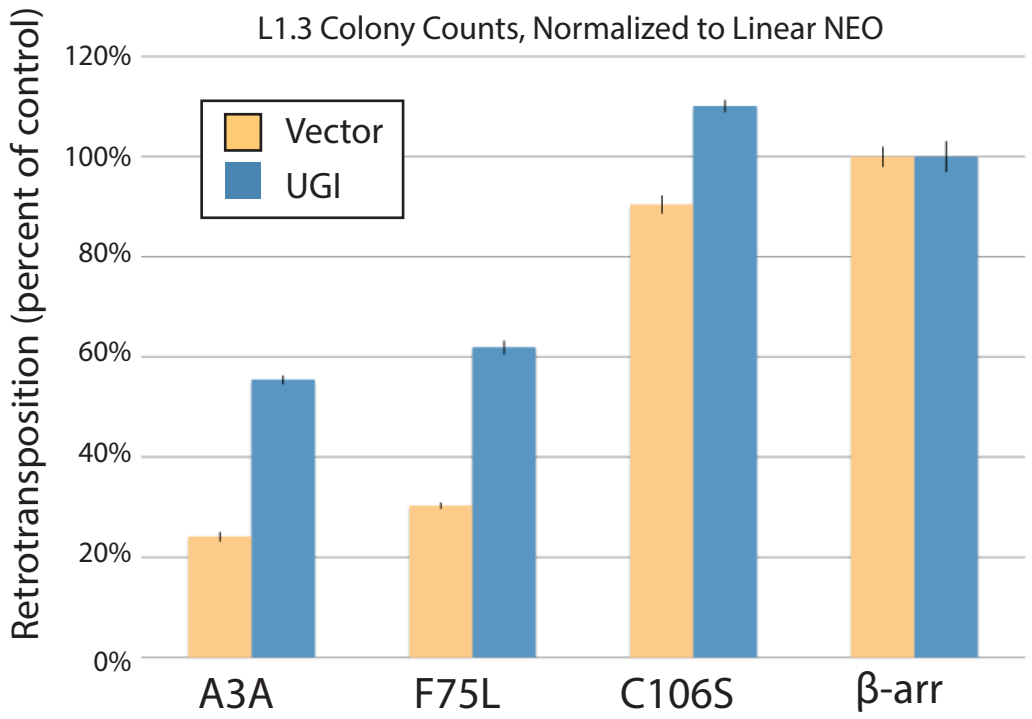


Figure 4.5: UGI alleviates A3A_F75L-mediated retrotransposition inhibition

Approximately $1x10^5$ HeLa cells were plated per 10 cm dish and co-transfected with 1 µg each of L1 expression plasmid (JM101_L1.3) A3A or β-arr control vector, and pLGCX_UGI or pLGCX empty vector. The y-axis depicts % retrotransposition. For each experimental condition, the appropriate β-arr control was set to 100%. Gold bars indicate empty pLGCX vector; blue bars indicate pLGCX_UGI. Data was normalized to circular NEO control co-transfections. Error bars represent percent standard deviation between duplicate transfections. Data from the same experiment for A3A, A3A_C106S, and β-arr are also presented in Figure 3.12.

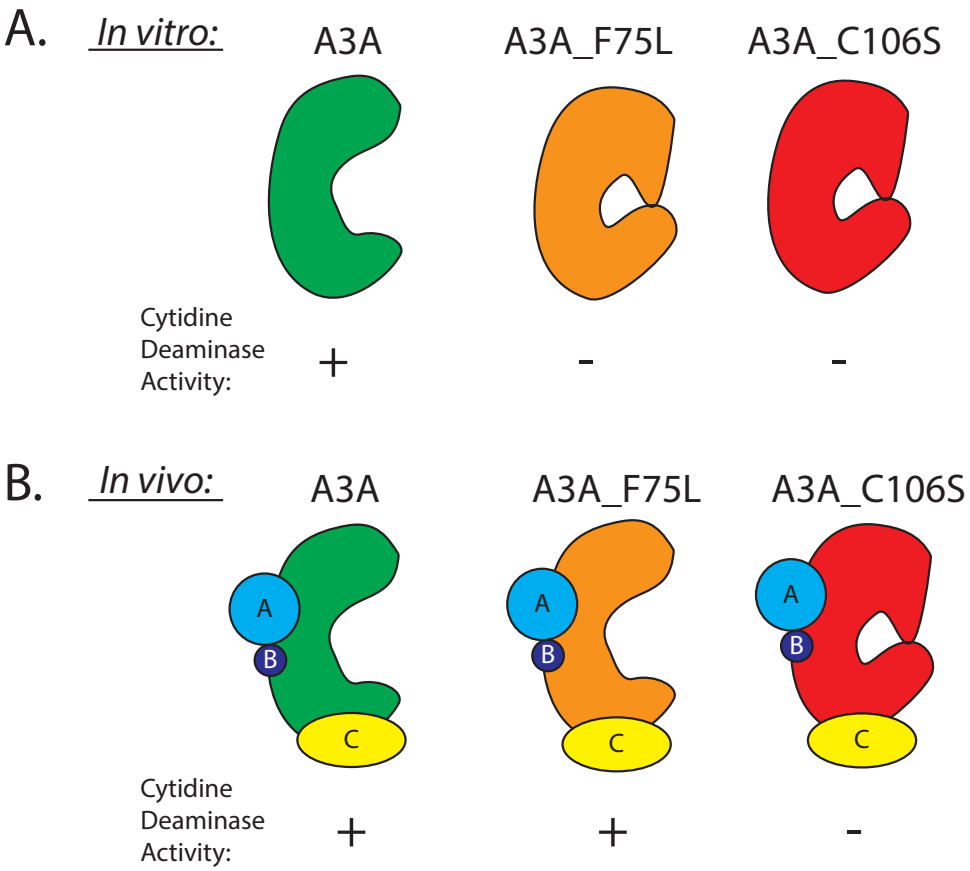


Figure 4.6: Speculative model for differential deaminase activity of A3A_F75L *in vitro* and *in vivo*

A. In vitro, wild-type A3A (green) retains deaminase activity. F75L (orange) is deficient in deaminase activity *in vitro,* possibly due to defects in protein folding or stability. C106S (red) is deficient in deaminase activity *in vitro*, due to disruption of Zn^{2+} ion coordination critical for deaminase activity.

B. In vivo, wild-type A3A (green) has deaminase activity and may interact with unknown cellular factors (A, B, C). A3A_F75L (orange) has some deaminase activity *in vivo;* stability or proper folding may be enhanced by cellular interacting factors. C106S (red) is deficient in deaminase activity *in vivo,* due to disruption of Zn^{2+} ion coordination critical for deaminase activity.

References

1. Narvaiza, I., et al., *Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase.* PLoS Pathog, 2009. **5**(5): p. e1000439.
2. Chiu, Y.L. and W.C. Greene, *The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements.* Annu Rev Immunol, 2008. **26**: p. 317-53.
3. Betts, L., et al., *Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex.* Journal of molecular biology, 1994. **235**(2): p. 635-56.
4. Navaratnam, N., et al., *Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site.* Cell, 1995. **81**(2): p. 187-95.
5. Chen, K.M., et al., *Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G.* Nature, 2008. **452**(7183): p. 116-9.
6. Prochnow, C., et al., *The APOBEC-2 crystal structure and functional implications for the deaminase AID.* Nature, 2007. **445**(7126): p. 447-51.
7. Chen, H., et al., *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.* Curr Biol, 2006. **16**(5): p. 480-5.
8. Bogerd, H.P., et al., *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8780-5.
9. Bogerd, H.P., et al., *APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.* Nucleic Acids Res, 2006. **34**(1): p. 89-95.
10. MacGinnitie, A.J., S. Anant, and N.O. Davidson, *Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity.* J Biol Chem, 1995. **270**(24): p. 14768-75.
11. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.* J Biol Chem, 2006. **281**(25): p. 16837-41.
12. Sugano, T., M. Kajikawa, and N. Okada, *Isolation and characterization of retrotransposition-competent LINEs from zebrafish.* Gene, 2006. **365**: p. 74-82.
13. Sassaman, D.M., et al., *Many human L1 elements are capable of retrotransposition.* Nat Genet, 1997. **16**(1): p. 37-43.
14. Dombroski, B.A., A.F. Scott, and H.H. Kazazian, Jr., *Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element.* Proc Natl Acad Sci U S A, 1993. **90**(14): p. 6513-7.
15. Goodier, J.L., et al., *A novel active L1 retrotransposon subfamily in the mouse.* Genome Res, 2001. **11**(10): p. 1677-85.
16. Han, J.S. and J.D. Boeke, *A highly active synthetic mammalian retrotransposon.* Nature, 2004. **429**(6989): p. 314-8.

17. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
18. Wei, W., et al., *A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events.* Anal Biochem, 2000. **284**(2): p. 435-8.
19. Bishop, K.N., et al., *APOBEC3G inhibits elongation of HIV-1 reverse transcripts.* PLoS Pathog, 2008. **4**(12): p. e1000231.
20. Kulpa, D.A. and J.V. Moran, *Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles.* Nat Struct Mol Biol, 2006. **13**(7): p. 655-60.
21. Kaiser, S.M. and M. Emerman, *Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G.* J Virol, 2006. **80**(2): p. 875-82.
22. Wang, Z. and D.W. Mosbaugh, *Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase.* The Journal of biological chemistry, 1989. **264**(2): p. 1163-71.
23. Savva, R. and L.H. Pearl, *Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex.* Nature structural biology, 1995. **2**(9): p. 752-7.
24. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-8.
25. Jarmuz, A., et al., *An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.* Genomics, 2002. **79**(3): p. 285-96.
26. Harris, R.S. and M.T. Liddament, *Retroviral restriction by APOBEC proteins.* Nat Rev Immunol, 2004. **4**(11): p. 868-77.

Chapter 5

Conclusions

Overview

In this thesis, I have undertaken a mechanistic examination of retroelement inhibition by the human APOBEC3 (A3) proteins. In Chapter 1, I provide a literature review discussing our current knowledge of retrotransposon control mechanisms in mammals, with a focus on the regulation of potentially heritable insertions in the mammalian germline and early embryo. Chapter 2 entails an initial survey of human A3 activity against autonomous LINE and non-autonomous SINE elements from various species. In Chapter 3, I elucidate a deaminase-dependent mechanism for APOBEC3A (A3A) mediated L1 inhibition, while in Chapter 4, I investigate putative deaminase-independent L1 inhibition by an A3A mutant, A3A_F75L. In this concluding Chapter, I reflect on the body of work presented in this thesis. I discuss the significance of results acquired in previous chapters, but focus on remaining questions, and provide suggestions for future experimental directions.

APOBEC3-Mediated Retrotransposition Inhibition in the Cultured Cell Assay: Reflections on our Experimental Approach

Why study L1 and A3A in HeLa Cells?

When considering the physiological context of L1 restriction by APOBEC3 proteins, it is important to note that endogenous A3A expression in humans is limited to peripheral blood lymphocytes [1-6], making it unlikely that A3A combats heritable retrotransposition events *in vivo*. However, I view A3A as a surrogate for investigating a general mechanism of deaminase-dependent retroelement restriction, and from an experimental standpoint, studying A3A-mediated L1 inhibition in cultured human HeLa cells has distinct advantages. First, HeLa cells do not highly express endogenous full-length L1 RNA [7] or L1 ORF1p [7, 8], yet they accommodate high levels of L1 retrotransposition from ectopically expressed retrotransposition indicator plasmids [9, 10]. Similarly, HeLa cells do not express endogenous A3A [11]. Thus, A3A-mediated L1 inhibition can be studied in these cells in a controlled manner, with minimal influence from endogenous L1 or A3A expression. In contrast, HeLa cells express endogenous A3B, and siRNA-mediated A3B knockdown increases engineered L1 retrotransposition in these cells [11]. Thus, a mechanistic examination of A3B-mediated L1 restriction in HeLa cells may be confounded by endogenous A3B expression, a complication which we avoid when studying A3A. Insights gained by studying A3A, however, can now be applied to examining the mechanism of L1 restriction by A3B (discussed further below).

From an evolutionary standpoint, it is also interesting to note that the ancestral AID/APOBEC-like protein, from which the modern-day AID/APOBEC family expanded, was almost certainly a single deaminase domain protein (as is A3A) which arose early in vertebrate evolution [12]. The closest extant relative to this AID/APOBEC predecessor is found in the sea lamprey, a jawless vertebrate, where it has been implicated in a form of adaptive immunity involving variable lymphocyte receptors [13]. I therefore speculate that ancestral AID/APOBEC proteins acquired a role in innate immunity against endogenous mobile elements by a deaminase-dependent mechanism similar to the mechanism we have uncovered for A3A-mediated L1 inhibition. I regard this deaminase-dependent restriction as a relatively "primitive" defense against transposable elements. Indeed, as illustrated by the cytotoxic effects of ectopic A3A expression in cultured cells (discussed below) [14], deaminase-dependent modes of transposable element restriction likely take a toll on genome stability. During the expansion and evolution of the mammalian APOBEC3 family, a subset of A3 genes fortuitously acquired a double deaminase domain structure [15]. I speculate that this double-domain configuration provided flexibility in terms of protein function and specificity, and that deaminase-independent retroelement and viral restriction represents a relatively recent adaptation of the AID/APOBEC family.

A3A Cytotoxicity in the Cultured Cell Retrotransposition Assay

It is now widely accepted that A3A expression has “off-target” effects on cell viability that are independent of L1 retrotransposition inhibition [2, 14, 16]. When I first chose to study the mechanism of A3A-mediated L1 inhibition, however, such effects had not been reported. In fact, several published studies included control experiments to rule out nonspecific cytotoxicity as a cause for apparent A3A-mediated L1 inhibition, and reported that A3A expression has no off-target effects [17, 18]. Upon initiating the APOBEC3 project, I performed GFP co-transfections to assess transfection efficiency in parallel with each retrotransposition assay, and I consistently observed fewer GFP positive cells in the presence of A3A as compared to control (β-arrestin) co-transfections. I became concerned that A3A expression was killing or inhibiting the growth of transfected cells, and that such cytotoxicity could be responsible for the reduction in G418-resistant colonies we observe in the retrotransposition assay in the presence of APOBEC3 expression.

Transfection efficiency is assessed at 72 hours post-transfection, while the retrotransposition assay is carried out for 12-14 days. I therefore developed a more stringent control for A3A-mediated off-target effects, by carrying out L1-A3 retrotransposition assays in parallel with circular or linearized NEO-A3 co-transfections (Figure 3.2 A). Like the L1-A3 retrotransposition assays, G418 selection on NEO-A3 co-transfections was initiated at 72 hours post-transfection and carried out for 12-14 days. Despite the increased stringency of this control, A3A demonstrates significant L1 inhibition after normalization of results to NEO co-transfections (Figure 3.2 B and C).

In retrospect, having found that A3A deaminates single-stranded DNA in L1 TPRT intermediates, I propose that A3A-mediated toxicity likewise stems from opportunistic deamination of single-stranded DNA. I hypothesize that A3A can access single-stranded genomic DNA transiently exposed at structures such as transcription bubbles and replication forks, leading to a DNA damage response and cell cycle perturbation [14]. Indeed, evidence of A3A-mediated editing of genomic and mitochondrial DNA has been uncovered [19], using a technique known as differential DNA denaturation PCR (3D-PCR) [20]. The 3D-PCR technique is a nested PCR strategy that entails one round of standard PCR using primers to a genomic or mitochondrial sequence, followed by a second round of PCR using a nested set of primers. Several PCR reactions are done in parallel, with denaturation temperatures reduced in a stepwise manner [20]. Lower denaturation temperatures allow G-to-A edited products (*i.e.*, products with a lower GC content) to denature, while the parental non-edited products (higher GC content) remain annealed and are not amplified. The 3D-PCR technique has also been employed to uncover evidence for A3A-mediated editing of transfected plasmid DNA [2]. Although this finding raises concerns that the editing we observe within genomic L1 insertions in the presence of A3A could have resulted from deamination of plasmid DNA, we observe a complete strand bias for editing of the minus-strand L1 cDNA, while editing of plasmid DNA is reported to occur on both strands [2]. If A3A-mediated editing was occurring on the autonomously-replicating episome (JM140/L1.3/Δ^2/k7; [21]) from which we generate L1

insertions, we would expect our L1 insertions to reflect editing of both strands of plasmid DNA.

Taken together, A3A-mediated editing of genomic DNA and plasmid DNA may account for the reduction in G418-resistant colonies observed in NEO co-transfection control assays (Figure 5.1). Normalization of L1 retrotransposition assays to NEO control co-transfections, however, reveals marked restriction of L1 retrotransposition that cannot be explained by off-target A3A-mediated editing alone.

APOBEC3-Mediated Retrotransposition Inhibition: Remaining Questions and Future Experimental Directions

Why is L1SM resistant to A3A-mediated inhibition?

We consistently observe that the synthetic mouse element, L1SM [22], is partially refractory to A3A-mediated inhibition, compared to human L1.3 and natural mouse [23] and zebrafish [24] elements (Figure 2.3 and Figure 3.1). We hypothesize that L1SM generates more insertions per cell, on average, than other elements. Indeed, preliminary data generated in the lab by a previous rotation student, Sean Ferris, suggests that this is the case. The model presented in Chapter 3 for A3A-mediated L1 inhibition suggests that inhibition occurs at the level of the individual insertion. As one insertion per cell is sufficient to confer G418-resistance in the retrotransposition assay, A3A expression would therefore have a stronger effect on elements which produce

fewer insertions per cell, while elements that generate many insertions per cell would be less potently inhibited. To test this hypothesis, one could generate a panel of clonal cell lines using L1.3 and L1SM, in the presence and absence of A3A (Figure 5.2 A). Southern blotting using a probe to the *mneoI* reporter gene would reveal the number of insertions per clonal cell line, and therefore per cell, for each condition. I predict that for L1.3 as well as L1SM, A3A would reduce the mean number of insertions per cell (Figure 5.2 B). However, such a shift for L1.3 would result in zero insertions in most cells (and hence, a loss of G418-resistant colonies), while for L1SM, more cells would ultimately harbor at least one insertion.

The ability of L1SM to partially escape A3A-mediated inhibition also provides an indirect argument that the reduction in colonies in the cultured-cell assay in the presence of A3A is not solely due to an A3A-triggered DNA damage response and resulting cell-cycle arrest [14], or degradation of plasmid DNA [2]. If A3A caused a loss of G418-resistant colonies entirely independently of L1 retrotransposition, then all elements should appear to be inhibited to the same extent.

How does APOBEC3G inhibit Zfl2-2 retrotransposition?

Perhaps the most unexpected result from our initial survey of A3-mediated inhibition of LINE elements is that the zebrafish LINE-2 element, Zfl2-2, is potently restricted by A3G (Figure 2.3). This is in contrast to human and mouse LINE-1 elements, which are inhibited by A3A and A3B, but not A3G. The data

presented in Chapter 3 indicate that A3A inhibits L1 retrotransposition by deaminating L1 TPRT intermediates, which is likely the same mechanism by which A3A inhibits mouse and zebrafish LINE elements. However, A3G is a cytoplasmic protein which presumably does not have access to TPRT, and is therefore unlikely to function like A3A to restrict Zfl2-2 mobilization. Comparisons to other elements which are inhibited by A3G may therefore provide more relevant insight about how A3G restricts Zfl2-2.

Zfl2-2, like the mouse and human LINE-1 elements, is an autonomous retrotransposon which encodes the enzymatic machinery for its own mobilization [24]. Unlike human and mouse L1s, however, Zfl2-2 lacks an ORF1p equivalent, and encodes only one ORF, analogous to ORF2p [24]. In this way, Zfl2-2 is similar to nonautonomous retroelements like human Alu and the mouse SINEs B1 and B2, which do not have protein-coding capacity, and which do not require L1 ORF1p for their mobilization [25-27]. A3G inhibits Alu retrotransposition [28], as well as B1 and B2 mobility (Figure 2.4). Thus, we observe a correlation wherein elements that encode ORF1p are refractory to A3G-mediated inhibition, while those lacking ORF1p are susceptible.

A3G-mediated Alu inhibition occurs independently of A3G cytidine deaminase activity [28]. Likewise, we find that A3G inhibits Zfl2-2 in a deaminase-independent manner (Figure 2.5). The prevailing model for Alu restriction by A3G is based on the observation that A3G expression leads to the inclusion of Alu RNAs in high molecular mass (HMM) A3G complexes [29]. This sequestration is proposed to inhibit Alu retrotransposition by preventing Alu

RNAs from accessing the L1 protein machinery. Notably, the data supporting this model do not distinguish correlation from causation: it is not apparent whether A3G-dependent inclusion of Alu RNA in HMM complexes is directly and solely responsible for A3G-mediated retrotransposition inhibition. Nevertheless, based on a sequestration hypothesis, one may speculate that L1 elements are protected from sequestration by A3G by the presence of ORF1p in RNP intermediates, whereas ORF1p-lacking elements are vulnerable to sequestration by A3G (Figure 2.7). As a first step to investigate this hypothesis, one could co-transfect cells with HA-tagged A3G and Zfl2-2, and ask whether Zfl2-2 mRNA is found in HMM A3G complexes, with Alu/A3G co-transfections serving as a positive control. There is currently no antibody available for the Zfl2-2 encoded protein; however, generation of a functional epitope-tagged (*i.e.*, TAP, HA, *etc.*) Zfl2-2 construct would allow detection of Zfl2-2 protein in HMM A3G complexes, as well. I would also employ tagged L1 constructs (*i.e.*, pAD3TE1, which bears a T7 tag on ORF1p, a TAP tag on ORF2p, and an MS2 tag on the L1 RNA [30]), to ask whether L1 RNA and protein are recruited to A3G HMM complexes. Co-localization of L1 and Zfl2-2 protein and RNA with A3G could also be examined by immunofluorescence microscopy. Inclusion of Zfl2-2 RNA and/or protein, but not L1 RNA and proteins, in A3G HMM complexes would strengthen the hypothesis that A3G restricts certain retroelements by a sequestration mechanism. However, if the RNA and protein components of L1, which is not inhibited by A3G, are recruited to HMM A3G complexes, it would suggest that

retroelement interaction with A3G HMM complexes is not indicative of susceptibility to restriction.

What is the fate of the L1 RNA following first-strand cDNA synthesis?

In Chapter 3, we found that in the *in vitro* LEAP assay, the L1 RNA and first-strand cDNA remain annealed in an RNA-DNA heteroduplex following first-strand L1 cDNA synthesis. The persistence of the L1 RNA protects the L1 cDNA from deamination by purified recombinant A3A protein (rA3A); the addition of recombinant RNaseH is required to render the L1 cDNA vulnerable to editing (Figure 3.10). This result led us to hypothesize that A3A effects L1 inhibition by deaminating single-stranded DNA substrates transiently exposed during TPRT. In a broader sense, this result also invites speculation about the fate of the L1 RNA during TPRT in cells. L1 does not encode an RNaseH activity [31], suggesting that the L1 RNA is either degraded by a cellular factor, or displaced during second-strand cDNA synthesis, as has been demonstrated *in vitro* for the related R2 element of *Bombyx mori* [32].

Our analysis of L1 insertions generated in the presence of A3A and UGI, where we find evidence of A3A-mediated deamination within the retrotransposed cDNA sequence, supports the hypothesis that a cellular factor may remove the L1 RNA prior to second-strand cDNA synthesis. Cellular RNaseH2, which cooperates with flap endonuclease (FEN-1) to remove RNA primers during lagging strand DNA replication [33], may be responsible for degradation of the L1 RNA. Unpublished data from the laboratory of Dr. Dan Stetson reveals that L1

retrotransposition efficiency can be up- and down-modulated by manipulating RNaseH2 expression levels: shRNA-mediated knockdown of RNaseH2 enhances L1 retrotransposition, while overexpression of RNaseH2 decreases L1 retrotransposition in the cultured cell assay (Dr. Dan Stetson, personal communication). Thus, I hypothesize that the fate of the L1 RNA varies on a cell-by-cell and perhaps insertion-by-insertion basis, depending on the efficiency of RNaseH2 activity. If cellular RNaseH2 gains access to the L1 RNA-cDNA hybrid before second-strand L1 cDNA synthesis takes place, the L1 RNA is removed, rendering the L1 cDNA vulnerable to cellular factors that may degrade single-stranded DNA and thereby prevent completion of a new L1 insertion. Alternately, if RNaseH2 does not access the TPRT intermediate, the first-strand cDNA remains protected until the putative RNA-displacement activity of the L1 enzymatic machinery displaces the L1 RNA upon second-strand cDNA synthesis.

In Chapter 3, we successfully employed UGI to preserve evidence of A3A-mediated editing of retrotransposed L1 first-strand cDNAs. I now propose that A3A-mediated editing provides a molecular signature of single-stranded DNA during TPRT, that can be used to discern whether RNaseH2 is responsible for degradation of the L1 cDNA. The reagents required for this experiment are a panel of HeLa cell lines which stably express UGI, and also stably over-express RNaseH2 (HeLa.UGI.RNaseH2+), express normal levels of RNaseH2 (HeLa.UGI.control) or contain a stable shRNA-mediated RNaseH2 knockdown (HeLa.UGI.RNaseH2-) (Figure 5.3). To confirm that any phenotype arising from

RNaseH2 knockdown can be reversed by RNaseH2 expression, I also propose to generate a stable cell line harboring an RNaseH2 shRNA, as well as an engineered shRNA-resistant RNaseH2 expression vector (HeLa.UGI.RH2.rescue)(Figure 5.3). On each of these backgrounds, I propose to generate L1 insertions in the presence of ectopic A3A expression, and analyze the retrotransposed sequences for evidence of A3A-mediated editing. The traditional L1 recovery assay used in Chapter 3 is labor and time-intensive; therefore, PCR amplification of the retrotransposed NEO reporter gene from a pool of insertion-harboring colonies, followed by next-generation sequencing, may provide a high-throughput and more efficient way to detect editing. If RNaseH2 is responsible for the removal of the L1 RNA, I would expect robust editing of HeLa.UGI.RNaseH2+ insertions, moderate editing of HeLa.UGI.control insertions (comparable to levels seen in Figure 3.19), and reduced editing of HeLa.UGI.RNaseH- insertions (Figure 5.3 B). HeLa.UGI.RH2.rescue insertions would be expected to harbor similar levels of deamination to HeLa.UGI.control insertions, provided that the engineered shRNA-resistant RNaseH2 expression vector generates similar RNaseH2 levels to the endogenous RNaseH2 gene.

It is also possible that RNaseH2 is not directly responsible for degradation of the L1 RNA, and that the L1 RNA-cDNA heteroduplex normally persists until second-strand cDNA synthesis. In this scenario, A3A may specifically recruit cellular factors that expose the L1 first-strand cDNA during TPRT. Notably, activation-induced deaminase (AID) has been demonstrated to deaminate both strands of actively transcribed DNA regions (entailing RNA-DNA heteroduplexes

on the template strand), to generate somatic hypermutation during immunoglobulin gene diversification [34-36]. RNaseH2 has not been found to participate in this process. However, the RNA exosome, a cellular RNA processing and degradation complex, has recently been identified to associate with AID, and facilitate its access to template strand DNA [37]. As AID and A3A are related, it is tempting to speculate that A3A could also associate with the RNA exosome, allowing A3A to access the first-strand L1 cDNA in the RNA-cDNA heteroduplex during TPRT.

Is there a deaminase-independent component of A3A-mediated L1 inhibition?

In Chapter 3, we uncover a deaminase-dependent mechanism for A3A-mediated L1 inhibition, supported by direct evidence of A3A-mediated editing of L1 insertions, and the observation that expression of uracil glycosylase inhibitor protein (UGI) alleviates inhibition. Notably, alleviation of inhibition by UGI is efficient, but consistently fails to restore retrotransposition to 100% of control levels (Figure 3.12 C and 3.17 B). There are a number of possible explanations for this observation. The first stems from the fact that we rely on the expression of the spliced *mneoI* reporter gene to detect retrotransposition. A3A-mediated editing of the *mneoI* region of the L1 cDNA likely causes deleterious missense or nonsense mutations which incapacitate the neomycin phosphotransferase gene a certain percentage of the time. Thus, these insertions would not be detected as successful retrotransposition events in the cultured cell assay, even in the presence of UGI. Second, mammalian cells express two nuclear uracil DNA

glycosylases with activity on single-stranded DNA: UNG2 [38], which is inhibited by UGI [39], and SMUG1, which is not affected by UGI [40]. Thus, SMUG1 could be responsible for degrading edited L1 substrates when UNG2 is blocked by UGI. This scenario is unlikely, at least in stable UGI-expressing U2OS cells, as previously published control experiments demonstrate a lack of detectable uracil DNA glycosylase activity in these cells [14].

Finally, it is possible that A3A effects a deaminase-independent mode of L1 inhibition, in addition to its deaminase-dependent activity. Notably, inhibition of *vif*-deficient HIV infection by A3G likely involves both deaminase-dependent and deaminase-independent mechanisms [41-49]. A3G robustly edits minus-strand viral cDNAs, which can give rise to defective progeny viruses. It appears that both UNG and SMUG-1 are dispensable for A3G-mediated inhibition of viral infectivity, suggesting that excision of uracil bases and APE-mediated degradation of HIV cDNAs is not responsible for this inhibition [40, 50]. Conversely, deaminase-deficient A3G mutants robustly inhibit *vif*-deficient HIV infectivity, and a recent model suggests that A3G functions as a physical "road-block" on the HIV RNA, to impede elongation of HIV reverse transcripts [42]. The relative contribution of deaminase-dependent and independent mechanisms to A3G-mediated HIV restriction remains unclear.

As discussed in Chapter 4, mutating Zn^{2+}-coordinating residues within the single CDA active site in A3A may compromise the structural integrity of the protein, abrogating deaminase activity as well as other putative activities that could effect deaminase-independent L1 inhibition. In contrast, A3G is a double

deaminase domain protein [51, 52]. Mutational analysis reveals that only the C-terminal domain of A3G has deaminase activity [53]; the N-terminal domain bears the consensus CDA active site, but is deficient in deaminase activity and may be important for other enzymatic functions. Therefore, I hypothesize that in contrast to A3A, mutations in the A3G CDA-competent active site are less likely to compromise the overall structural integrity of the A3G protein, making separation of deaminase activity from antiretroviral activity readily attainable.

In Chapter 4, we employ an A3A mutant, A3A_F75L, previously reported to lack deaminase activity but retain the ability to inhibit adeno-associated virus replication [54]. We find that this mutant potently restricts L1 retrotransposition, so, based on previous reports [54], we initially regarded it as a separation-of-function mutant indicative of a deaminase-independent mode of A3A-mediated L1 restriction. However, we ultimately found that A3A_F75L-mediated inhibition is alleviated by UGI expression (Figure 4.5), and that A3A_F75L has greater deaminase activity than A3A_C106S in a bacterial mutator assay (Weitzman and Kohli, personal communication). The most parsimonious explanation for the data is that A3A_F75L lacks deaminase activity *in vitro*, but retains deaminase activity when expressed in the cellular context (mammalian or prokaryotic). Recovering L1 insertions generated in the presence of UGI and A3A_F75L, as done in Figure 3.16 and 3.19 for wild-type A3A, would be the most effective way to determine whether A3A_F75L has deaminase activity in cells. Alternately, in a potentially more efficient approach, a PCR-based strategy could be used to amplify the retrotransposed NEO gene from a pool of cells harboring retrotransposition

events that took place in the presence of A3A_F75L. Evidence of deamination could then be detected by next-generation sequencing.

Overall, our mechanistic examination of A3A-mediated L1 inhibition has been a source of insight about how deaminase-dependent retroelement inhibition can occur. Due to the inherent difficulties of A3A mutational analysis, continuing to pursue a deaminase-independent component of A3A-mediated L1 inhibition represents a challenging undertaking. Indeed, I argue that a mechanistic examination of L1 inhibition by APOBEC3B (A3B) is more likely to provide insight about deaminase-independent L1 inhibition.

How does APOBEC3B inhibit L1 retrotransposition?

Like A3A, APOBEC3B (A3B) potently restricts L1 retrotransposition in cultured cells [1, 17, 53, 55-57]. Indeed, we find that A3B acts in a sequence-independent manner, as it is capable of restricting LINE elements from mouse and zebrafish (Figure 2.3). Unlike A3A, however, mutation of the A3B deaminase-competent active site, which resides in the C-terminal domain of the protein, does not compromise A3B-mediated L1 inhibition [53, 55]. Furthermore, UGI expression modestly alleviates L1 inhibition by wild-type A3B, but has no effect on inhibition by deaminase-deficient A3B mutants (Figure 3.14). Notably, A3B contains a nuclear import signal, and has the ability to shuttle between the cytoplasm and nucleus, giving it access to L1 TPRT [53]. I therefore hypothesize that, similar to what has been reported for A3G-mediated HIV restriction [41-49], A3B inhibits L1 retrotransposition by deaminating single-stranded DNA within

TPRT intermediates, and also functions in a deaminase-independent manner to physically block the elongation of the L1 first-strand cDNA on its RNA template.

To investigate the action of A3B during L1 TPRT, one could generate purified recombinant A3B protein (rA3B), and include it in the LEAP reaction. As deaminase active site A3B mutants retain the ability to inhibit L1, I suggest using rA3A as a negative control for direct L1 RT inhibition. If rA3B blocks the procession of L1 RT, full-length LEAP products will likely be diminished. Therefore, to detect the presence of shorter LEAP products, I suggest using a set of tiled L1 3' end primers in the PCR amplification step. As reported for A3G-mediated HIV RT inhibition [42], if A3B blocks L1 RT procession I would expect an inverse correlation between LEAP product length and abundance in the presence of rA3B. Alternately, if A3B does not directly block L1 RT procession, robust full-length LEAP products will be generated in the presence of rA3B. Regardless of whether A3B diminishes the quantity and length of LEAP products, I would expect sequencing to reveal signatures of rA3B-mediated deamination.

If rA3B does not block L1 RT procession, then A3B likely functions in a deaminase-independent manner at a different point during the L1 retrotransposition process. One possibility is that A3B, which is a shuttle protein and can translocate between the cytoplasm and the nucleus [53], sequesters the L1 RNA or RNP in the cytoplasm, as proposed for A3G-mediated Alu inhibition [29]. Indeed, it would be interesting to determine by co-immunoprecipitation experiments whether A3B can interact with L1 ORF1p or ORF2p, and if such an interaction is direct or mediated by an RNA bridge.

What are the physiological roles of A3A and A3B in humans?

It is well demonstrated that A3A potently restricts L1 retrotransposition when ectopically expressed in cultured cells [1, 17, 53, 56, 57]. However, A3A expression is restricted to peripheral blood lymphocytes [1, 53] and is induced by interferon [2]. Therefore, it is unlikely that A3A plays a role in the restriction of heritable L1 retrotransposition events in humans, which must occur in the germline or very early embryo prior to germline specification. Similarly, A3A also restricts adeno-associated virus (AAV) replication [1, 54]; however, AAV is not known to be pathogenic in humans and therefore may not represent a natural target of A3A. Thus, the physiological role of A3A in humans likely remains to be uncovered. The fact that A3A expression is induced by interferon and can mutate transfected DNA [2] suggests that A3A may play a role in an innate immune response to an as-yet unidentified exogenous pathogen. On the other hand, recent evidence that A3A can deaminate host cell genomic DNA [14, 19] opens up the possibility that, similar to AID, endogenous A3A may target the genome to create somatic mosaicism as part of a presently undescribed cellular process. Interestingly, a new report indicates that endogenous A3A in monocytes localizes to the cytoplasm, and does not edit cellular genomic DNA [58]. The A3A protein has a molecular weight of <40 kD and therefore is small enough to diffuse into the nucleus, and in cultured cells ectopically expressed A3A localizes diffusely throughout the nucleus and cytoplasm [1, 53]. Therefore, the mechanism by which endogenous A3A is restricted to the cytoplasm remains

to be determined. Indeed, it will be interesting to see whether future studies uncover a defined physiological role for this potent cytidine deaminase.

In contrast to A3A, A3B is expressed in a wide variety of tissues [52, 53], and endogenous A3B has been implicated in L1 restriction in human embryonic stem cells (hESCs) [11]. Thus, A3B is more likely than A3A to represent a physiologically relevant retroelement restriction factor. I therefore propose that a mechanistic examination of A3B-mediated L1 inhibition, as described above, represents an important step in understanding the regulation of potentially heritable L1 retrotransposition events.

Notably, recent studies have uncovered evidence for endogenous L1 mobility in the human brain and neuronal cell types [59-61], where A3B expression has not explicitly been demonstrated [52]. It is interesting to speculate that A3B expression may be repressed in neuronal cell types, allowing L1 retrotransposition to generate somatic mosaicism. It is also interesting to note that a common deletion polymorphism exists in humans, in which the last exon of the *APOBEC3A* gene and the first seven exons of the *APOBEC3B* gene are deleted [62]. This deletion is predicted to generate a fusion transcript that produces a protein identical to A3A, and results in a complete loss of A3B. This polymorphism is nearly fixed in Oceanic populations, exists at high allele frequencies in East Asian and American Indian populations, and is rare in Africans and Europeans [63]. It would be interesting to determine whether Oceanic populations have a higher frequency of heritable L1-mediated insertions when compared to Africans and Europeans. Furthermore, one could investigate

whether induced pluripotent stem cells (iPSCs) [64-66] derived from individuals bearing the deletion accommodate higher levels of retrotransposition from engineered retrotransposon vectors when compared to iPSCs from individuals with functional *APOBEC3* alleles. Likewise, it would be interesting to examine various somatic tissues from individuals bearing the *APOBEC3* deletion to determine the prevalence of endogenous L1-mediated retroelement insertions. Of course, it will be important to recognize the inherent difficulties in establishing positive and negative controls, when comparing cell lines and tissues from individuals with different genetic backgrounds. Overall, as sequencing technologies improve, allowing detailed examination of endogenous L1 insertions, future studies will no doubt elucidate the role of A3B and other cellular factors in regulating heritable and somatic retrotransposition events.

Summary

The mobility of transposable elements such as L1 represents an ancient and ongoing threat to genome stability, and the evolution of the APOBEC3 family of cytidine deaminases is hypothesized to have been driven by genetic conflict with endogenous retroelements. Indeed, certain members of the APOBEC3 family, most potently A3A, restrict L1 retrotransposition in cultured cells. Previously, little was known about the molecular mechanisms of A3-mediated L1 inhibition. In this thesis, I have uncovered a deaminase-dependent mechanism for A3A-mediated L1 restriction, and in doing so gained insight into various aspects of APOBEC3 and L1 biology. Future studies will no doubt further

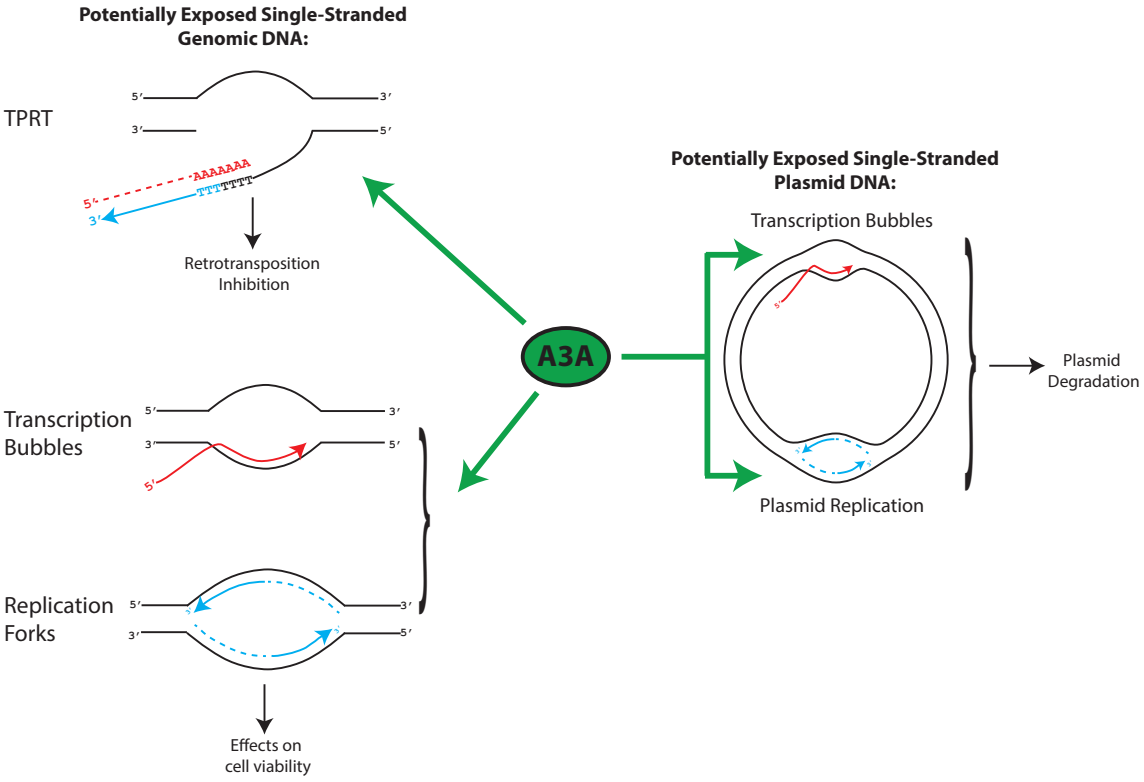
elucidate the dynamic interplay between endogenous retroelements and host restriction factors.

Acknowledgements

I would like to acknowledge members of the Moran lab for helpful discussion. Sean Ferris performed preliminary Southern blot experiments to investigate the number of insertions per cell generated by various L1 elements. Dr. Dan Stetson kindly shared his preliminary data on the effects of RNaseH2 overexpression and knockdown on L1 retrotransposition.

Figure 5.1: Possible explanations for "off-target" effects of A3A expression in cultured cells

A3A has been demonstrated to act upon transiently single-stranded DNA during TPRT (top left). A3A may also access single-stranded genomic DNA at transcription bubbles and replication forks. The downstream consequences of genomic DNA deamination likely include a DNA damage response and cell cycle arrest (bottom left) [14]. In addition, A3A may access single-stranded DNA on transfected plasmids, either during transcription from the plasmid or during plasmid replication (right). These activities could potentially lead to degradation of the transfected plasmid.


Potentially Exposed Single-Stranded Genomic DNA:
TPRT
5′
3′
AAAAAAA
TTTTTTTT
Retrotransposition Inhibition
Transcription Bubbles
Replication Forks
Effects on cell viability
A3A
Potentially Exposed Single-Stranded Plasmid DNA:
Transcription Bubbles
Plasmid Replication
Plasmid Degradation

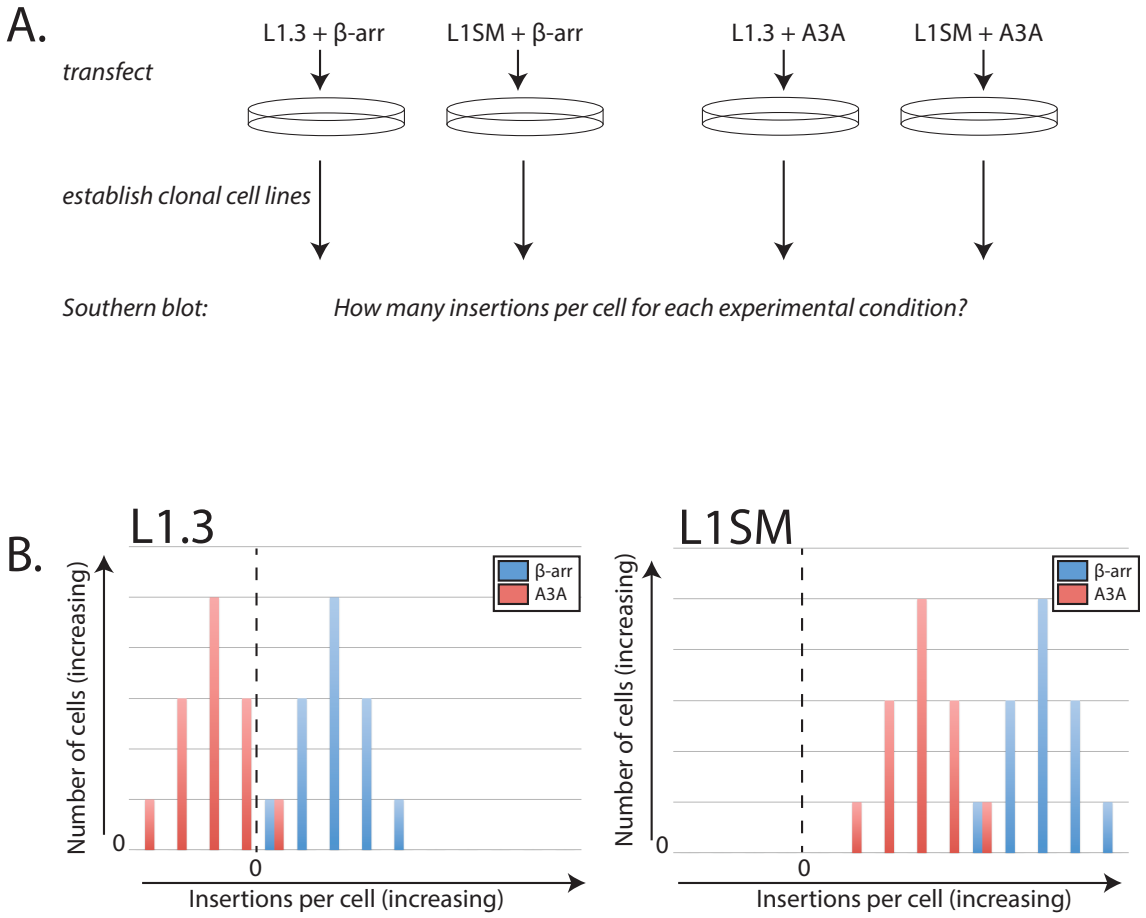
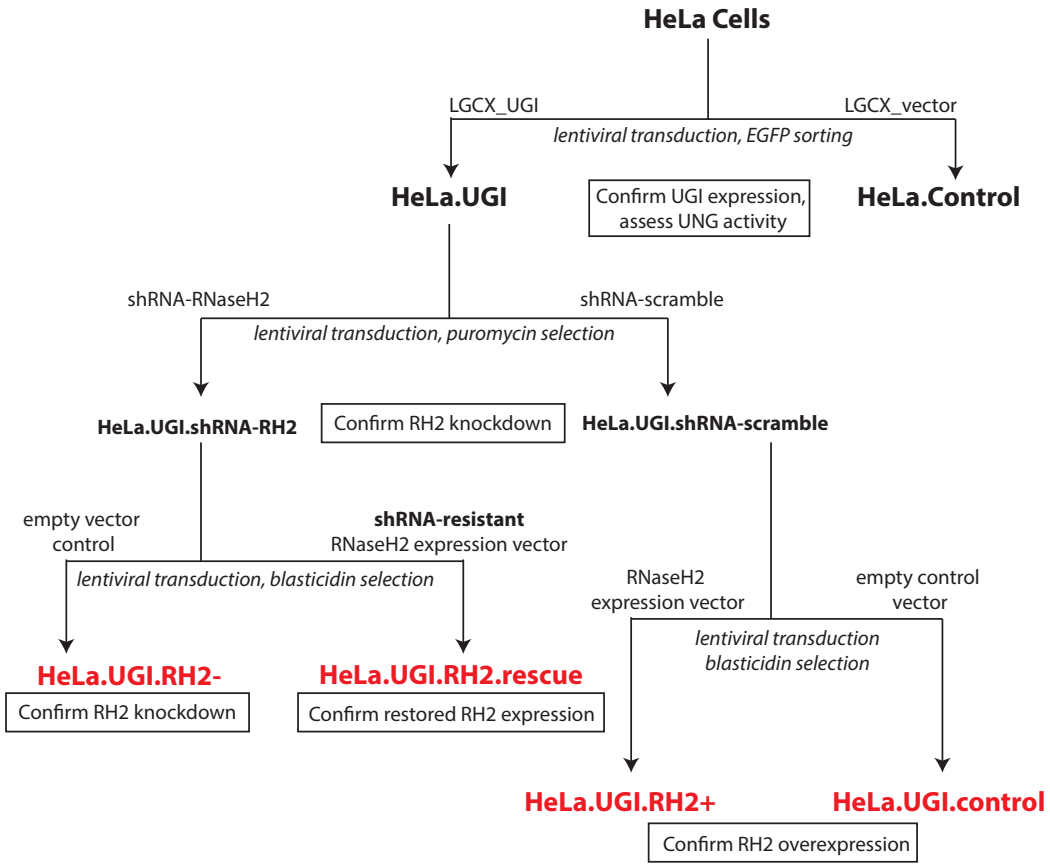


Figure 5.2: Proposed experiment to examine the impact of A3A expression on the number of L1 insertions per cell

A. Generation and characterization of clonal cell lines. Above, HeLa cells are co-transfected with L1.3 or L1SM, and A3A or β-arr control plasmid. Clonal cell lines are generated by selection of G418 and expansion of the resulting colonies. Genomic DNA from each cell line is subjected to Southern blotting with a probe to the NEO gene.

B. Predicted distributions of L1 insertions per cell. L1.3 is shown at left, L1SM, at right. The x-axes represent the number of insertions per cell; the dashed line highlights the "zero" point. The y-axes represent the number of cells. Blue bars indicate β-arr control distributions; red bars indicate the predicted leftward shift in the presence of A3A.

A.



B.

HeLa.UGI Cell Lines:

	shRNA	RNaseH2	Expected Deamination Level
HeLa.UGI.control	scramble shRNA	vector only	moderate
HeLa.UGI.RH2+	scramble shRNA	RNaseH2 expression	high
HeLa.UGI.RH2-	RNaseH2 shRNA	vector only	low
HeLa.UGI.RH2.rescue	RNaseH2 shRNA	shRNA-resistant RNaseH2 expression	moderate*

* Depends on effeciency of ectopic RNaseH2 expression

Figure 5.3: Experiment to test L1 cDNAs for single-strandedness during target-primed reverse transcription (TPRT)

A. HeLa cells are transduced with a lentiviral vector harboring the humanized UGI gene (LGCX_UGI) or empty vector control. Cells are flow-sorted on GFP to generate a polyclonal population of GFP and UGI-expressing cells (HeLa.UGI). Next, sequential rounds of lentiviral transduction and selection are used to generate HeLa.UGI.control, HeLa.UGI.RH2+, and HeLa.UGI.RH2- cell lines.

B. The table indicates the genotype of each engineered cell line, as well as the expected relative levels of A3A-mediated deamination within retrotransposed L1 sequences.

References

1. Chen, H., et al., *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.* Curr Biol, 2006. **16**(5): p. 480-5.
2. Stenglein, M.D., et al., *APOBEC3 proteins mediate the clearance of foreign DNA from human cells.* Nat Struct Mol Biol, 2010. **17**(2): p. 222-9.
3. Koning, F.A., et al., *Target cell-mediated editing of HIV-1 cDNA by APOBEC3 proteins in human macrophages.* Journal of virology, 2011. **85**(24): p. 13448-52.
4. Refsland, E.W., et al., *Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction.* Nucleic acids research, 2010. **38**(13): p. 4274-84.
5. Berger, G., et al., *APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells.* PLoS pathogens, 2011. **7**(9): p. e1002221.
6. Thielen, B.K., et al., *Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms.* The Journal of biological chemistry, 2010. **285**(36): p. 27753-66.
7. Wissing, S., et al., *Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility.* Hum Mol Genet, 2011.
8. Garcia-Perez, J.L., et al., *LINE-1 retrotransposition in human embryonic stem cells.* Hum Mol Genet, 2007. **16**(13): p. 1569-77.
9. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
10. Wei, W., et al., *A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events.* Anal Biochem, 2000. **284**(2): p. 435-8.
11. Wissing, S., et al., *Endogenous APOBEC3B Restricts LINE-1 Retrotransposition in Transformed Cells and Human Embryonic Stem Cells.* J Biol Chem, 2011. **286**(42): p. 36427-37.
12. Conticello, S.G., *The AID/APOBEC family of nucleic acid mutators.* Genome Biol, 2008. **9**(6): p. 229.
13. Rogozin, I.B., et al., *Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase.* Nature immunology, 2007. **8**(6): p. 647-56.
14. Landry, S., et al., *APOBEC3A can activate the DNA damage response and cause cell-cycle arrest.* EMBO Rep, 2011. **12**(5): p. 444-50.
15. Conticello, S.G., et al., *Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases.* Mol Biol Evol, 2005. **22**(2): p. 367-77.
16. Bulliard, Y., et al., *Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities.* J Virol, 2011. **85**(4): p. 1765-76.

17. Muckenfuss, H., et al., *APOBEC3 proteins inhibit human LINE-1 retrotransposition.* J Biol Chem, 2006. **281**(31): p. 22161-72.
18. Bogerd, H.P., et al., *APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.* Nucleic acids research, 2006. **34**(1): p. 89-95.
19. Suspene, R., et al., *Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism.* Proc Natl Acad Sci U S A, 2011. **108**(12): p. 4858-63.
20. Suspene, R., et al., *Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR.* J Gen Virol, 2005. **86**(Pt 1): p. 125-9.
21. Gilbert, N., et al., *Multiple fates of L1 retrotransposition intermediates in cultured human cells.* Mol Cell Biol, 2005. **25**(17): p. 7780-95.
22. Han, J.S. and J.D. Boeke, *A highly active synthetic mammalian retrotransposon.* Nature, 2004. **429**(6989): p. 314-8.
23. Goodier, J.L., et al., *A novel active L1 retrotransposon subfamily in the mouse.* Genome Res, 2001. **11**(10): p. 1677-85.
24. Sugano, T., M. Kajikawa, and N. Okada, *Isolation and characterization of retrotransposition-competent LINEs from zebrafish.* Gene, 2006. **365**: p. 74-82.
25. Dewannieux, M., C. Esnault, and T. Heidmann, *LINE-mediated retrotransposition of marked Alu sequences.* Nat Genet, 2003. **35**(1): p. 41-8.
26. Dewannieux, M. and T. Heidmann, *L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells.* J Mol Biol, 2005. **349**(2): p. 241-7.
27. Weiner, A.M., P.L. Deininger, and A. Efstratiadis, *Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information.* Annual review of biochemistry, 1986. **55**: p. 631-61.
28. Hulme, A.E., et al., *Selective inhibition of Alu retrotransposition by APOBEC3G.* Gene, 2007. **390**(1-2): p. 199-205.
29. Chiu, Y.L., et al., *High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15588-93.
30. Doucet, A.J., et al., *Characterization of LINE-1 ribonucleoprotein particles.* PLoS Genet, 2010. **6**(10).
31. Piskareva, O. and V. Schmatchenko, *DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro.* FEBS letters, 2006. **580**(2): p. 661-8.
32. Kurzynska-Kokorniak, A., et al., *DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon.* J Mol Biol, 2007. **374**(2): p. 322-33.
33. Zheng, L. and B. Shen, *Okazaki fragment maturation: nucleases take centre stage.* Journal of molecular cell biology, 2011. **3**(1): p. 23-30.

34. Muramatsu, M., et al., *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.* Cell, 2000. **102**(5): p. 553-63.
35. Muramatsu, M., et al., *Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells.* J Biol Chem, 1999. **274**(26): p. 18470-6.
36. Arakawa, H., J. Hauschild, and J.M. Buerstedde, *Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion.* Science, 2002. **295**(5558): p. 1301-6.
37. Basu, U., et al., *The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates.* Cell, 2011. **144**(3): p. 353-63.
38. Nilsen, H., et al., *Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene.* Nucleic acids research, 1997. **25**(4): p. 750-5.
39. Karran, P., R. Cone, and E.C. Friedberg, *Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase.* Biochemistry, 1981. **20**(21): p. 6092-6.
40. Langlois, M.A. and M.S. Neuberger, *Human APOBEC3G can restrict retroviral infection in avian cells and acts independently of both UNG and SMUG1.* J Virol, 2008. **82**(9): p. 4660-4.
41. Anderson, J.L. and T.J. Hope, *APOBEC3G restricts early HIV-1 replication in the cytoplasm of target cells.* Virology, 2008. **375**(1): p. 1-12.
42. Bishop, K.N., et al., *APOBEC3G inhibits elongation of HIV-1 reverse transcripts.* PLoS Pathog, 2008. **4**(12): p. e1000231.
43. Browne, E.P., C. Allers, and N.R. Landau, *Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent.* Virology, 2009. **387**(2): p. 313-21.
44. Iwatani, Y., et al., *Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G.* Nucleic Acids Res, 2007. **35**(21): p. 7096-108.
45. Li, X.Y., et al., *APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription.* J Biol Chem, 2007. **282**(44): p. 32065-74.
46. Mangeat, B., et al., *Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.* Nature, 2003. **424**(6944): p. 99-103.
47. Miyagi, E., et al., *Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1.* J Virol, 2007. **81**(24): p. 13346-53.
48. Zhang, H., et al., *The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.* Nature, 2003. **424**(6944): p. 94-8.
49. Newman, E.N., et al., *Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity.* Curr Biol, 2005. **15**(2): p. 166-70.
50. Kaiser, S.M. and M. Emerman, *Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute*

to the antiviral effects of the cytidine deaminase Apobec3G. J Virol, 2006. **80**(2): p. 875-82.
51. Li, J., M.J. Potash, and D.J. Volsky, *Functional domains of APOBEC3G required for antiviral activity.* J Cell Biochem, 2004. **92**(3): p. 560-72.
52. Jarmuz, A., et al., *An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.* Genomics, 2002. **79**(3): p. 285-96.
53. Bogerd, H.P., et al., *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8780-5.
54. Narvaiza, I., et al., *Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase.* PLoS Pathog, 2009. **5**(5): p. e1000439.
55. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.* J Biol Chem, 2006. **281**(25): p. 16837-41.
56. Kinomoto, M., et al., *All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.* Nucleic Acids Res, 2007. **35**(9): p. 2955-64.
57. Niewiadomska, A.M., et al., *Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association.* J Virol, 2007. **81**(17): p. 9577-83.
58. Land, A.M., et al., *Endogenous APOBEC3A DNA Cytosine Deaminase Is Cytoplasmic and Nongenotoxic.* The Journal of biological chemistry, 2013. **288**(24): p. 17253-60.
59. Coufal, N.G., et al., *L1 retrotransposition in human neural progenitor cells.* Nature, 2009. **460**(7259): p. 1127-31.
60. Baillie, J.K., et al., *Somatic retrotransposition alters the genetic landscape of the human brain.* Nature, 2011.
61. Muotri, A.R., et al., *Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition.* Nature, 2005. **435**(7044): p. 903-10.
62. Kidd, J.M., et al., *Population stratification of a common APOBEC gene deletion polymorphism.* PLoS genetics, 2007. **3**(4): p. e63.
63. Kidd, J.M., et al., *Population stratification of a common APOBEC gene deletion polymorphism.* PLoS Genet, 2007. **3**(4): p. e63.
64. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.
65. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
66. Wernig, M., et al., *A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types.* Nat Biotechnol, 2008. **26**(8): p. 916-24.

Appendix

Characterizing Engineered L1 Retrotransposition Events in Cultured Human Embryonic Stem Cells

Introduction

As discussed in Chapter 1, the early embryo, before germline specification, represents a critical window for the generation of heritable L1 insertions. Human embryonic stem cells (hESCs) reprsent an ideal model to study L1 retrotransposition in early human development. Previous studies have demonstrated that L1 can retrotranspose in hESCs in culture [1-3] and that L1 retrotransposition can occur *in vivo* during early human embryonic development [4].

Here, we generate a library of L1 insertions in H9 and H13B hESCs using a “second-generation” L1 retrotransposition indicator plasmid, pUB-RAM-LRE3. Characterizing these insertions allows us to observe the structural hallmarks, putative target-site preference, and alterations to target-site DNA associated with

L1 retrotransposition in the early embryo. Furthermore, unique features of the pUB-RAM-LRE3 indicator plasmid allows manipulation of the L1-delivered reporter cassette in clonal insertion-bearing cell lines.

Results/Discussion

The pUB-RAM-LRE3 retrotransposition indicator plasmid

To facilitate the generation and manipulation of L1 insertions in clonal hESC cell lines, we engineered a "second-generation" L1 retrotransposition indicator plasmid termed pUB-RAM-LRE3 (Figure A.1). This construct consists of the human L1 element LRE3 [5], driven by its 5'UTR, as well as the ubiquitin C ligase (UBC) promoter [6]. The UBC promoter is intended to maximize L1 expression. The construct bears the standard *mneoI* retrotransposition indicator cassette [7, 8], followed by the ColE1 bacterial orgin of replication, which allows recovery of retrotransposed L1 insertions as autonomously replicating plasmids in bacteria, as previously described [9, 10]. The entire *mneoI*-ColE1 cassette is flanked by heterologous loxP sites [11, 12], which in turn are flanked by FRT sites [13]. After integration, the FRT sites are designed to allow Flp recombinase-mediated removal of the *mneoI*-ColE1 cassette. The heterologous loxP sites are designed to facilitate replacement of the integrated *mneoI*-ColE1 cassette with any gene of choice.

Generation of clonal insertion-bearing cell lines

Clonal insertion-bearing cell lines were generated by transfecting H9 and H13B hESCs with the pUB-RAM-LRE3 plasmid, and imposing G418 selection for 14 days. The resulting G418-resistant colonies were then expanded to generate clonal cell lines. We have generated 28 cell lines, 15 in H13B (designated RAM 1-15) and 13 in H9 hESCs (designated RAM A-M). A subset of cell lines were validated as bona-fide hESCs by immunostaining for pluripotency markers (Figure A.2), and the presence of the spliced *mneoI* cassette delivered by retrotransposition was confirmed for all cell lines by PCR (Figure A.3).

Four pUB-RAM-LRE3 insertions from H9 hESCs have been characterized (Figure A.4). These insertions bear the typical hallmarks of L1 retrotransposition, including target-site duplications and poly-A tails. Interestingly, one insertion was apparently prematurely polyadenylated within the SV40 promoter that drives expression of the *mneoI* indicator gene (Figure A.4, line RAM-I). This insertion was characterized by inverse PCR, as premature polyadenylation resulted in omission of the ColE1 bacterial origin of replication from the insertion.

L1-delivered reporter genes may undergo silencing in hESCs

The generation of G418-resistant colonies harboring pUB-RAM-LRE3-derived retrotransposition events proved to be a slow process, with fewer than five G418-resistant colonies arising from a typical transfection. We therefore speculate that the L1-delivered *mneoI* reporter gene may become epigenetically silenced during or after insertion in hESCs. We chose twelve H9 clonal cell lines

(RAM A through L), harboring the retrotransposed L1 indicator cassette, as verified by PCR, that had been maintained on irradiated feeder MEFs in the absence of selection for 4-8 passages. We re-subjected these cells to G418 selection at a range of concentrations (50-150 μg/ml) for seven days, with untransfected H9 hESCs serving as a negative control. Interestingly, different cell lines displayed varying degrees of resistance to G418 selection, with some lines appearing fully resistant (ie RAM-D; Figure A.5, top row, fourth well), while others were much more susceptible to G418 selection (ie RAM-J; Figure A.5, bottom row, second well). Thus, while these insertion-bearing hESCs clearly maintained *mneoI* reporter gene expression through the initial selection process to generate clonal cell lines, in some cases expression from the *mneoI* cassette appears to have become silenced with subsequent passaging in culture.

Materials and Methods

hESC culture:

H9 and H13b hESCs (WiCell) were grown on gelatin-coated plates on pre-irradiated mouse embryonic fibroblast (MEF) feeder cells (GlobalStem or ChemiCon). hESC culture media contains DMEM F12 Media (Gibco), 20% Knockout Serum Replacer (KOSR; Gibco), 4 ng/ml FGF-2 (fibroblast growth factor) (Invitrogen), 1mM L-glutamine (Gibco), 50 μM β-mercaptoethanol, and 0.1 mM non-essential amino acids (Gibco). Cells were manually passaged using the StemPro EZ passage Cell Passaging Tool (Invitrogen) every 3 or 4 days.

Plasmids:

All plasmids were grown in DH5α (F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1) competent *E. coli* (Invitrogen; Carlsbad, CA. Prepared in house as described in [14]). Plasmids wereprepared using the Qiagen Plasmid Midi Kit (QIAGEN; Hilden, Germany) according to the manufacturer's protocol.

pUB-RAM-LRE3: consists of the pBSKS II backbone containing the LRE3 element driven by a ubiquitin C ligase (UBC) promoter. The 3'UTR of the element harbors an engineered retrotransposition cassette, *mneoI,* and the ColE1 bacterial origin of replication. These sequences are flanked by heterologous LoxP sites (LoxP and Lox511), which in turn are flanked by FRT sites.

Transfection and selection of hESCs:

1-2 hours prior to transfection, hESC media was removed from cells and replaced with hESC media supplemented with 10 μM ROCK inhibitor Y-27632. In the meantime, 10 cm dishes (Corning) were coated with matrigel (BD biosciences) diluted 1:15 in cold DMEM-F12 media and incubated for 1-2 hours at room temperature. For each round of transfection, one plate of hESCs was sacrified for counting purposes: cells were dissociated using trypsin, and cell density determined by counting on a hematocytometer. For cells to be transfected, hESCs were dissociated using TrypLE Select (Invitrogen) and

washed once with 1X PBS. For each transfection, approximately ~2x10^6 cells were resuspended in 100 µl of V-kit solution (Amaxa). This suspension was mixed with 4 µg of plasmid DNA, and transferred to an Amaxa nucleofection cuvette. Cells were nucleofected using Amaxa program A-23. Following nucleofection, 500 µl of pre-warmed RPMI was added to the cuvette and the entire suspension was then transferred to a microcentrifuge tube and incubated at 37°C for 15-30 minutes. Each transfection was seeded on a Matrigel-coated plate and maintained in MEF-conditioned media (made in house) supplemented with 20 ng/ml of FGF-2 and 10 µM Y-27632 ROCK inhibitor (MCM20-iROCK).

Transfected cells were fed daily with MCM20-iROCK for 5 days. Selection with 50 µg G418 (Invitrogen) was initiated at day 6, and continued for one week. Selection with 100 µg G418 was then initiated at day 14, and continued for one week. G418-resistant colonies were manually dissected and replated onto irradiated feeder MEFs, and expanded to generate clonal cell lines. Genomic DNA was prepared using the Qiagen Blood and Cell Culture Midi Kit, according to the manufacturer's instructions.

Immunocytochemistry:

Approximately 5x10^5 hESCs were plated per well of a 12-well dish, on irradiated feeder MEFs. Cells were allowed to grow for four days, at which point they were fixed using 2% paraformaldehyde. Cells were washed twice with 1X D-PBS (Invitrogen), and permeabilized using permeabilization buffer (0.1% triton, 0.1% sodium citrate). Cells were then washed once with serum wash (9.8 mL D-

PBS, 100 µl normal donkey serum, 100 µl 10 % sodium azide), and incubated in blocking buffer (1 mL normal donkey serum, 50 µl Triton X-100, 100 µl 10% sodium azide, 8.85 ml D-PBS) for 30 minutes in a humidified chamber at room temperature. OCT4 antibody (Santa Cruz, SC-8628) was diluted 1:400 in antibody dilution buffer (100 µl normal donkey serum, 50 µl Triton X-100, 100 µl 10% sodium azide, 9.75 ml D-PBS). Cells were incubated in primary OCT4 antibody overnight at 4°C in a humidified chamber. The next day, cells were washed with serum wash. The secondary antibody (Texas Red Donkey anti-Goat antibody, Affinipure, 705-075-147) was diluted 1:800 in antibody dilution buffer. Cells were incubated for 30 minutes at room temperature in a humidity chamber. Cells were washed with serum wash in the dark, then washed with D-PBS in the dark. Cells were then stained with Hoechst for 5 minutes, then rinsed once with H_2O and once with D-PBS. Cells were visualized using fluorescence microscopy.

NEO PCR:

Primers:

437s: GAGCCCCTGATGCTCTTCGTCC [7]

1808as: CATTGAACAAGATGGATTGCACGC [7]

For each PCR reaction, 300 ng of hESC genomic DNA was used as the template. PCR reactions were performed Roche Expand High-Fidelity PCR system. Cycling conditions are as follows: 95° C for 4 minutes, followed by 35

cycles of 95° C, 45s; 52°C, 20s; 72°C, 1 min., and a ten-minute extension at 72° C. Products were visualized on a 1% agarose gel.

Recovery of L1 insertions:

The L1 insertion recovery procedure

Insertions were recovered generally as previously described [10]. Briefly, 8 µg of genomic DNA was digested with excess restriction enzyme (*Hind*III, *Ssp*I, *Bgl*II, *Bam*HI: NEB; *Bcl*I: Promega) overnight at 37° C. The following morning, an additional 2 µl of enzyme was added to the digest, and incubated at 37° C for 2 hours. Digestion reactions were heat-inactivated (65° C, 15 minutes), or in the case of *Bgl*II digest, cleaned up with the Wizard DNA clean-up kit (Promega). The entire digest was then ligated with 8 µl T4 DNA ligase (NEB) under dilute conditions (500 µl total volume) overnight at 16° C. The next morning, 2 µl additional T4 DNA ligase was added to the reaction and incubated at room temperature for 4 hours. Ligations were concentrated on an Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-100 membrane (Millipore; Billerica, Massachussets) at 8.0xg for 5 minutes, followed by a wash with 500µl dH_2O, for 5 minutes at 8.0 g. 10-100 µl of concentrated DNA was recovered with a short spin. The entire ligation was then added to 500 µl of XL-10 gold ultra-competent *E. coli* (Stratagene; prepared in-house as described in [14]) and incubated on ice for 1-3 hours. Transformations were heat-shocked at 42° C for 38-45 seconds, and allowed to recover on ice for 2 minutes. 2 ml of warm (37° C) LB (no

antibiotic) was added to each transformation. Transformations were incubated overnight at room-temperature on an orbital shaker at 100 rpm. The next morning, transformations were pelleted (300 g) and gently resuspended in 500 µl fresh LB. About 400 µl of this resuspension was plated on a 15 cm Kanamycin plate (30 µg/ml), and the remaining ~100 µl was used to seed a 2 ml liquid culture (30 µg/ml Kanamycin). Plates and liquid cultures were incubated for 18-24 hours at 37° C. Mini-preps (Promega SV mini-prep kit) were prepared from 2 ml cultures. From 15 cm plates, individual colonies were picked to inoculate 2 ml cultures (30 µg/ml Kanamycin), grown overnight at 37° C, and plasmid DNA was prepared by mini-prep the following day. Plasmid DNA recovered from mini-preps was digested with the original enzyme used for recovery, to confirm intramolecular ligation. To characterize flanking genomic DNA, recovered insertions were sequenced using primers annealing to the 5' (NEOasReco) and 3' (Rescue3seq) ends of the *mneoI*_ColE1 recovery cassette, as well as an oligo dA primer (polyAseq). The genomic locations of insertions were determined by aligning flanking sequence to the human genome (Feb. 2009; GRCh37/hg19), using the BLAT function of the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgBlat).

Inverse PCR

Primers:

NEO210as: GACCGCTTCCTCGTGCTTTACG

NEO1720s: TGCGCTGACAGCCGGAACACG

NEO173as: CATCGCCTTCTATCGCCTTCTTG

NEO1808s: GCGTGCAATCCATCTTGTTCAATG

4-5 µg of genomic DNA was digested with restriction endonuclease (HindIII, BglII, or SspI, NEB). Reactions were heat-inactivated or, when heat inactivation was not possible (BglII), cleaned up using the Promega Wizard DNA clean-up kit. Fragments were ligated under dilute conditions, with a final volume of 1 ml and 3200 units of T4 DNA ligase (Invitrogen). DNA was recovered using phenol extraction, and precipitated using sodium acetate and ethanol. First and second-round PCR reactions were performed using the Roche Long Template PCR system, with buffer 3. Reaction conditions include 10 pMol each primer, 20 nMol each dNTP, and 2.5 units of DNA polymerase mix. 4 µl of the ligation was used as template for the first round of PCR. Primers used were NEO210as, NEO1720s. Cycling conditions were 2 minutes at 95C, followed by 30 cycles of 94° C, 15s; 64° C, 30s; 68° C, 15 minutes; followed by a 30-minute extension at 68° C. 1-5 µl of the first-round PCR product was used for the second-round PCR. Primers used were NEO173as, NEO 1808s. Cycling conditions were identical to the first-round PCR. Products from the second-round PCR were ligated into the pGEM-T Easy vector (Promega) and characterized.

Re-imposing selection on clonal cell lines:

Clonal H9 cell lines generated as described above were passaged on irradiated feeder MEFs in the absence of selection for 2-4 weeks (4-8 passages).

Cells were dissociated using the StemPro EZ Passage Cell Passaging Tool and ~$5x10^5$ cells were plated per well of Matrigel-coated 12-well dishes (BD Biosciences). Twenty-four hours after plating, selection with 50 µg/ml, 100 µg/ml, or 150 µg/ml of G418 was initiated and carried out for seven days.

Acknowledgements:

I would like to acknowledge Jose L. Garcia-Perez, who generated the pUB-RAM-LRE3 construct and established clonal H13B cell lines RAM-1 and RAM-2. The Michigan Center for hES Research provided laboratory space and generated irradiated MEFs, conditioned media, and other reagents for the generation of the H13b clonal cell lines. Nancy Leff and Shuen Hon generated these reagents for the generation of the H9 clonal cell lines. Nancy Leff also assisted greatly with the maintenance of hESC cell lines.

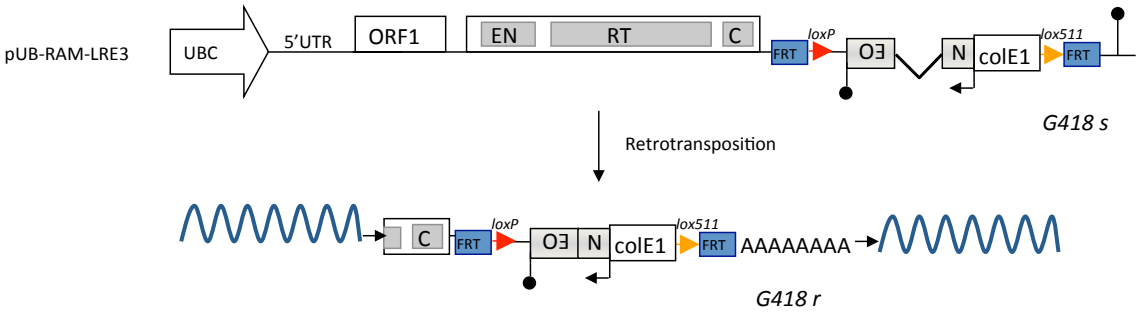


Figure A.1: The pUB-RAM-LRE3 retrotransposition indicator

The human L1 element LRE3 [5] is drived by its 5' UTR, and a ubiquitin C ligase (UBC) promoter [6] to ensure high expression leves. The construct bears the *mneoI* retrotransposition inidcator cassette. The black arrow and black lollipop represent the *mneoI* engineered heterologous promoter and polyadenylation signal, respectively. The *mneoI* indicator gene is followed by the ColE1 bacterial orgin of replication [9, 10]. The entire *mneoI*-ColE1 cassette is flanked by heterologous loxP sites [11, 12], which in turn are flanked by FRT sites [13]. This figure was generated by Jose Garcia-Perez.

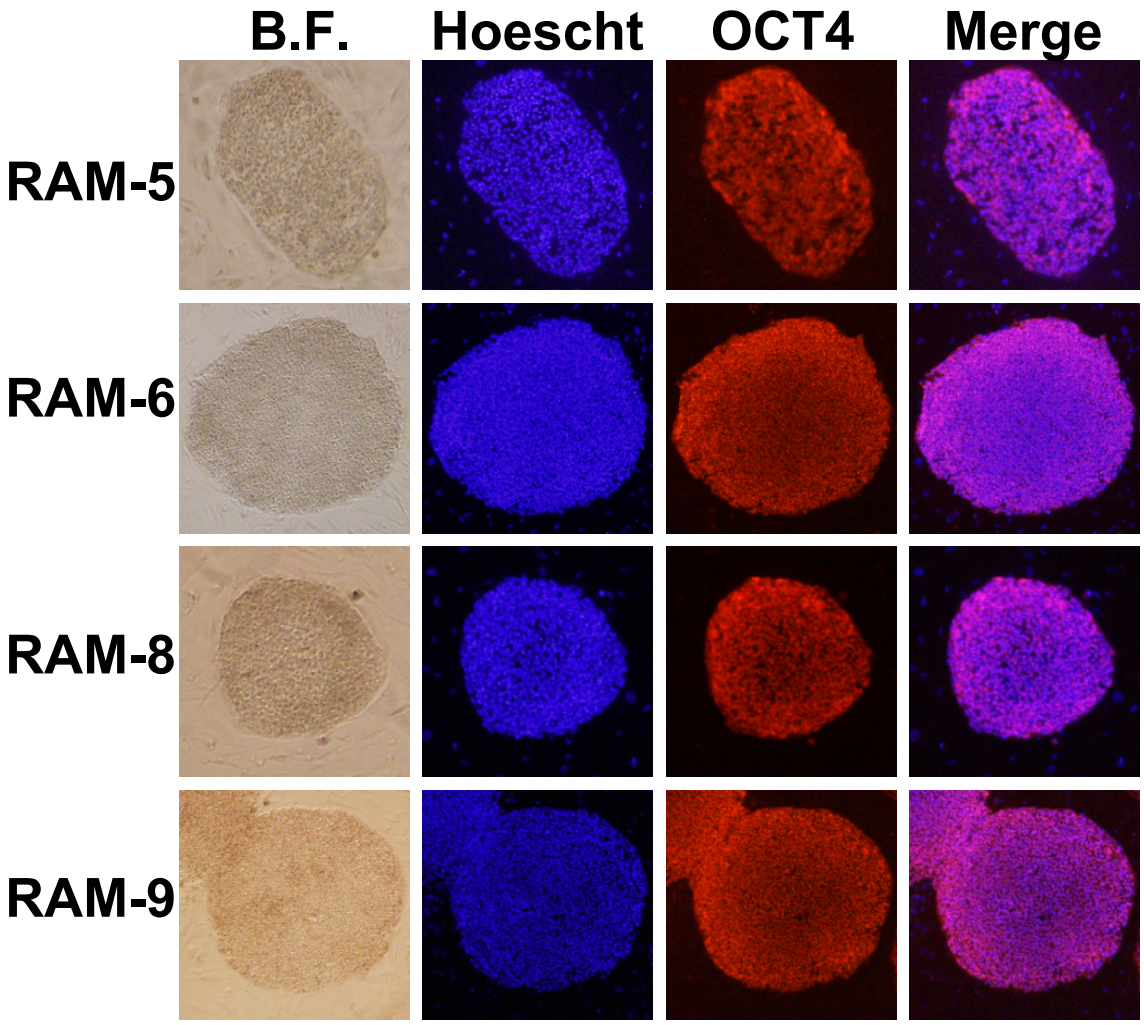


Figure A.2: Immunostaining of H13B hESC cell lines harboring L1 retrotransposition events

Representative colonies from four clonal H13B cell lines harboring pUB-RAM-LRE3 L1 insertions are shown. Brightfield (B.F.) images show typical hESC colony morphology. Hoechst stains the nuclei of hESCs, as well as feeder MEFs. OCT4 is a marker of pluripotency.

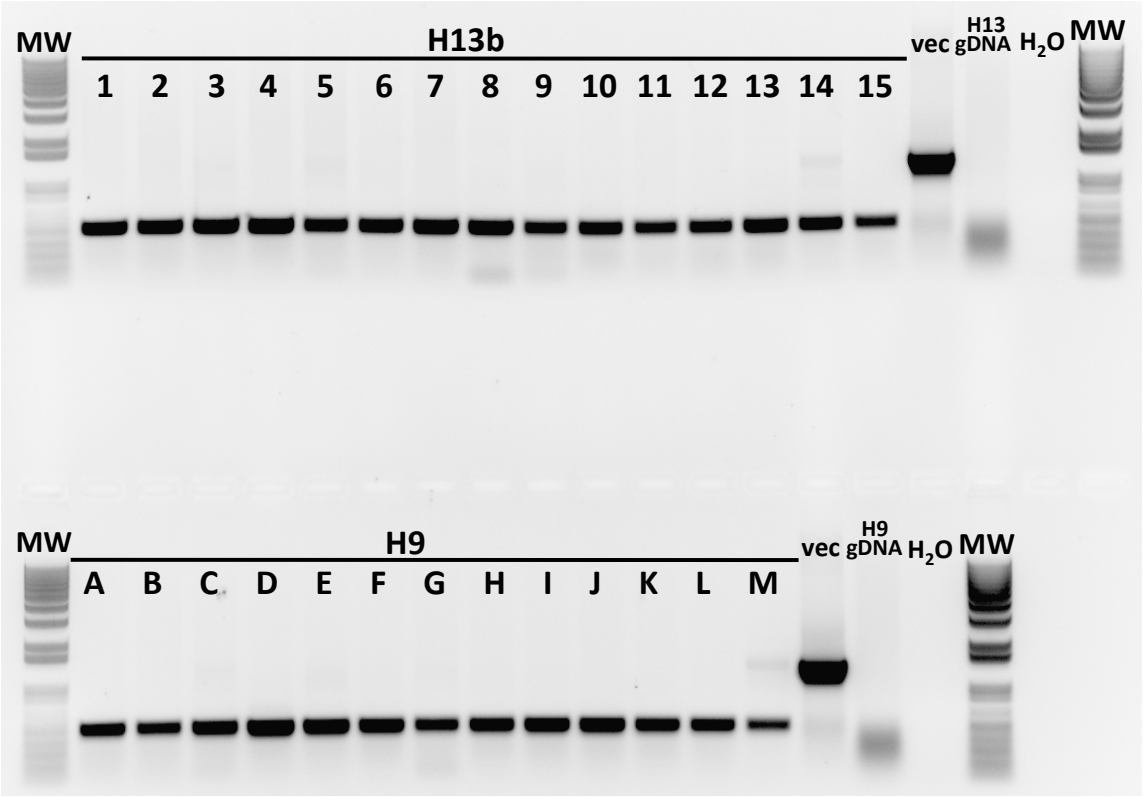


Figure A.3: PCR across the spliced NEO gene confirms L1 insertions in H13b and H9 hESCs

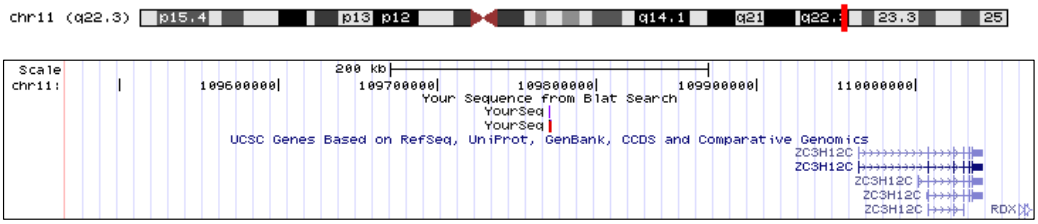
Primers flanking the intron within the NEO retrotransposition indicator cassette (437s, 1808as) were used to amplify the spliced pUB-RAM-LRE3-delivered NEO gene from 15 H13b (RAM 1-15, top row) and 13 H9 (RAM A-M, bottom row) hESC clonal cell lines. The spliced (retrotransposed) product is 493 bp; the unspliced (vector) product is 1396 bp. For each cell type, untransfected genomic DNA was used as a negative control.

Figure A.4: Engineered L1 retrotransposition events in cultured human embryonic stem cells

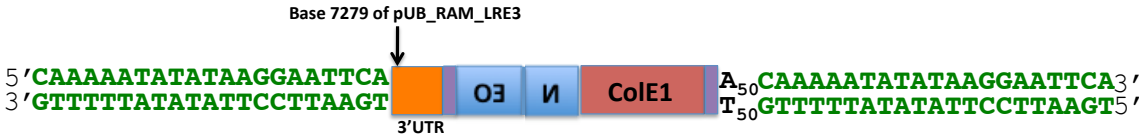
Above: for each insertion, the chromosomal location is depicted as viewed on the UCSC genome brower (http://genome.ucsc.edu/cgi-bin/hgGateway). If the insertion resides within a gene, the location within the gene (*ie,* exon or intron number) and relative transcriptional orientation to the gene, are listed.

Below: The sequence of the empty site for each insertion is shown, with the region that ultimately became the target-site duplication (TSD) in green. The filled site is depicted beneath. TSDs are highlighted in green, and the structure of the insertion, including truncation point and polyA tail length, is indicated

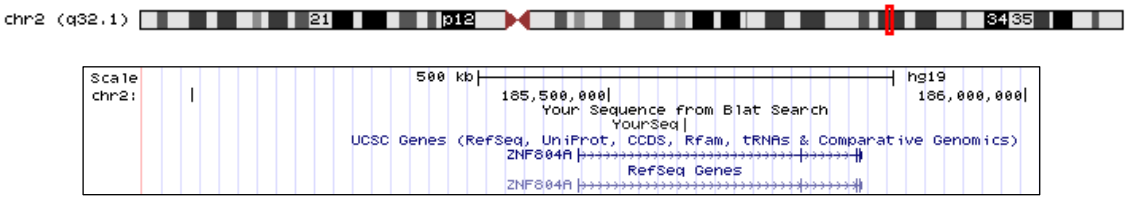
H9 hESC Line RAM_C



5′AATAT**CAAAAATATATAAGGAATTCA**AACAA3′
3′TTATA**GTTTTTATATATTCCTTAAGT**TTGTT5′

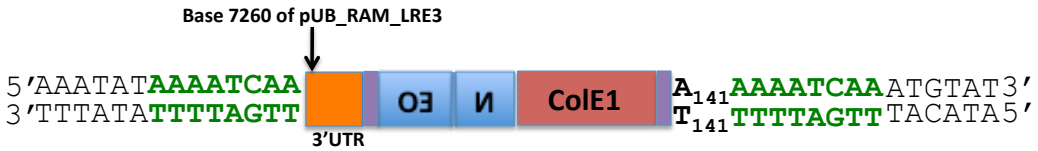


H9 hESC Line RAM_E

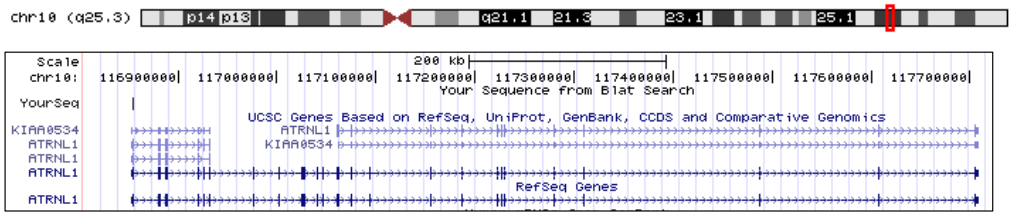


Intron 1 of ZNF804A, opposite transcriptional orientation

5′AAATAT**AAAATCAA**ATGTAT3′
3′TTTATA**TTTTAGTT**TACATA5′

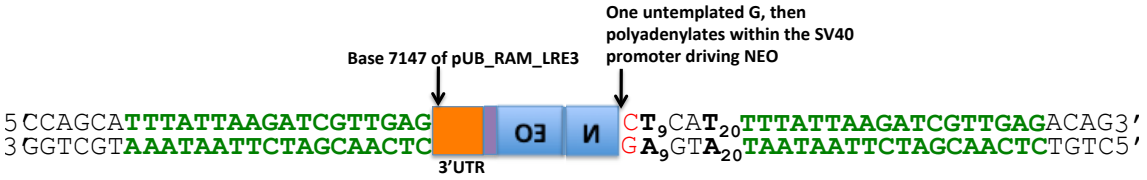


H9 hESC Line RAM_I

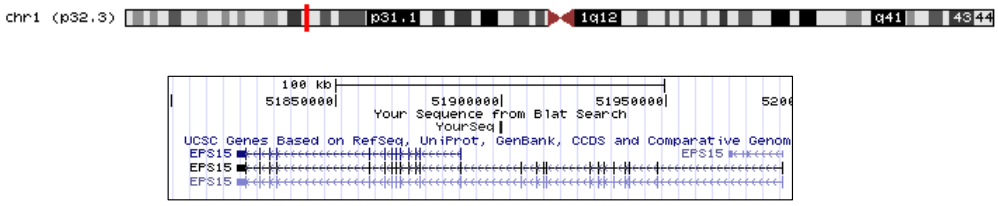


Intron 1 of ATRNL1 gene, opposite transcriptional orientation

5′CCAGCATTTATTAAGATCGTTGAGACAGA3′
3′GGTCGTAAATAATTCTAGCAACTCTGTCT5′

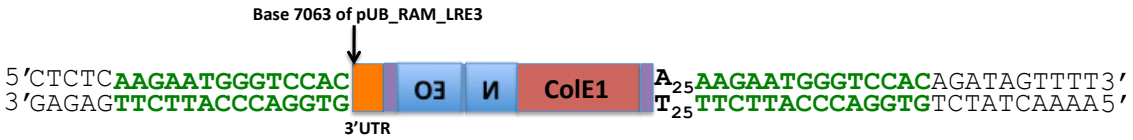


H9 hESC Line RAM_N



Intron 13 of EPS15, opposite transcriptional orientation

5′CTGTCACAGCAACAGCTCTCAAGAATGGGTCCACAGATAGTTTTTAAGTG 3′
3′GACAGTGTCGTTGTCGAGAGTTCTTACCCAGGTGTCTATCAAAAATTCAC 5′



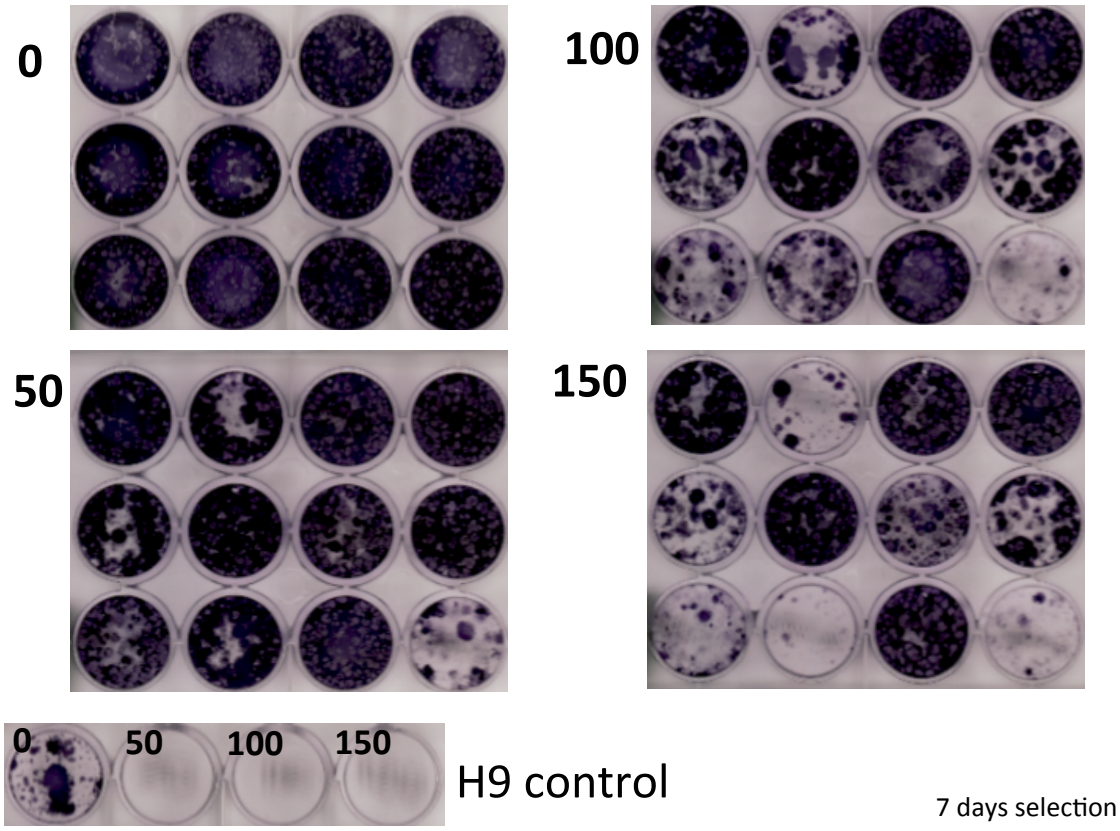


Figure A.5: A NEO reporter gene delivered by the pUB-RAM-LRE3 retrotransposition indicator can undergo silencing in hESCs

Approximately $5x10^5$ H9 hESCs from twelve different clonal cell lines harboring insertions generated using pUB-RAM-LRE3 were plated in quadruplicate in 12-well matrigel-coated dishes and fed with MEF-conditioned media. 24 hours after plating, cells were subjected to selection with zero, 50, 100, or 150 µg/ml G418. Selection was carried out for seven days; cells were then fixed and stained with crystal violet. H9 hESC clonal lines are RAM A through L; in each 12-well dish: top row A-D, middle row E-H, bottom row I-L. Untransfected H9 hESCs serve as a negative control.

References

1. Garcia-Perez, J.L., et al., *LINE-1 retrotransposition in human embryonic stem cells.* Hum Mol Genet, 2007. **16**(13): p. 1569-77.
2. Wissing, S., et al., *Endogenous APOBEC3B Restricts LINE-1 Retrotransposition in Transformed Cells and Human Embryonic Stem Cells.* J Biol Chem, 2011. **286**(42): p. 36427-37.
3. Wissing, S., et al., *Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility.* Hum Mol Genet, 2011.
4. van den Hurk, J.A., et al., *L1 retrotransposition can occur early in human embryonic development.* Hum Mol Genet, 2007. **16**(13): p. 1587-92.
5. Brouha, B., et al., *Evidence consistent with human L1 retrotransposition in maternal meiosis I.* Am J Hum Genet, 2002. **71**(2): p. 327-36.
6. Gill, D.R., et al., *Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1alpha promoter.* Gene therapy, 2001. **8**(20): p. 1539-46.
7. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
8. Wei, W., et al., *A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events.* Anal Biochem, 2000. **284**(2): p. 435-8.
9. Gilbert, N., et al., *Multiple fates of L1 retrotransposition intermediates in cultured human cells.* Mol Cell Biol, 2005. **25**(17): p. 7780-95.
10. Gilbert, N., S. Lutz-Prigge, and J.V. Moran, *Genomic deletions created upon LINE-1 retrotransposition.* Cell, 2002. **110**(3): p. 315-25.
11. Sternberg, N. and D. Hamilton, *Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites.* Journal of molecular biology, 1981. **150**(4): p. 467-86.
12. Feng, Y.Q., et al., *Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange.* Journal of molecular biology, 1999. **292**(4): p. 779-85.
13. McLeod, M., S. Craft, and J.R. Broach, *Identification of the crossover site during FLP-mediated recombination in the Saccharomyces cerevisiae plasmid 2 microns circle.* Molecular and cellular biology, 1986. **6**(10): p. 3357-67.
14. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-8.